Certifying Random Polynomials Over the Unit Sphere via Sum-of-Squares Hierarchy

Vijay Bhattiprolu*  Venkatesan Guruswami†  Euiwoong Lee‡

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213.

Abstract

Given a random $n$-variate degree-$d$ homogeneous polynomial $f$, we study the following problem:

$$\max_{\|x\|=1} f(x).$$

Besides being a fundamental problem in its own right, this random model has received attention recently due to its connection to tensor PCA. We study how the degree-$q$ Sum of Squares (SoS) hierarchy performs on this problem. Our results include:

- When $f$ is a degree-$q$ polynomial with independent random coefficients, we prove that there is a constant $c$ such that with high probability, degree-$q$ SoS satisfies

  $$\left(\frac{n}{q^{c+o(1)}}\right)^{q/4} \leq \text{SoS-Value} \leq \left(\frac{n}{q^{c-o(1)}}\right)^{q/4}.$$  

  Our upper bound improves a result of Montanari and Richard [MR14] when $q$ is large. (It is known that $\max_{\|x\|=1} f(x) \lesssim \sqrt{nq \log q}$ w.h.p.)

- When $f$ is a random degree-3 or 4 polynomial, we prove that with high probability, the degree-$q$ SoS-Value is at most $\tilde{O}(\frac{n^{3/4}}{q^{1/4}})$, $\tilde{O}(\frac{n^{1/2}}{\sqrt{q}})$ respectively. This improves on a result of Hopkins, Shi, and Steurer [HSS15] when $q$ is growing.

- Lower bounds on the degree-$q$ SoS-Value of the 2-norm of the $q$-parity-tensor of a random graph.

In providing these results, we develop tools that we believe to be useful in analysing the SoS hierarchy on other optimization problems over the sphere.
1. Introduction

We study the problem of optimizing polynomials over the unit sphere through the lens of the sum of squares hierarchy. Formally, given an \( n \)-variate degree-\( d \) homogeneous polynomial \( f \), the goal is to compute

\[
\max_{\|x\|=1} f(x).
\]

Besides being a natural and fundamental problem in its own right, it has connections to quantum information theory [BH13, BKS14], the Small Set Expansion Hypothesis (SSEH) and the Unique Games Conjecture (UGC) (via \( 2 \to 4 \) norm, see [BBH+12, BKS14]), tensor decomposition [BKS15, GM15], tensor PCA [MR14, HSS15], and planted clique (via the parity tensor, see [FK08, BV09]).

One of the popular approaches to the above problem, called the \textit{Sum of Squares Hierarchy} (SoS), proceeds by replacing a system of non-negativity constraints by a suitable sum of squares decomposition. Algorithms based on this framework are parametrized by the degree \( q \) of their SoS decomposition. Optimization over \( S^{n-1} \) via SoS has been given attention in the optimization community, where for a fixed number of variables \( n \) and degree \( d \) of the polynomial, it is known that the estimates produced by the SoS hierarchy get arbitrarily close to the true optimal solution as \( q \) increases. We refer the reader to the recent work of Doherty and Wehner [DW12] and de Klerk, Laurent, and Sun [dKLS14] and references therein for more information on convergence results. By using semidefinite programming (SDP), these algorithms run in time \( n^{O(q)} \), which is polynomial for constant \( q \). Unfortunately, known convergence results often give a non-trivial bound only when the degree parameter \( q \) is linear in \( n \).

It is natural to ask how well polynomial time algorithms can \textit{approximate} these polynomial optimization problems over a compact set. When the compact set is \( S^{n-1} \), Nesterov [Nes03] gave a reduction from Maximum Independent Set to optimizing a homogeneous cubic polynomial over \( S^{n-1} \). Formally, given a graph \( G \), there exists a homogeneous cubic polynomial \( f(G) \) such that

\[
\sqrt{1 - \frac{\alpha(G)}{\alpha(G)}} = \max_{\|x\|=1} f(x).
\]

Combined with the hardness of Maximum Independent Set [Hås96], this rules out an FPTAS for optimization over the unit sphere. Assuming the Exponential Time Hypothesis, Barak et al. [BBH+12] proved that computing \( 2 \to 4 \) norm of a matrix, a special case when \( f \) is a degree-4 homogeneous polynomial, is hard to approximate within a factor \( \exp(\log^{1/2-\varepsilon}(n)) \) for any \( \varepsilon > 0 \). See the survey of de Klerk [DK08] for approximability for other compact sets such as simplices and hypercubes.

In computer science, much attention has been given to the regime \( d \leq q \ll n \), so that the resulting algorithm runs in at most subexponential time. For deterministic polynomials, approximation guarantees have been proved for special cases including \( 2 \to q \) norms [BBH+12], nonnegative polynomials [BKS14], and some polynomials that arise in quantum information theory [BH13]. As such, there is considerable interest in tightly characterizing the approximation guarantee of the SoS hierarchy on many of these special cases.

Of particular interest to us are random polynomials, for which known results include approximation of \( 2 \to 4 \) norm of a random matrix [BBH+12], tensor decomposition [BKS15, GM15], and the 2-norm of random tensors [HSS15] (with applications to tensor-PCA). In this work, we primarily focus on the performance of SoS on optimizing random polynomials over \( S^{n-1} \). In particular for any even \( q \) (SoS is only defined for even degree), we give essentially tight upper and lower bounds on the value of the degree-\( q \) SoS relaxation of maximizing a degree-\( q \) homogeneous polynomial \( f \) with independent rademacher/gaussian coefficients over \( S^{n-1} \). We also provide upper bounds on the degree-\( q \) SoS value of random low-degree polynomials (3 and 4) (over \( S^{n-1} \)), and our results easily extend to polynomials of any degree. These results extend and improve results of Montanari and Richard [MR14] and Hopkins, Shi, and Steurer [HSS15].

Frieze and Kannan [FK08] showed a connection between the planted clique problem and the two-norm of the 3-parity-tensor, and Brubaker and Vempala [BV09] extended this to the general \( q \)-parity-tensor (with the guarantees improving with \( q \)). This hints at a natural hierarchy of relaxations for planted clique: run
degree-\(q\) SoS on the two-norm of the \(q\)-parity tensor. Our lower bound techniques extend to this setting, allowing us to show that the above hierarchy of relaxations is significantly weaker than the natural SoS formulation of planted clique.

1.1. Our Results

Given a random polynomial \(f\) of degree \(d\) it is known that \(\max_{\|x\|=1} f(x) \lesssim \sqrt{nd \log d}\) (see [TS14]). Montanari and Richard [MR14] presented a \(n^{O(d)}\)-time algorithm that can certify that the optimal value is atmost \(O(n^{d/2})\) with high probability. Hopkins, Shi, and Stuerer [HSS15] improved it to \(O(n^{4d/3})\) with the same running time. They also asked how many levels of SoS are required to certify a bound of \(n^{3/4-\delta}\) for \(d = 3\). We extend their work in two ways. First, for a random polynomial of degree \(q\) (even), we consider degree-\(q\) SoS and provide a tighter analysis. Our analysis asymptotically improves the previous results when \(q\) is growing with \(n\), and we prove an essentially matching lower bound. Secondly, we consider the case when \(d\) is fixed to be 3 or 4, and give improved results for the performance of degree-\(q\) SoS (for large \(q\)), thus answering in part, a question posed by Hopkins, Shi and Steurer [HSS15].

Note. We will state our results in the setting of random tensors because prior work has been written as such (the only difference between this model and the model of polynomials with identical, independent unit variance coefficients is that the variances of the tensor polynomial’s coefficients differ slightly as a function of \(q\)). It is however the case that all our techniques extend almost verbatim to the i.i.d. coefficient setting and one can get similarly tight results.

Model. Let \(\mathcal{A} \in (\mathbb{R}^n)^{\otimes q}\) be a \(q\)-tensor with independent, centred, sub-gaussian entries. The polynomial \(\mathcal{A}(x)\) and the 2-norm of \(\mathcal{A}\) are defined as,

\[
\mathcal{A}(x) := \sum_{i_1, \ldots, i_q \in [n]} \mathcal{A}[i_1, \ldots, i_q] x_{i_1} x_{i_2} \cdots x_{i_q} \quad \|\mathcal{A}\|_2 := \max_{\|x\|=1} \mathcal{A}(x).
\]

Degree-\(q\) Polynomials. We study the performance of degree-\(q\) SoS hierarchies on a random polynomial of degree \(q\).

Theorem 1.1 (Informal). For any any even \(q \leq n\), let \(\mathcal{A} \in (\mathbb{R}^n)^{\otimes q}\) be a \(q\)-tensor with independent, centred, sub-gaussian entries. With high probability, degree-\(q\) SoS relaxation of \(\|\mathcal{A}\|_2\) satisfies

\[
\left( \frac{n}{q^{1+o(1)}} \right)^{q/4} \leq \text{SoS-Value} \leq \left( \frac{n}{q^{1-\alpha(1)}} \right)^{q/4}
\]

This improves upon the \(O(n^{q/4})\) upper bound by Montanari and Richard [MR14].

Low Degree Polynomials. For low order tensors, we prove the following results, whose proofs together allow one to easily extend these results to tensors (read. random polynomials) of any degree.

Theorem 1.2 (Informal). Let \(\mathcal{A} \in (\mathbb{R}^n)^{\otimes 4}\) be a 4-tensor with independent, rademacher (resp. sub-gaussian) entries. Then for any even \(q \leq n\) (resp. \(q \leq n^{2/3}\)), with high probability, the degree-\(q\) SoS relaxation of \(\|\mathcal{A}\|_2\) has value at most \(O(n/\sqrt{q})\).

\[^{1}\text{[TS14] proves this for injective norm and it is not hard to specialize their proof for the two-norm case and obtain a better bound of } \sqrt{n \log d}\]
Theorem 1.3 (Informal). Let $\mathcal{A} \in (\mathbb{R}^n)^{\otimes 3}$ be a 3-tensor with independent, rademacher (resp. sub-gaussian) entries. Then for any even $q \leq n$ (resp. $q \leq n^k$ for some universal const. $\delta > 0$), with high probability, the degree-$q$ SoS relaxation of $\|\mathcal{A}\|_2$ has value at most $\widetilde{O}(n^{3/4}/q^{1/4})$.

By taking $q = n^\varepsilon$ for some $\varepsilon > 0$, we prove that $n^\varepsilon$-degree SoS can certify the optimal value is $\widetilde{O}(n^{3/4-\varepsilon/4})$. This answers a question of Hopkins, Shi, and Steurer [HSS15] regarding 3-tensors.

Remark 1.4. Our techniques for 3-tensor and 4-tensor, readily extend to the case of tensors of arbitrary order $d$ (resp. polynomials of arbitrary degree), to yield an upper bound of the form

$$2^{O(d)}(n \cdot \text{poly log})^{d/4}/q^{d/4-1/2}. \tag{1.1}$$

Combining our degree-3 and degree-4 upper bounds with the work of [HSS15] would yield improved tensor-PCA guarantees on higher levels of SoS.

Remark 1.5. Raghavendra, Rao, and Schramm [RRS16] have independently obtained similar results to Eq. (1.1). They additionally provided tight SoS upper bounds for refuting random D-Lin instances by showing SoS upper bounds for certifying the norms of certain families of random tensors.

Parity Tensor Formulation of Planted Clique. Our lower bound techniques allow us to give lower bounds for a hierarchy of tensor two-norm formulations of planted clique.

Theorem 1.6 (Informal). For any any even $q \leq n$, let $\mathcal{A} \in (\mathbb{R}^n)^{\otimes q}$ be the $q$-parity-tensor of a graph $G$ (see Section 5 for definition) With high probability, degree-$q$ SoS relaxation of $\|\mathcal{A}\|_2$ cannot distinguish between the cases (1) $G$ is drawn from $G_{n,1/2}$ and (2) $G$ is drawn from $G_{n,1/2}$ and has a planted clique of size $\sim \sqrt{n}/\log n$.

1.2. Preliminaries

1.2.1. Notation

We denote multisets with square brackets and multiset union with the square cup ($\sqcup$) symbol. $\binom{n}{k}$ represents “$n$ multichoose $k$” and $\binom{[n]}{k}$ represents the set of multisets of size $k$. We define $\binom{n}{\leq k}$ and $\binom{[n]}{\leq k}$ similarly for all multisubssets of size at most $k$. For variables $x_1, \ldots, x_n$ and a multiset $S \in \binom{[n]}{\leq k}$, $x^S$ denotes the monomial $\prod_{i \in S} x_i$. Since $S$ is a multiset, the exponent of some variables may be bigger than 1. For a multiset $S$, we denote it’s orbit, i.e. the set of all tuples obtained by permuting the elements of $S$, by $\mathcal{O}(S)$.

For a tuple $s \in [n]^k$, let $\text{multiset}(t)$ denote the multiset of it’s elements. $\oplus$ denotes tuple-concatenation. For a tuple $t$, we denote it’s orbit by $\mathcal{O}(t)$. A tuple-indexed (resp. multiset-indexed) matrix $A$ is called (degree-$q$) SoS-symmetric if for any $i, j, i', j' \in [n]^{q/2}$, $A[i, j] = A[i', j']$ (resp. $A[\text{multiset}(i), \text{multiset}(j)] = A[\text{multiset}(i'), \text{multiset}(j')]$) whenever $\text{multiset}(i) \sqcup \text{multiset}(j) = \text{multiset}(i') \sqcup \text{multiset}(j')$.

We use letters $\mathcal{A}, \mathcal{T} \in (\mathbb{R}^{[n]})^{\otimes k}$ to denote order-$k$ tensors, $A, B, M \in \mathbb{R}^{[n]^k \times [n]^k}$ for tuple-indexed matrices, and $A, B, M \in \mathbb{R}^{\binom{[n]}{\leq k} \times \binom{[n]}{\leq k}}$ for multiset-indexed matrices. Given an order-$2k$ tensor $\mathcal{A} \in (\mathbb{R}^{[n]})^{\otimes k}$, let $\text{Mat}(\mathcal{A}) \in [n]^k \times [n]^k$ be the tuple-indexed matrix defined by $\mathcal{A}[i_1, \ldots, i_{2k}] := \text{Mat}(\mathcal{A})([i_1, \ldots, i_k], (i_{k+1}, \ldots, i_{2k}))$. Similarly, given a tensor or matrix $A$, let $\text{Vec}(A)$ be the vector with the same entries. $J_{m \times n}$ denotes the $m \times n$ all-ones matrix.

1.2.2. Quotient Matrix

Definition 1.7. We say that a matrix $A \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$ is (degree-$q$) block-symmetric if for every $i, j \in [n]^{q/2}$ and $i' \in \mathcal{O}(i), j' \in \mathcal{O}(j)$, it holds that $A[i, j] = A[i', j']$. 

3
We next define a quotient matrix formed by replacing each rectangular matrix block by its maximum singular value.

**Definition 1.8.** For a block-symmetric matrix $A \in \mathbb{R}^{[n/2]\times [n/2]}$, we define a multiset indexed matrix called the $(q\text{-SoS})$ quotient matrix of $A$ and denoted by $A_Q \in \mathbb{R}^{([n]/q)\times ([n]/q)}$ as, $A_Q[i, j] := A[i', j'] \sqrt{|\mathcal{O}(i)||\mathcal{O}(j)|}$ for all $i, j \in \left([\frac{n}{q}/2]\right)$, where $i'$ and $j'$ are any tuples in $\mathcal{O}(i)$ and $\mathcal{O}(j)$ respectively.

The quotient matrix has multiple useful properties, like preserving inner products, trace, and PSDness. The spectral norm of the quotient matrix is also an upper bound on the spectral norm of its corresponding block-symmetric matrix. Refer to Section A for proofs and more details.

### 1.2.3. SoS Hierarchy

Let $\mathbb{R}[x]_{\leq q}$ be the vector space of polynomials with real coefficients in variables $x = (x_1, \ldots, x_n)$, of degree at most $q$. For an even integer $q$, the degree-$q$ pseudo-expectation operator is a linear operator $\bar{E} : \mathbb{R}[x]_{\leq q} \mapsto \mathbb{R}$ such that

1. $\bar{E}[1] = 1$ for the constant polynomial $1$.
2. $\bar{E}[p_1 + p_2] = \bar{E}[p_1] + \bar{E}[p_2]$ for any polynomials $p_1, p_2 \in \mathbb{R}[x]_{\leq q}$.
3. $\bar{E}[p^2] \geq 0$ for any polynomial $p \in \mathbb{R}[x]_{\leq q/2}$.

The pseudo-expectation operator $\bar{E}$ can be described by the **moment matrix** $M \in \mathbb{R}^{([n]/q)\times ([n]/q)}$ such that $M[S, T] = \bar{E}[x^S x^T]$.

For a homogeneous degree $d$ polynomial $p$, we say a matrix $M_p \in \mathbb{R}^{[n]^d/2\times [n]^d/2}$ is a degree-$d$ matrix representation of $p$ for all $x$, $p(x) = (x^\otimes d/2)^T M x^\otimes d/2$. For polynomials $p_1, p_2$, let $p_1 \preceq p_2$ denote that $p_1 - p_2$ is a sum of squares. It is easy to verify that if $p_1, p_2$ are homogeneous degree $d$ polynomials and there exist matrix representations $M_{p_1}$ and $M_{p_2}$ of $p_1$ and $p_2$ respectively, such that $M_{p_1} - M_{p_2} \succeq 0$, then $p_1 - p_2 \preceq 0$. We will also use the Pseudo-Cauchy-Schwarz inequality (see [BKS14] for a clean proof) which states that $\bar{E}[p_1 p_2] \leq (\bar{E}[p_1^2] \bar{E}[p_2^2])^{1/2}$ for any $p_1, p_2$ of degree at most $q/2$.

Given a polynomial $p$ of degree $d \leq q$, and the optimization problem $\max \|x\|_q = 1 p(x)$, we are interested in the relaxation: $\max \bar{E}[p] \ s.t. \ \bar{E}[\|x\|_q] = 1$, over all valid degree-$q$ pseudo-expectation operators $\bar{E}$.

### 1.3. Techniques

- **Quotient matrix.** When studying SoS upper and lower bounds one often considers the tuple-indexed SoS moment matrix. Every such matrix however, has a large common kernel prompting one to instead consider a smaller multiset or set indexed moment matrix. We introduce the simple notion of a quotient matrix that allows us to switch back and forth between an almost-SoS-symmetric matrix $A$ and a multiset indexed matrix $\tilde{A}$, while still preserving properties like trace, inner products, PSDness, and upper bounds on the spectrum of $A$. This notion is one of the main reasons we are able to give tight upper and lower bounds on random degree-$q$ polynomials.

- **Higher Order Mass-Shifting.** Our approach to upper bounds on a random low degree (say $d$) polynomial $f$, is through exhibiting a matrix representation of $f^{q/d}$ that has small operator norm. Such approaches have been used previously for low-degree SoS upper bounds. However when the SoS degree is constant, the set of SoS symmetric positions is also a constant and the usual approach is to shift all the mass towards the diagonal which is of little consequence when the SoS-degree is low. In contrast, when the SoS-degree is large, many non-trivial issues arise when shifting mass across SoS-symmetric positions, as there are many permutations with very large operator norm. In our setting, mass-shifting approaches like symmetrizing and diagonal-shifting fail quite spectacularly to provide
good upper bounds. For our upper bounds, we crucially exploit the existence of certain good shifts, as well as the large size of the set of SoS-symmetric positions. To our knowledge, our upper bounds are the first instance of such a ‘matrix-representation’ upper bound that exploits the large number of SoS-symmetric positions.

- **Square Moments of Wigner Semicircle Distribution.** Often when one is giving SoS lower bounds, one has a linear functional that is not necessarily PSD and a natural approach is to fix it by adding a pseudo-expectation operator with large value on square polynomials (under some normalization). Finding such operators however, is quite a non-trivial task when the SoS-degree is growing. We show that if \(x_1, \ldots, x_n\) are independently drawn from the Wigner semicircle distribution, then for any polynomial \(p\) of any degree, \(E[p^2]\) is large (with respect to the degree and coefficients of \(p\)). Our proof crucially relies on knowledge of the Cholesky decomposition of the moment matrix of the univariate Wigner distribution. This tool was useful to us in giving tight \(q\)-tensor lower bounds, as well as in giving degree-\(q\) lower bounds for the \(q\)-parity-tensor, and we believe it to be generally useful for high degree SoS lower bounds.

2. **2-norm of a Random \(q\)-Tensor**

2.1. **Degree-\(q\) SoS Formulation for \(q\)-Tensor 2-Norm**

Let \(q\) be even and \(\mathcal{A} \in (\mathbb{R}^{n})^{\otimes q}\) be a \(q\)-tensor with independent, centred, sub-gaussian entries of sub-gaussian norm at most \(K_2\). The degree-\(q\) SoS formulation is:

\[
\max \ E \left[ \langle \mathcal{A}, x^{\otimes q} \rangle \right] \quad s.t. \quad E \left[ ||x||_2^q \right] = 1
\]

2.2. **Upper Bound**

2.2.1. **Analysis Overview**

Let \(A := \text{Mat}(\mathcal{A}) \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}\). Let \(\tilde{E}\) be the pseudo-expectation operator returned by the program above. Let \(f(x) := \langle \mathcal{A}, x^{\otimes q} \rangle\). Our goal now is to show that there is some matrix representation \(B\) of \(f\), such that w.h.p. \(||B||_2\) is small. Motivated by ideas from our 4-tensor upper bound (Section 3) we define our matrix representation \(B\) as follows:

\[
\forall i, j \in [n]^{q/2}, \quad B[i, j] := \frac{1}{(q/2)!} \sum_{\pi, \sigma \in S_q} A[\pi(i), \sigma(j)] = \frac{1}{|O(i)||O(j)|} \sum_{\tau \in O(i), \rho \in O(j)} A[i', j']
\]

In particular, note that the entries of \(B_0\) are independent - a fact we will use later. We shall show that \(||B||_2 \lesssim n^{q/4}/q^{q/4}\) with high probability. Then we would have \(||B||_2 I - B \succeq 0\), which implies that \(||B||_2 \langle x \rangle^{q} - f(x) \succeq 0\). Thus we would obtain w.h.p. \(\tilde{E}[f] \lesssim \frac{n^{q/4}}{q^{q/4}} \tilde{E}[||x||^q] = \frac{n^{q/4}}{q^{q/4}}\) as desired.

2.2.2. **Analysis**

**Lemma 2.1.** There exist constants \(C, c > 0\) (which depend only on \(K_2\)), such that with probability \(1 - e^{-ct^2}\), one has

\[
||B||_2 \leq C \left( \left( \frac{n}{q/2} \right) \right)^{1/2} + t.
\]

**Proof.** Observe that \(B\) is block-symmetric and thus the quotient matrix \(B_0\) is well defined. Observe that by Lemma B.3, \(B_0\) has independent centred sub-gaussian entries with sub-gaussian norm at most \(O(K_2)\). Thus, combining known results about \(||B_0||_2\) with Lemma A.6 yields the claim. \(\blacksquare\)
Thus we obtain

**Theorem 2.2.** For any even \( q \leq n \), let \( \mathcal{A} \in \mathbb{R}^{[n]^q} \) be a \( q \)-tensor with independent centred sub-gaussian entries of sub-gaussian norm at most \( K_2 \). Then there exist constants \( C, c \) depending only on \( K_2 \), such that w.p. \( 1 - e^{-ct^2} \), the degree-\( q \) SoS program in Section 2.1 certifies that

\[
\| \mathcal{A} \|_2 \lesssim C \left( \left( \frac{n}{q/2} \right) \right)^{1/2} + t.
\]

### 2.3. Lower Bound

#### 2.3.1. Overview of Approach

On a high level we follow the philosophy of Hopkins et al. [HSS15] who gave a degree-4 SoS lower bound on the two-norm of a 4-tensor. The approach is to start with a linear functional that has good objective value by design but is not necessarily PSD, and fix it by adding an appropriate pseudo-expectation operator with large square moments. While it is fairly clear what such a distribution is when the operator is a degree-4 operator, this problem becomes quite non-trivial when we would like degree-\( q \) operators for growing \( q \).

We next give a detailed breakdown of the approach.

1. Given a random tensor \( \mathcal{A} \) and its matrix version \( A := \text{Mat}(\mathcal{A}) \), we construct a moment matrix \( M \) which is (1) degree-\( q \) SoS symmetric, (2) PSD, and (3) has a good inner product with \( A \). This is easily extended to the desired degree-\( q \) pseudo-expectation operator.
2. We will define \( M \) by defining its quotient matrix \( M_Q \) and leveraging the fact that the quotient matrix preserves trace, inner products and PSDness. (This step is crucial to our obtaining tight lower bounds)
3. We start with the natural observation that (symmetrized version of) \( A \) has a good inner product with itself, prompting us to consider \( A_Q \) as a choice for \( M_Q \).
4. However \( A_Q \) has negative eigenvalues. To fix this, one would like to add the identity matrix to increase all eigenvalues, but unfortunately, the standard identity matrix does not satisfy SoS-symmetry. This prompts the search for an SoS-symmetric matrix \( W \), such that \( A_Q + \alpha \cdot W \) is PSD.
5. Lastly, we need to normalize by trace so that we satisfy the unit sphere constraint. Thus our choice of \( M_Q \) is of the form \( (A_Q + \alpha \cdot W) / \text{Tr}(A_Q + \alpha \cdot W) \). Thus we need an SoS-symmetric matrix \( W \) that has large ratio of minimum eigenvalue to trace.

#### 2.3.2. Wigner Moment Matrix

In this section, we construct an SoS-symmetric matrix \( W \in \mathbb{R}^{(\begin{pmatrix} n \\ q/2 \end{pmatrix}) \times (\begin{pmatrix} n \\ q/2 \end{pmatrix})} \) such that \( \lambda_{\min}(W) / \text{Tr}(W) \geq 1 / (2^{q+1} \cdot \left( \begin{pmatrix} n \\ q/2 \end{pmatrix} \right) \cdot \left( \begin{pmatrix} q/2 \\ q/2 \end{pmatrix} \right) ) \), i.e. the ratio of the minimum eigenvalue to the average eigenvalue is at least \( 1 / 2^{q+1} \).

**Theorem 2.3.** For any positive integer \( n \) and any positive even integer \( q \), there exists a matrix \( W \subseteq \mathbb{R}^{\begin{pmatrix} n \\ q/2 \end{pmatrix} \times \begin{pmatrix} n \\ q/2 \end{pmatrix}} \) that satisfies the following three properties: (1) \( W \) is degree-\( q \) SoS symmetric. (2) The minimum eigenvalue of \( W \) is at least \( \frac{1}{2} \). (3) Each entry of \( W \) is in \( [0, 2^q] \).

Theorem 2.3 is proved by explicitly constructing independent random variables \( x_1, \ldots, x_n \) such that for any \( n \)-variate polynomial \( p(x_1, \ldots, x_n) \) of degree at most \( \frac{q}{2} \), \( \mathbb{E}[p^2] \) is bounded away from 0. The proof consists of three parts. The first part shows the existence of a desired distribution for one variable \( x_i \). The second part uses induction to prove that \( \mathbb{E}[p^2] \) is bounded away from 0. The third part constructs \( W \subseteq \mathbb{R}^{\begin{pmatrix} n \\ q/2 \end{pmatrix} \times \begin{pmatrix} n \\ q/2 \end{pmatrix}} \) from the distribution defined.
Wigner Semicircle Distribution and Hankel Matrix. Let $k$ be a positive integer. In this part, the rows and columns of all $(k+1) \times (k+1)$ matrices are indexed by $\{0, 1, \ldots, k\}$. Let $T$ be a $(k+1) \times (k+1)$ matrix where $T[i, j] = 1$ if $i - j = 1$ and $T[i, j] = 0$ otherwise. Let $e_0 \in \mathbb{R}^{k+1}$ be such that $(e_0)_i = 1$ and $(e_0)_i = 0$ for $1 \leq i \leq k$. Let $R \in \mathbb{R}^{(k+1) \times (k+1)}$ be defined by $R := [e_0, Te_0, T^2 e_0, \ldots, T^k e_0]$. Let $R_0, \ldots, R_k$ be the columns or $R$ so that $R_i = T^i e_0$. It turns out that $R$ is closely related to the number of ways to consistently put parentheses. Given a string of parentheses '(' or ')', we call it consistent if any prefix has at least as many '(' as ')'. For example, (((() is consistent, but ()(() is not.

Claim 2.4. $R[i, j]$ is the number of ways to place $j$ parentheses '(' or ')' consistently so that there are $i$ more '(' than ')'.

The proof appears in Appendix C. Easy consequences of the above claim are (1) $R[i, i] = 1$ for all $0 \leq i \leq k$, and $R[i, j] = 0$ for $i > j$, and (2) $R[i, j] = 0$ if $i + j$ is odd, and $R[i, j] \geq 1$ if $i \leq j$ and $i + j$ is even.

Given a sequence of $m_0 = 1, m_1, m_2, \ldots$ of real numbers, the Hamburger moment problem asks whether there exists a random variable $W$ supported on $\mathbb{R}$ such that $\mathbb{E}[W^i] = m_i$. It is well-known that there exists a unique such $W$ if for all $k \in \mathbb{N}$, the Hankel matrix $H \in \mathbb{R}^{(k+1) \times (k+1)}$ defined by $H[i, j] := \mathbb{E}[W^{i+j}]$ is positive definite [Sim98]. Since our construction of $H \in \mathbb{R}^{(k+1) \times (k+1)}$ ensures its positive definiteness for any $k \in \mathbb{N}$, there exists a unique random variable $W$ such that $\mathbb{E}[W^i] = 0$ if $i$ is odd, $\mathbb{E}[W^i] = C_i^2$ if $i$ is even. It is known as the Wigner semicircle distribution with radius $R = 2$.

Remark 2.5. Some other distributions (e.g., Gaussian) will give an asymptotically weaker bound. Let $G$ be a standard Gaussian random variable. The quantitative difference comes from the fact that $\mathbb{E}[W^{2i}] = C_i = \frac{1}{\pi} \left( \frac{2i}{i + 1} \right) \leq 4^i$ while $\mathbb{E}[G^{2i}] = (2i - 1)!! > 2^{\Omega(\log i)}$.

Multivariate Distribution. Fix $n$ and $q$. Let $k = \frac{q}{2}$. Let $H \in \mathbb{R}^{(k+1) \times (k+1)}$ be the Hankel matrix defined as above, and $W$ be a random variable sampled from the Wigner semicircle distribution. Consider $x_1, \ldots, x_n$ where each $x_i$ is an independent copy of $\frac{W}{\sqrt{n}}$ for some large number $N$ to be determined later. Our $W$ is later defined to be $W[S, T] = \mathbb{E}[x_1^{q/2}] \cdot N^q$ so that the effect of the normalization by $N$ is eventually cancelled, but large $N$ is needed to prove the induction that involves non-homogeneous polynomials.

We study $\mathbb{E}[p(x)^2]$ for any $n$-variate (possibly non-homogeneous) polynomial $p$ of degree at most $k$. For a multivariate polynomial $p = \sum_{S \in \binom{[n]}{\frac{q}{2}}} p_S x^S$, let $\|p\|^2 := \sum_S p_S^2$. For $0 \leq m \leq n$ and $0 \leq l \leq k$, let $\sigma(m, l) := \inf_p \mathbb{E}[p(x)^2]$ where the infimum is taken over polynomials $p$ such that $\|p\| = 1$, $\deg(p) \leq l$, and $p$ depends only on $x_1, \ldots, x_m$.

Lemma 2.6. There exists $N := N(n, k)$ such that $\sigma(m, l) \geq \frac{(1 - \frac{m}{N})}{N^m}$ for all $0 \leq m \leq n$ and $0 \leq l \leq k$.

Proof. We prove the lemma by induction on $m$ and $l$. When $m = 0$ or $l = 0$, $p$ becomes the constant polynomial 1 or $-1$, so $\mathbb{E}[p^2] = 1$.

Fix $m, l > 0$ and a polynomial $p = p(x_1, \ldots, x_m)$ of degree at most $l$. Decompose $p = \sum_{i=0}^{l-1} p_i x_m^i$ where each $p_i$ does not depend on $x_m$. The degree of $p_i$ is at most $l - i$.

$$\mathbb{E}[p^2] = \mathbb{E}[\left( \sum_{i=0}^{l-1} p_i x_m^i \right)^2] = \sum_{0 \leq i, j \leq l} \mathbb{E}[p_i p_j] \mathbb{E}[x_m^{i+j}].$$
Let $\Sigma = \text{diag}(1, \frac{1}{2}, \ldots, \frac{1}{N}) \in \mathbb{R}^{(l+1) \times (l+1)}$. Let $H_l \in \mathbb{R}^{(l+1) \times (l+1)}$ be the submatrix of $H$ with the first $l + 1$ rows and columns. The rows and columns of $(l + 1) \times (l + 1)$ matrices are still indexed by $\{0, \ldots, l\}$. Define $R_t \in \mathbb{R}^{(l+1) \times (l+1)}$ similarly from $R$, and $r_t (0 \leq t \leq l)$ be the $t$th column of $(R_t)^T$. Note $H_l = (R_t)^T R_l = \sum_{t=0}^l r_t r_t^T$. Let $H' = \Sigma H_l \Sigma$ such that $H'[i, j] = \mathbb{E}[x_m^t]$. Finally, let $\mathcal{P} \in \mathbb{R}^{(l+1) \times (l+1)}$ be defined such that $P[i, j] = \mathbb{E}[p_t p_j]$. Then $\mathbb{E}[p_t^2]$ is equal to

$$
\text{Tr}(PH') = \text{Tr}(P\Sigma H_l \Sigma) = \text{Tr} \left( P \Sigma \left( \sum_{t=0}^l r_t r_t^T \right) \right) = \sum_{t=0}^l \mathbb{E} \left[ \frac{1}{N^t} + \frac{(r_t)_{t+1}}{N^{t+1}} + \cdots + \frac{(r_t)_{t}}{N^t} \right],
$$

where the last step follows from the fact that $(r_t)_j = 0$ if $j < t$ and $(r_t)_t = 1$. Consider the polynomial

$$q_t := \frac{1}{N^t} + \frac{(r_t)_{t+1}}{N^{t+1}} + \cdots + \frac{(r_t)_{t}}{N^t}.
$$

Since $p_t$ is of degree at most $l - i$, $q_t$ is of degree at most $l - t$. Also recall that each entry of $R$ is bounded by $2^k$. By the triangle inequality,

$$\|q_t\| \geq \frac{1}{N^t} \left( \|p_t\| - \|p_{t+1}\| \frac{(r_t)_{t+1}}{N^l} + \cdots + \|p_t\| \frac{(r_t)_{t}}{N^t} \right) \geq \frac{1}{N^t} \left( \|p_t\| - \frac{k2^k}{N} \right),
$$

$$\|q_t\|^2 \geq \frac{1}{N^{2t}} \left( \|p_t\|^2 - \frac{k2^k}{N} \right).$$

Finally,

$$\mathbb{E}[p_t^2] = \sum_{t=0}^l \mathbb{E}[q_t^2] \geq \sum_{t=0}^l \|q_t\|^2 \geq \sum_{t=0}^l \sigma(m - 1, l - t) \cdot \frac{1}{N^{2t}} \left( \|p_t\|^2 - \frac{k2^k}{N} \right) \geq \frac{1}{N^{2l}} \left( 1 - \frac{m - 1}{2N} \right) \cdot \sum_{t=0}^l \frac{1}{N^{2t}} \left( \|p_t\|^2 - \frac{k2^k}{N} \right) \geq \frac{1}{N^{2l}} \cdot \left( 1 - \frac{2k2^k}{N} \right).$$

Take $N := 4nk2k$ so that $(1 - \frac{m - 1}{2N}) \cdot (1 - \frac{2k2^k}{N}) \geq 1 - \frac{m - 1}{2N} - \frac{2k2^k}{N} = 1 - \frac{m}{N}$. This completes the induction and proves the lemma.

**Construction of $\mathcal{W}$.** We now prove Theorem 2.3. Given $n$ and $q$, let $k = \frac{q}{2}$, and consider random variables $x_1, \ldots, x_n$ above. Let $\mathcal{W} \in \mathbb{R}\left[n^{\frac{q}{2}} \times \left(n^{\frac{q}{2}}\right)\right]$ be such be that for any $S, T \in [n]^k$, $\mathbb{W}_{ST} = \mathbb{E}[x^S \cdot x_T] \cdot N^{2k}$. By definition, $\mathcal{W}$ is degree-$q$ SoS symmetric. Since each entry of $\mathcal{W}$ corresponds to a monomial of degree exactly $q$ and each $x_i$ is drawn independently from the Wigner semicircle distribution, each entry of $\mathcal{W}$ is at most the $\frac{q}{2}$th Catalan number $C_{\frac{q}{2}} \leq 2^q$. For any unit vector $p = (p_s)_{S \subseteq [n]^{\frac{q}{2}}} \in \mathbb{R}\left([n]^{\frac{q}{2}}\right)$, Lemma 2.6 shows $p^T \mathcal{W} p = \mathbb{E}[p^2] \cdot N^{2k} \geq \frac{1}{2}$ where $p$ also represents a degree-$k$ homogeneous polynomial $p(x_1, \ldots, x_n) = \sum_{S \subseteq [n]^{\frac{q}{2}}} p_S x_S$. Therefore, the minimum eigenvalue of $\mathcal{W}$ is at least $\frac{1}{2}$.

**2.3.3. Candidate Moment Matrix**

Given a 4-tensor $\mathcal{A}$ with centred sub-gaussian entries of norm $O(1)$, let $A := \text{Mat}(\mathcal{A})$, and define $B \in \mathbb{R}\left(n^{\frac{q}{2}} \times n^{\frac{q}{2}}\right)$ and $A, B, M \in \mathbb{R}[n^{\frac{q}{2}} \times n^{\frac{q}{2}}]$ as,

$$\forall i, j \in [n]^{\frac{q}{2}}, \quad A[i, j] := \frac{1}{|O(i \oplus j)|} \sum_{i \oplus j \in O(i \oplus j)} A[i', j']$$

(symmetrizing $A$)
\[ \mathcal{B} := \overline{A}_0 + C 2^q \left( \binom{n}{q/2} \right) \frac{1}{2} \mathcal{W} \]  

(C to be determined shortly)

\[ \text{B is defined s.t. } B_Q := \mathcal{B}, \quad M := \frac{B}{\text{Tr}(B)} \]

M is the candidate moment matrix that we will extend to a pseudo-expectation operator.

### 2.3.4. PSDness

**Lemma 2.7 (Proof in Appendix: Lemma C.2).** There exist universal constants \( C, c > 0 \), such that with probability \( 1 - e^{-c^2} \), one has \( \|A_0\| \leq C 2^{O(q)} n^{q/4}/q^{q/4} + t \).

**Observation 2.8.** By Lemma 2.7 and properties of \( \mathcal{W} \), there exists a constant \( C > 0 \) such that with probability \( 1 - e^{-n^{O(q)}} \), \( \mathcal{B} \) is PSD.

**Lemma 2.9.** \( M \) is PSD with high probability.

**Proof.** It suffices to show that \( B \) is PSD w.h.p. since \( \text{Tr}(B) \) would then be non-negative w.h.p. Now by combining Observation 2.8 and Lemma A.7, we see that \( B \) is PSD w.h.p.

### 2.3.5. Objective Function’s Value on \( M \)

**Observation 2.10.** \( \text{Tr}(B) = 2^{\Theta(q)} (n/q)^{3q/4} \) with high probability.

**Proof.** Since each entry of \( \mathcal{W} \) is bounded by \( 2^q \), \( \text{Tr}(\mathcal{B}) = \text{Tr}(\overline{A}_0) + 2^{\Theta(q)} (n/q)^{3q/4} \). We then obtain that \( \text{Tr}(\mathcal{B}) = 2^{\Theta(q)} (n/q)^{3q/4} \) since \( \text{Tr}(\overline{A}_0) \) is bounded by \( O(n^{q/4}) \) with high probability. Lastly, the claim follows by Observation A.4.

**Theorem 2.11.** For any even \( q \leq n \), let \( \mathcal{A} \in \mathbb{R}^{n^q} \) be a \( q \)-tensor with independent centred sub-gaussian entries of sub-gaussian norm \( O(1) \). Then the SoS value of the formulation from Section 2.1 is at least \( 2^{-O(q)} n^{q/4}/q^{q/4} \) with high probability.

**Proof.** \( M \) is PSD by Lemma 2.9 and \( \text{Tr}(M) = 1 \) by construction. Lastly we have,

\[
\langle \overline{A}, M \rangle = \frac{\langle \overline{A}, B \rangle}{\text{Tr}(B)} = \frac{\langle \overline{A}_0, B_Q \rangle}{\text{Tr}(\mathcal{B})} \geq \frac{\langle \overline{A}_0, \overline{A}_0 \rangle + 2^{O(q)} n^{q/4} \langle \overline{A}_0, \mathcal{W} \rangle}{\text{Tr}(\mathcal{B})} \geq \frac{2^{-O(q)} Q(n/q)^q}{\text{Tr}(\mathcal{B})} \quad \text{w.h.p.} \\
\geq \frac{2^{-O(q)} n^{q/4}}{q^{q/4}} \quad \text{w.h.p.} \quad \text{(by Observation A.5)}
\]

Now we construct the pseudo-expectation operator \( \overline{E} \). For any \( S \in \left( \left[ \frac{n}{q} \right] \right) \), let \( \overline{E}[x^S] = M[i, j] \) for any \( i, j \in \left[ \frac{n}{q} \right]^{q/2} \) such that \( S = \text{multiset}(i) \sqcup \text{multiset}(j) \). It is well-defined since \( M \) is degree-\( q \) SoS symmetric. Finally, \( \overline{E}[x^S] = 0 \) for all \( S \in \left( \left[ \frac{n}{q} \right] \right) \). \( \text{Tr}(M) = 1 \) implies \( \overline{E}[\|x\|_2^2] = 1 \), and \( \overline{E}[\langle \mathcal{A}, x^{\otimes q} \rangle] = \langle A, M \rangle = \langle A, \overline{A}_0 \rangle = \langle A, M \rangle = 2^{-O(q)} n^{q/4}/q^{q/4} \) w.h.p. This completes the lower bound.
3. 2-Norm of a Random 4-Tensor

3.1. Degree-\(q\) SoS Formulation for 4-Tensor 2-Norm

Let \(A \in \mathbb{R}^{[n]^4}\) be a 4-tensor with independent rademacher or centred sub-gaussian entries (with the maximum of the sub-gaussian norms being \(K_2\)). The degree-\(q\) SoS formulation for certifying an upper bound on \(\|A\|_2\) is: \(\max E[\langle A, x_{\otimes 4}\rangle] \text{ s.t. } E[\|x\|_2^2] = 1\)

3.2. Analysis Overview

Assume \(q/4\) is a power of 2 as this only changes our claims by constants. Let \(\mathbb{R}^{[n]^2 \times [n]^2} \ni T := \text{Mat}(A)\). Let \(T = \widetilde{T} + \bar{T}/2\), a matrix with symmetric independent (above diagonal) centred entries. Let \(f(x) := \langle A, x_{\otimes 4}\rangle\) and let \(A := T_{\otimes q/4}\). Then \(A\) is a matrix representation of \(f_{q/4}\).

Let \(\tilde{E}\) be the pseudo-distribution returned by the program above. Our goal now is to show that there is some matrix representation \(B\) of \(f_{q/4}\), such that \(\tilde{E}\) is small. It turns out that the choice of matrix representation here is critical. It is not hard to see that \(\text{Vec}(T)_{\otimes q/8} (\text{Vec}(T)_{\otimes q/8})^T\) is a representation that has spectral norm \(\sim n^{q/2}\), while \(A\) is a representation that has spectral norm \(\sim n^{q/4}\). However, this is still not enough for our purposes (we require \(\sim n^{q/4}/q^{q/8}\)). To obtain a representation with the desired spectral norm, the natural approach would be to reduce the variance of each entry of the representation by averaging the entries across all SoS-symmetric positions. Another natural approach is to use a representation with all the mass being shifted ‘close’ to the diagonal. It turns out that both these approaches fail for similar reasons, namely, these representations are highly correlated with low-rank representations of \(f_{q/4}\). To elaborate, since trace is fixed upto \(q^{\Theta(q)}\) factors across all representations, lower rank would imply bigger eigenvalues. Thus we need a representation that is ‘far’ from being low-rank. It turns out that the right notion of ‘far’ here, is that specifying either just a row or just a column of an entry \(e\) in the representation, should not by itself reveal any entries of \(T\) that the entry \(e\) depends on (note that any entry of a representation of \(f_{q/4}\) would be a sum over products of \(q/4\) entries of \(T\)). At the same time, we would like to average over a sufficiently large subset of the SoS-symmetric positions, in order to reduce variance.

To this end, consider the following matrix \(B\):

\[
\forall i, j \in [n]^{q/2}, \quad B[i, j] := \frac{1}{(q/2)!} \sum_{\pi, \sigma \in [q/2]} A[\pi(i), \sigma(j)] = \frac{1}{|O(i)||O(j)|} \sum_{\pi \in E(i), j \in E(j)} A[i', j']
\]

Above, it is critical that we use \(A = T_{\otimes q/4}\) as the starting point of the averaging and not \(\text{Vec}(T)_{\otimes q/8} (\text{Vec}(T)_{\otimes q/8})^T\). We shall show \(\|B\|_2 = \tilde{O}(n/\sqrt{q})^{q/4}\) w.h.p. which would imply that \(\tilde{E}[f_{q/4}] \leq \tilde{O}(n/\sqrt{q})^{q/4}\). Then by Pseudo-Cauchy-Schwarz one obtains \(\tilde{E}[f] \leq \tilde{O}(n/\sqrt{q})^{q/4}\) as desired. We employ the trace method to bound the spectral norm of \(B\). This however, requires some graph theoretic ideas and careful counting of special types of permutations. For any \(i^1, \ldots, i^p \in [n]^{q/2}\), let \(#(i^1, \ldots, i^p)\) denote the number of distinct elements of \([n]\) that occur in \(i^1, \ldots, i^p\).

At a high level, the key steps in the analysis are as follows:

A) If \(\tilde{E}[B[i^1, i^2] B[i^2, i^3] \ldots B[i^p, i^1]] \neq 0\), then \(#(i^1, \ldots, i^p)\) is small (uses connectivity of an appropriately defined graph).

B) If \(#(i^1, \ldots, i^p)\) is small, then \(\tilde{E}[B[i^1, i^2] B[i^2, i^3] \ldots B[i^p, i^1]]\) is small (uses careful estimates on certain special permutations). This is the most technical part of the analysis.
3.3. Analysis

For any $i^1, \ldots, i^p \in [n]^{q/2}$ let $e_k$ denote the $k$-th smallest element in $\bigcup_{\ell \in j} \{ i^\ell \}$. For any $c^1, \ldots, c^s \in [q/2]^p$, let

$$C(c^1 \ldots c^s) := \left\{ (i^1, \ldots, i^p) \mid \#(i^1, \ldots, i^p) = s, \forall k \in [s], \ell \in [p], e_k \text{ appears } c^\ell_k \text{ times in } i^\ell \right\}$$

**Observation 3.1.**

$$\left| \left\{ (c^1, \ldots, c^s) \in ([q/2]^p)^s \mid C(c^1, \ldots, c^s) \neq \emptyset \right\} \right| \leq 2^{O(pq)} p^{pq/2}$$

if $C(c^1, \ldots, c^s) \neq \emptyset$ then $|C(c^1, \ldots, c^s)| \leq \frac{n^s}{s!} \prod_{\ell \in [p]} c^\ell_1! \ldots c^\ell_s!$

**Lemma 3.2.** Consider any $i^1, \ldots, i^p \in [n]^{q/2}$ such that $(i^1, \ldots, i^p) = s$. Assume $T$ has independent centred sub-gaussian entries with the maximum of the sub-gaussian norms being $K_2$. Then for any $j^1, k^1 \in O(i^1), \ldots, j^p, k^p \in O(i^p)$, we have

$$\mathbb{E} \left[ \prod_{\ell \in [p]} T \left[ j^\ell_1 j^\ell_2, k^\ell_1 k^\ell_2 \right] \prod_{\ell \in [p]} \frac{J_{q/2-\ell} j^\ell_q \cdots k^\ell_q}{q_{q/2-\ell}} \right] \leq 2^{O(pq)} K_2^{pq/4} (pq)^{pq/8 - s/4}$$

**Proof.** Observe that the expectation is taken over a product of $pq/4$ terms. It is easy to verify that $s$ of these terms consist of at least $s/2$ distinct terms of multiplicity at most two. Thus there are $pq/4 - s$ remaining terms in the product with possibly higher multiplicities. Thus the LHS in the claim is upper bounded by

$$\max_{s_1 + \cdots + s_t \leq pq/4 - s/2} \prod_{\ell \in [p]} \mathbb{E} \left[ X^{s_\ell + 2} \right] \leq 2^{O(pq)} K_2^{pq/4} (pq)^{pq/8 - s/4} \quad \text{(by Definition B.1)}$$

where $X$ is a sub-gaussian r.v. with sub-gaussian norm at most $K_2$.

**Lemma 3.3.** Consider any $c^1, \ldots, c^s \in [q/2]^p$ and $(i^1, \ldots, i^p) \in C(c^1, \ldots, c^s)$. We have

$$\mathbb{E} \left[ \prod_{\ell \in [p]} B \left[ j^\ell_1, j^\ell_2 \right] B \left[ k^\ell_1, k^\ell_2 \right] \ldots B \left[ i^p, i^1 \right] \right] \leq 2^{O(pq)} \frac{p^{5pq/8}}{q^{3pq/8}} \prod_{\ell \in [p]} c^\ell_1! \ldots c^\ell_s! \quad \text{(rademacher entries in T)}$$

$$\mathbb{E} \left[ \prod_{\ell \in [p]} B \left[ j^\ell_1, j^\ell_2 \right] B \left[ k^\ell_1, k^\ell_2 \right] \ldots B \left[ i^p, i^1 \right] \right] \leq K_2^{O(pq)} \frac{p^{5pq/8}}{q^{3pq/8}} (pq)^{pq/8 - s/2} \prod_{\ell \in [p]} c^\ell_1! \ldots c^\ell_s! \quad \text{(sub-gaussian entries in T)}$$

**Proof.** Consider any $c^1, \ldots, c^s \in [q/2]^p$ and $(i^1, \ldots, i^p) \in C(c^1, \ldots, c^s)$. We have

$$\mathbb{E} \left[ \prod_{\ell \in [p]} B \left[ j^\ell_1, j^\ell_2 \right] B \left[ k^\ell_1, k^\ell_2 \right] \ldots B \left[ i^p, i^1 \right] \right] = \frac{\prod_{\ell} c^\ell_1! \cdots c^\ell_s!}{((q/2)!)^{2p}} \sum_{j^1, k^1 \in O(i^1), \ldots, j^p, k^p \in O(i^p)} \mathbb{E} \left[ \prod_{\ell \in [p]} A \left[ j^\ell_1, k^\ell_1 \right] A \left[ j^\ell_2, k^\ell_2 \right] \ldots A \left[ j^\ell_p, k^\ell_p \right] \right]$$

$$= \frac{\prod_{\ell} c^\ell_1! \cdots c^\ell_s!}{((q/2)!)^{2p}} \sum_{\forall \ell, j^\ell, k^\ell \in O(i^\ell)} \mathbb{E} \left[ \prod_{\ell \in [p]} T \left[ j^\ell_1 j^\ell_2, k^\ell_1 k^\ell_2 \right] \ldots T \left[ j^\ell_q j^\ell_q, k^\ell_q k^\ell_q \right] \right]$$

$$= \frac{\prod_{\ell} c^\ell_1! \cdots c^\ell_s!}{((q/2)!)^{2p}} \left| \mathcal{F}(i^1, \ldots, i^p) \right| \quad \text{(rademacher)} \quad \text{(3.1)}$$

OR
Thus it remains to estimate the size of $\mathcal{S}(i^1, \ldots, i^p)$. We begin with some notation. For a tuple $t$ and a subsequence $t_1$ of $t$, let $t \setminus t_1$ denote the subsequence of elements in $t$ that are not in $t_1$. For a tuple of 2-sets of ordered pairs $t = \{(a_1, b_1), (c_1, d_1), \ldots, (a_m, b_m), (c_m, d_m)\}$, let $\text{atomize}(t)$ denote the tuple $(a_1, b_1, c_1, d_1, \ldots, a_m, b_m, c_m, d_m)$ (we assume $\forall i, (a_i, b_i) < (c_i, d_i)$ lexicographically). Observe that “atomize” is invertible.

For any $\bigoplus_{\ell=1} (j^\ell, k^\ell) \in \mathcal{S}(i^1, \ldots, i^p)$, observe that $\mathcal{S}\left(\bigoplus_{\ell=1} (j^\ell, k^\ell)\right)$ (which is of length $pq/4$) contains a subsequence $I$ of length $pq/8$, such that $\text{multiset}(I) = \text{multiset}\left(\mathcal{S}\left(\bigoplus_{\ell=1} (j^\ell, k^\ell)\right) \setminus I\right)$. Now we know

$$\text{multiset}\left(\text{atomize}\left(\mathcal{S}\left(\bigoplus_{\ell=1} (j^\ell, k^\ell)\right)\right)\right) = \bigcup_{\ell=1} \text{multiset}\left(j^\ell \oplus k^\ell\right) = \bigcup_{\ell=1} \text{multiset}\left(i^\ell \oplus i^\ell\right)$$

this implies $\text{multiset}(\text{atomize}(1)) = \text{multiset}\left(\text{atomize}\left(\mathcal{S}\left(\bigoplus_{\ell=1} (j^\ell, k^\ell)\right) \setminus 1\right)\right)$

$$= \bigcup_{\ell=1} \text{multiset}\left(i^\ell\right).$$

Thus, $\exists \pi \in \mathbb{S}_{pq/2}, s.t. \text{atomize}(1) = \pi(i^1 \oplus \cdots \oplus i^p)$ (3.3)

For tuples $t, t'$, let $\text{intrvl}(t, t')$ denote the set of all tuples obtained by interleaving the elements in $t$ and $t'$. By Eq. (3.3), we obtain that for any $\bigoplus_{\ell=1} (j^\ell, k^\ell) \in \mathcal{S}(i^1, \ldots, i^p)$,

$$\exists \pi \in \mathbb{S}_{pq/2}, \sigma \in \mathbb{S}_{pq/8}, s.t. \mathcal{S}\left(\bigoplus_{\ell=1} (j^\ell, k^\ell)\right) \in \text{intrvl}(1, \sigma(1))$$

where $I = \text{atomize}^{-1}(\pi(i^1 \oplus \cdots \oplus i^p))$ (3.4)

For any $j \in [s]$, let $c^j := \sum_{\ell \in [p]} c_{\ell}^j$. Now since $|\text{intrvl}(t, t')| \leq 2|t|+|t'|$, by Eq. (3.4) we have

$$|\mathcal{S}(i^1, \ldots, i^p)| \leq 2^{pq/4} |\mathbb{S}_{pq/8}| \times |i^1 \oplus \cdots \oplus i^p|$$

$$\leq 2^{pq/4} \frac{(pq/2)!}{c^1! \cdots c^p!}$$

$$\Rightarrow \mathbf{E}[B[i^1, i^2]B[i^2, i^3] \cdots B[i^p, i^1]] \leq 2^{pq/4} \frac{(pq/2)!}{c^1! \cdots c^p!} \prod_{\ell=1} c_{\ell}^{1,2} \cdots c_{\ell}^{p,1}$$

by Eq. (4.1)

$$\leq 2^{pq/4} \frac{(pq/2)!}{(pq/2)!} \prod_{\ell=1} c_{\ell}^{1} \cdots c_{\ell}^{p}$$

$$= \frac{2^{O(pq)}}{q^{3pq/8}} \prod_{\ell=1} c_{\ell}^{1} \cdots c_{\ell}^{p}$$

Lemma 3.4. For all $i^1, \ldots, i^p \in [n]^{q/2}$, we have

$$\mathbf{E}[B[i^1, i^2]B[i^2, i^3] \cdots B[i^p, i^1]] \geq 0$$
(2)  \[ \mathbb{E}[B^{i_1,i_2}B^{i_2,i_3}...B^{i_p,i_1}] \neq 0 \quad \Rightarrow \quad \#(i_1,...,i_p) \leq \frac{pq}{4} + \frac{q}{2} \]

**Proof.** The first claim follows immediately on noting that one is taking expectation of a polynomial of independent centered random variables with all coefficients positive.

For the second claim, note that \( \mathbb{E}[B^{i_1,i_2}B^{i_2,i_3}...B^{i_p,i_1}] \neq 0 \) implies that \( \mathcal{S}(i_1,...,i_p) \neq \emptyset \).

Therefore there exists \( \oplus_{\ell}(j^\ell,k^\ell) \) (where \( j^\ell,k^\ell \in \mathcal{O}(\hat{\ell}) \)) such that every element in \( \mathcal{S}(\oplus_{\ell}(j^\ell,k^\ell)) \) has even multiplicity. We now define a graph induced by \( \oplus_{\ell}(j^\ell,k^\ell) \) and denoted by \( \mathcal{P}_{\mathcal{G}}(\oplus_{\ell}(j^\ell,k^\ell)) \) as,

\[
\mathcal{P}_{\mathcal{G}}\left(\oplus_{\ell}(j^\ell,k^\ell)\right) := (V,E) \quad \text{where,}
\]

\[
V := \bigcup_{\ell \in [p]} \bigcup_{k \in [q/2]} \left\{ g^\ell_k \right\} = \bigcup_{\ell \in [p]} \bigcup_{k \in [q/2]} \left\{ t_k^\ell \right\}
\]

\[
E := \bigcup_{m \in [q/2]} \left\{ \{ j_m^1, k_m^1 \}, \{ j_m^2, k_m^3 \}, ..., \{ j_m^n, k_m^1 \} \right\}
\]

The even multiplicity condition on \( \mathcal{S}(\oplus_{\ell}(j^\ell,k^\ell)) \) implies that every element in \( E \) has even multiplicity and consequently \( |E| \leq pq/4 \). We next show that \( \mathcal{P}_{\mathcal{G}}(\oplus_{\ell}(j^\ell,k^\ell)) \) is the union of \( q/2 \) paths. To this end, we construct \( g^1 \in \mathcal{O}(i^1),...,g^q \in \mathcal{O}(i^q) \) as follows:

1. Let \( g^2 := k^2 \)
2. For \( 3 \leq \ell \leq p \) do:

   Since \( g^\ell \in \mathcal{O}(j^\ell) \), there exists \( \pi \in \mathbb{S}_{q/2} \) s.t. \( \pi(j^\ell) = g^\ell \)

   Let \( g^{(\ell+1)q-p} := \pi(k^{\ell+1}) \)

We observe that by construction,

\[
\bigcup_{m \in [q/2]} \left\{ \{ j_m^1, s_m^2 \}, \{ s_m^2, s_m^3 \}, ..., \{ s_m^n, s_m^1 \} \right\} = \bigcup_{m \in [q/2]} \left\{ \{ j_m^1, t_m^2 \}, \{ t_m^2, t_m^3 \}, ..., \{ t_m^n, t_m^1 \} \right\} = E
\]

which establishes that \( \mathcal{P}_{\mathcal{G}}(g^1,...,g^q) \) is a union of \( q/2 \) paths.

Now since \( \mathcal{P}_{\mathcal{G}}(\oplus_{\ell}(j^\ell,k^\ell)) \) is the union of \( q/2 \) paths, one need add at most \( q/2 - 1 \) edges to it and obtain a connected graph \( G' = (V,E') \) on the same vertex set. Since \( G' \) is connected, we have \( |V| \leq |E'| + 1 \leq pq/4 + q/2 \). But \( \#(i^1,...,i^p) = |V| \), which completes the proof.

**Lemma 3.5.** When \( T \) has rademacher (resp. sub-gaussian) entries, for any \( q \leq n \) (resp. \( q \leq n^{2/3}/\log n \)) we have

\[
\|B\|_2 \leq 2^{O(q)} \frac{n^{q/4} \log^{9q/8} n}{q^{q/8}} \quad \text{(rademacher)} \quad \|B\|_2 \leq K_2^{O(q)} \frac{n^{q/4} \log^{9q/8} n}{q^{q/8}} \quad \text{(sub-gaussian)} \quad \text{w.h.p.}
\]

**Proof.** We proceed by trace method. (Note that since \( T \) is symmetric, so are \( A \) and \( B \)). We assume \( T \) has rademacher entries. The case for sub-gaussian entries is identical.

\[
\mathbb{E}[\text{Tr}(B^p)] = \sum_{i_1,...,i_p \in [n]^{q/2}} \mathbb{E}[B^{i_1^1,i_2^2}B^{i_2^3,i_3^4}...B^{i_p^q,i_1^1}]
\]

\[
= \sum_{s \in [pq/4+q/2]} \sum_{\#(i^1,...,i^p) = s} \mathbb{E}[B^{i_1^1,i_2^2}B^{i_2^3,i_3^4}...B^{i_p^q,i_1^1}] \quad \text{(by Lemma 3.4)}
\]

\[
= \sum_{s \in [pq/4+q/2]} \sum_{c_1,...,c_t \in [q/2]^p} \sum_{(i^1,...,i^p) \in \mathcal{E}(c_1,...,c_t)} \mathbb{E}[B^{i_1^1,i_2^2}B^{i_2^3,i_3^4}...B^{i_p^q,i_1^1}]
\]

\[
= \sum_{s \in [pq/4+q/2]} \sum_{c_1,...,c_t \in [q/2]^p} \sum_{(i^1,...,i^p) \in \mathcal{E}(c_1,...,c_t)} \mathbb{E}[B^{i_1^1,i_2^2}B^{i_2^3,i_3^4}...B^{i_p^q,i_1^1}]
\]
Let $\mathcal{A} \in \mathbb{R}^{[n]^4}$ be a 4-tensor with independent rademacher (resp. centred sub-gaussian) entries, then for any even $q$ such that $q \leq n$ (resp. $q \leq n^{2/3}/\log n$), w.h.p. the degree-$q$ SoS program in Section 3.1 certifies that

$$\|\mathcal{A}\|_2 \lesssim \frac{n \log^{9/2} n}{\sqrt{q}} \quad \text{(resp. } \frac{K_2^{O(1)} n \log^{9/2} n}{\sqrt{q}} \text{)}$$

4. 2-Norm of a Random 3-Tensor

4.1. Degree-$q$ SoS Formulation for 3-Tensor 2-Norm

Let $\mathcal{A} \in \mathbb{R}^{[n]^3}$ be a 3-tensor with i.i.d. uniform ±1 entries (the extension to sub-gaussian variables is similar to the proof in Section 3).

The degree-$q$ SoS formulation is:

$$\max \mathbb{E} \left[ \langle \mathcal{A}, x^{\otimes 3} \rangle \right] \text{ s.t. } \mathbb{E} \left[ \|x\|_2^2 \right] = 1$$

4.2. Analysis Overview

Assume $q/4$ is a power of 2 as this only changes our claims by constants. For $\ell \in [n]$ let $T_\ell$ be an $n \times n$ matrix with i.i.d. uniform ±1 entries, such that we have

$$f(x) := \langle \mathcal{A}, x^{\otimes 3} \rangle = \sum_{\ell \in [n]} x_\ell (x^T T_\ell x) = \sum_{\ell \in [n]} x_\ell (x^T T_\ell x).$$

Let $T := (T_i + T_i^T)/2$. Following Hopkins et. al. [HSS15], let $T_i := \sum_{i=1}^n T_i \otimes T_i$. Let $E \in \mathbb{R}^{[n]^2 \times [n]^2}$ be the matrix such that $E[(i,j), (j,k)] = T_i[(i,j),(k,j)]$ for any $i,j \in [n]$ and $E[(i,j),(k,l)] = 0$ otherwise. Let $E' \in \mathbb{R}^{[n]^2 \times [n]^2}$ be the matrix such that $E'[i,j],[j,k)] = E[(i,j),(j,k)] + E[(j,k),(i,j)]$ for any $i,j \in [n]$ and $E'[i,j],[k,l)] = 0$ otherwise.

Let $T := T - E \in \mathbb{R}^{[n]^2 \times [n]^2}$ and $A := T^{-\otimes 4/4}$. Let $g(x) := (x^{\otimes 2})^T T x^{\otimes 2}$ and $h(x) := (x^{\otimes 2})^T E x^{\otimes 2} = (x^{\otimes 2})^T E' x^{\otimes 2}$. Let $\mathbb{E}$ be the pseudo-expectation operator returned by the program above.

We would like to show that there is some matrix representation $B$ of $A$, such that w.h.p. $\max_{\|y\|_1 = 1} y^T B y$ is small. To this end, consider the following mass shift procedure that we apply to $A$ to get $B$:

$$\forall i, j \in [n]^{q/2}, \quad B[i, j] = \frac{1}{(q/2)^{12}} \sum_{\pi, \sigma \in S_{q/2}} A[\pi(i), \sigma(j)] = \frac{1}{|O(i)| |O(j)|} \sum_{\ell \in O(i), f \in O(j)} A[\ell, f]$$
Below the fold we shall show that \( \|B\|_2^{4/q} = \tilde{O}(n^{3/2}/\sqrt{q}) \) w.h.p. This is sufficient to obtain the desired result since we have

\[
\|B\|_2 I - B \succeq 0
\]

\[
\Rightarrow \|B\|_2 \|x\|^q - \langle x^\otimes q/2, B x^\otimes q/2 \rangle \succeq 0
\]

\[
\Rightarrow \|B\|_2 \|x\|^q - \langle x^\otimes 2, T x^\otimes 2 \rangle q/4 \succeq 0
\]

\[
\Rightarrow \|B\|_2 \|x\|^q - (g(x) - h(x))q/4 \succeq 0
\]

\[
\Rightarrow \mathbb{E} \left[ (g(x) - h(x))q/4 \right] \leq \|B\|_2
\]

\[
\Rightarrow \mathbb{E} [g(x) - h(x)] \leq \|B\|_2^{4/q} \quad \text{(Pseudo-Cauchy-Schwarz)}
\]

\[
\Rightarrow \mathbb{E} [g(x)] \leq \|B\|_2^{4/q} + \mathbb{E} [h(x)]
\]

\[
\Rightarrow \mathbb{E} [g(x)] \leq \|B\|_2^{4/q} + 5n \quad \text{(5nI - E' \succeq 0)}
\]

\[
\Rightarrow \mathbb{E} [g(x)] = \tilde{O}(n^{3/2}/\sqrt{q}).
\]

Now \( \mathbb{E} [f(x)] = \mathbb{E} \left[ \sum_{\ell \in [n]} x_\ell (x^T T_\ell x) \right] \quad \text{(Following [HSS15])} \)

\[
\leq \mathbb{E} \left[ \|x\|^2 \right]^{1/2} \mathbb{E} \left[ \sum_{\ell \in [n]} (x^T T_\ell x)^2 \right]^{1/2} \quad \text{(Pseudo-Cauchy-Schwarz)}
\]

\[
\leq \mathbb{E} \left[ \sum_{\ell \in [n]} (x^T T_\ell x)^2 \right]^{1/2} \quad \text{(Pseudo-Cauchy-Schwarz)}
\]

\[
\leq \mathbb{E} \left[ \langle x^\otimes 2, T x^\otimes 2 \rangle \right]^{1/2}
\]

\[
= \mathbb{E} [g(x)]^{1/2} = \tilde{O}(n^{3/4}/q^{1/4})
\]

4.3. Analysis

For any \( i^1, \ldots, i^p \in [n]^{q/2} \) let \( e_k \) denote the k-th smallest element in \( \bigcup_{\ell,j} \{i^\ell_\ell \} \) and let

\[
\#(i^1, \ldots, i^p) := \left| \bigcup_{\ell \in [p]} \bigcup_{j \in [n]} \{i^j_\ell \} \right|.
\]

For any \( c^1, \ldots, c^p \in [q/2]^{p}, \) let

\[
\mathcal{C}(c^1 \ldots c^p) := \left\{ (i^1, \ldots, i^p) \mid \#(i^1, \ldots, i^p) = s, \forall k \in [s], \ell \in [p], e_k \text{ appears } c^\ell_k \text{ times in } i^\ell \right\}
\]

Observation 4.1.

\[
\left| \left\{ (c^1, \ldots, c^p) \in ([q/2]^{p})^s \mid \mathcal{C}(c^1, \ldots, c^p) \neq \phi \right\} \right| \leq 2^{O(pq)} p^{pq/2}
\]

if \( \mathcal{C}(c^1, \ldots, c^p) \neq \phi \) then \( \mathcal{C}(c^1, \ldots, c^p) \)\]
Lemma 4.2. Consider any $c^1, \ldots, c^t \in [q/2]^p$ and $(i^1, \ldots, i^p) \in \mathcal{C}(c^1, \ldots, c^t)$. We have

\[
E[B[i^1, i^2]B[i^2, i^3] \ldots B[i^p, i^1]] \leq 2^{O(pq)} \frac{p^{5pq/8}}{q^{3pq/8}} \prod_{\ell \in [p]} c^\ell_1 \cdots c^\ell_t!
\]

Proof. Consider any $c^1, \ldots, c^t \in [q/2]^p$ and $(i^1, \ldots, i^p) \in \mathcal{C}(c^1, \ldots, c^t)$. We have

\[
E[B[i^1, i^2]B[i^2, i^3] \ldots B[i^p, i^1]] = \prod_{\ell \in [p]} c^\ell_1 \cdots c^\ell_t^2 \sum_{\forall \ell, j^\ell, k^\ell \in \mathcal{O}(i^\ell)} \mathbb{E} \left[ \prod_{\ell \in [p]} T[j^\ell_1, j^\ell_2, \cdots, j^\ell_4] \right] \right]
\]

Thus it remains to estimate the size of $\mathcal{S}(i^1, \ldots, i^p)$. We begin with some notation. For a tuple $t$ and a subsequence $t_1$ of $t$, let $t \setminus t_1$ denote the subsequence of elements in $t$ that are not in $t_1$. For a tuple of 2-sets $t = \{(a_1, b_1), \ldots, (a_m, b_m)\}$, let atomize$(t)$ denote the tuple $(a_1, b_1, \ldots, a_m, b_m)$ (we assume $\forall i, a_i < b_i$). Observe that “atomize” is invertible.

For any $(\bigoplus_{\ell \in [p]} (j^\ell, k^\ell), (S_1 \ldots S_n)) \in \mathcal{S}(i^1, \ldots, i^p)$, observe that $\mathcal{S}_r(\bigoplus_{\ell \in [p]} (j^\ell, k^\ell))$ (which is of length $2|S_r|$) contains a subsequence $I_{S_r}$ of length $|S_r|$, such that $\text{multiset}(I_{S_r}) = \text{multiset}(\mathcal{S}_r(\bigoplus_{\ell \in [p]} (j^\ell, k^\ell)) \setminus I_{S_r})$. Now we know

\[
\text{multiset} \left( \bigoplus_{\ell \in [p]} \text{atomize} \left( \mathcal{S}_r(\bigoplus_{\ell \in [p]} (j^\ell, k^\ell)) \right) \right) = \bigsqcup_{t \in [p]} \text{multiset} \left( j^\ell \oplus k^\ell \right) = \bigsqcup_{t \in [p]} \text{multiset} \left( i^\ell \oplus i^\ell \right)
\]

Thus, $\exists \pi \in \mathbb{S}_{pq/2}, \exists \pi, \text{atomize}(I_{S_r}) = \pi(i^1 \oplus \cdots \oplus i^p)$ (4.2)
For tuples \( t, t' \), let \( \text{intrlv}(t, t') \) denote the set of all tuples obtained by interleaving the elements in \( t \) and \( t' \). By Eq. (4.2), we obtain that for any \((\oplus_{\ell}(j^{\ell}, k^{\ell}), (S_1 \ldots S_n)) \in \mathcal{S}(i^1, \ldots, i^p)\),

\[
\exists \pi \in S_{pq/2}, \text{s.t. } \forall r \in [n], \exists \sigma_r \in S_{|S_r|}, \text{s.t. } \mathcal{S}_\pi (\oplus_{\ell}(j^{\ell}, k^{\ell})) \in \text{intrlv}(I_{S_r}, \sigma_r(I_{S_r})) \quad (4.3)
\]

where \( I_{S_r} = \text{atomize}^{-1}(\pi(i^1 \oplus \cdots \oplus i^p)) \)

For any \( j \in [s] \), let \( \bar{c}_j = \sum_{\ell \in [\rho]} c_{\ell j}^j \). Now since \( |\text{intrlv}(t, t')| \leq 2^{r + |t'|} \), by Eq. (4.3) we have that for any \( \bigcup_{u \in [n]} S_u = [p] \times [q/4] \),

\[
\#(S_1 \ldots S_n) := \left\{ \left( \bigoplus_{\ell}(j^{\ell}, k^{\ell}), (S_1 \ldots S_n) \right) \in \mathcal{S}(i^1, \ldots, i^p) \right\} \\
\leq 2^{pq/4} |S_{|S_1|}| \times \cdots \times |S_{|S_n|}| \times |\bigoplus(i^1 \oplus \cdots \oplus i^p)| \\
\leq 2^{pq/4} |S_1|! \cdots |S_n|! \left( \frac{(pq/2)!}{\bar{c}_1! \cdots \bar{c}_s!} \right) \quad (4.4)
\]

For any \((\oplus_{\ell}(j^{\ell}, k^{\ell}), (S_1 \ldots S_n)) \in \mathcal{S}(i^1, \ldots, i^p)\), observe that the even multiplicity condition combined with the condition that \((j_{2^k-1}^{\ell}, k_{2^k-1}^{\ell+1}) \neq (j_{2^k}^{\ell}, k_{2^k}^{\ell+1})\), imply that for each \( r \in [n], |S_r| \neq 1 \). Thus every non-empty \( S_r \) has size at least 2, implying that the number of non-empty sets in \( S_1, \ldots, S_n \) is at most \( pq/8 \). Thus we have,

\[
|\mathcal{S}(i^1, \ldots, i^p)| = \sum_{U \subseteq [n], |U| \leq pq/8} \sum_{u \in U} \#(S_1, \ldots, S_n) \\
= \sum_{U \subseteq [n], |U| \leq pq/8} \sum_{u \in U} \sum_{u \in [p] \times [q/4]} \#(S_1, \ldots, S_n) \\
\leq 2^{pq/4} |S_1|! \cdots |S_n|! \left( \frac{(pq/2)!}{\bar{c}_1! \cdots \bar{c}_s!} \right) \quad (4.4)
\]

\[
\sum_{u \in [p/4]} 2^{O(pq)} \frac{(pq/2)!}{\bar{c}_1! \cdots \bar{c}_s!} \left( \frac{pq}{4} \right) \\
\leq \sum_{u \in [p/8]} 2^{O(pq+|U|)} \frac{(pq/2)!}{\bar{c}_1! \cdots \bar{c}_s!} \left( \frac{pq}{4} \right) \\
\leq \sum_{u \in [p/8]} 2^{O(pq)} \left( \frac{n}{u} \right) \frac{(pq/2)!}{\bar{c}_1! \cdots \bar{c}_s!} \left( \frac{pq}{4} \right) \\
\leq \sum_{u \in [p/8]} 2^{O(pq)} \left( \frac{pq}{4} \right) \frac{(pq/2)!}{\bar{c}_1! \cdots \bar{c}_s!} \left( \frac{pq}{4} \right) \\
\leq \sum_{u \in [p/8]} 2^{O(pq)} \frac{(pq)^{pq/8}}{\bar{u}^2} \left( \frac{pq/2}{\bar{c}_1! \cdots \bar{c}_s!} \right) \left( \frac{pq}{4} \right) \\
\leq \sum_{u \in [p/8]} 2^{O(pq)} \frac{(pq)^{pq/8}}{\bar{u}^2} \left( \frac{pq/2}{\bar{c}_1! \cdots \bar{c}_s!} \right) \left( \frac{pq}{4} \right) \quad (\text{since } pq < n)
\]
\begin{proof}
\textbf{Proof of Lemma 4.4.}

Therefore there exists
\begin{equation}
\sum_{\ell \in [p]} \frac{c_\ell! \ldots c_\ell!}{(q/2)!^2p^q} = \sum_{\ell \in [p]} \frac{c_\ell! \ldots c_\ell!}{(q/2)!^2p^q} = \sum_{\ell \in [p]} \frac{c_\ell! \ldots c_\ell!}{(q/2)!^2p^q} \leq 2^{O(pq)} n^{p/8} \frac{p^5q^8}{q^{3pq/8}} \prod_{\ell \in [p]} c_\ell! \ldots c_\ell!
\end{equation}
\end{proof}

\textbf{Lemma 4.3.} For all \(i^1, \ldots, i^p \in [n]^{q/2}\), we have

\begin{enumerate}
\item \(E \left[ B \left[ i^1, i^2 \right] B \left[ i^2, i^3 \right] \ldots B \left[ i^p, i^1 \right] \right] \geq 0 \)
\item \(E \left[ B \left[ i^1, i^2 \right] B \left[ i^2, i^3 \right] \ldots B \left[ i^p, i^1 \right] \right] \neq 0 \Rightarrow \#(i^1, \ldots, i^p) \leq \frac{pq}{4} + \frac{q}{2}
\end{enumerate}

\textbf{Proof.} The first claim follows immediately on noting that one is taking expectation of a polynomial of independent centered random variables with all coefficients positive.

For the second claim, note that \(E \left[ B \left[ i^1, i^2 \right] B \left[ i^2, i^3 \right] \ldots B \left[ i^p, i^1 \right] \right] \neq 0\) implies that \(\mathcal{S} \left( i^1, \ldots, i^p \right) \neq \phi\). Therefore there exists \(\oplus_{\ell} (j^\ell, k^\ell)\) (where \(j^\ell, k^\ell \in \mathcal{O}(j^\ell)\)) and \(\cup_{\ell \in [n]} S_u = [p] \times [q/4]\) such that every element in \(\oplus_{\ell \in [n]} \mathcal{S}, \oplus_{\ell} (j^\ell, k^\ell)\) has even multiplicity. The rest of the proof follows from the same ideas as in the proof of Lemma 3.4.

\textbf{Lemma 4.4.}

\[ \|B\|_2^{3/4} \leq \frac{n^{3/2} \log^5 n}{\sqrt{q}} \text{ w.h.p.} \]

\textbf{Proof.} We proceed by trace method. (Note that since \(T\) is symmetric, so are \(A\) and \(B\)).

\[ E[Tr(B^p)] = \sum_{i^1, \ldots, i^p \in [n]^{q/2}} E \left[ B \left[ i^1, i^2 \right] B \left[ i^2, i^3 \right] \ldots B \left[ i^p, i^1 \right] \right] \]
\[ = \sum_{s \in [pq/4+q/2]} \sum \sum_{(i^1, \ldots, i^p) \in S} E \left[ B \left[ i^1, i^2 \right] B \left[ i^2, i^3 \right] \ldots B \left[ i^p, i^1 \right] \right] \text{ by Lemma 4.3} \]
\[ = \sum_{s \in [pq/4+q/2]} \sum \sum_{(i^1, \ldots, i^p) \in S} E \left[ B \left[ i^1, i^2 \right] B \left[ i^2, i^3 \right] \ldots B \left[ i^p, i^1 \right] \right] \]
\[ = \sum_{s \in [pq/4+q/2]} \sum \sum_{(i^1, \ldots, i^p) \in S} E \left[ B \left[ i^1, i^2 \right] B \left[ i^2, i^3 \right] \ldots B \left[ i^p, i^1 \right] \right] \]
\[ \leq \sum_{s \in [pq/4+q/2]} \sum \sum_{(i^1, \ldots, i^p) \in S} 2^{O(pq)} n^{p/8} \frac{p^5q^8}{q^{3pq/8}} \prod_{s \in [p]} c_\ell! \ldots c_\ell! \text{ by Lemma 4.2} \]
\[ \leq \sum_{s \in [pq/4+q/2]} 2^{O(pq)} n^{p/8} \frac{p^5q^8}{q^{3pq/8}} \text{ by Observation 4.1} \]

18
Let \( A \) be a graph and \( \mathcal{A} \in (\mathbb{R}^r)^{r \times r} \) be the \( r \)-parity tensor of a graph with a clique of size \( P \), then \( \| \mathcal{A} \|_2 \geq P^{r/2} \). Thus if one can quickly compute \( \| \mathcal{A} \|_2 \), one can quickly detect cliques of size as low as \( \tilde{O}(n^{1/r}) \) ([BV09]) also shows how to recover the clique given the maximizer \( x \) of \( \| \mathcal{A} \|_2 \). More generally, if one can certify in time \( T \) (w.h.p.) an upper bound of \( U \) on the two-norm of the \( r \)-parity-tensor of a graph drawn from \( G_{n,1/2} \), then one can detect planted cliques of size \( \sim U^{2/r} \) in time \( T + \tilde{O}(1) \).

The above result hints at a natural hierarchy of SoS programs to detect (or recover) planted cliques, namely, given a graph \( G \), run the degree-\( q \) SoS two-norm relaxation on the \( q \)-parity-tensor of \( G \). The relaxation is:

### 5. Parity Tensor

#### 5.1. Preliminaries

For a graph \( G = (V,E) \) its \( r \)-parity tensor \( \mathcal{A} \) is defined as follows:

\[
\mathcal{A}[i_1, \ldots, i_r] = \begin{cases} 
-1 & \text{if } i_1, \ldots, i_r \text{ are distinct} \\
0 & \text{otherwise}
\end{cases}
\]

Brubaker and Vempala [BV09] established the following:

**Theorem 5.1.** There is a constant \( C_r \) such that with probability at least \( 1 - 1/n \) the norm of the \( r \)-parity tensor \( \mathcal{A} : [n]^r \rightarrow \{-1,1\} \) for the random graph \( G_{n,1/2} \) is bounded by

\[
\| \mathcal{A} \|_2 \leq C_r r^{(r-1)/2} \sqrt{n} \log^{(3r-1)/2} n.
\]

It is not hard to verify that if \( \mathcal{A} \) is the \( r \)-parity-tensor of a graph with a clique of size \( P \), then \( \| \mathcal{A} \|_2 \geq P^{r/2} \).

### 5.1.1. Degree-\( q \) SoS Formulation for \( q \)-Parity-Tensor 2-Norm

Let \( q \) be even, \( G \) be a graph and \( \mathcal{A} \in (\mathbb{R}^r)^{r \times r} \) be the \( q \)-parity-tensor of \( G \). The degree-\( q \) SoS formulation is:

\[
\max \quad \mathbb{E} \left[ \langle \mathcal{A}, x^{\otimes q} \rangle \right] \quad \text{s.t.} \quad \mathbb{E} \left[ |x|_2^2 \right] = 1
\]

In this section, we will show that the above program cannot detect cliques of size \( \sim \sqrt{n}/\log n \) which is evidence that the parity tensor formulation of planted clique is likely worse than the standard formulation that is used in [FK00], since that formulation allows one to find cliques of size \( \sqrt{n}/2^{q-1} \) in the \( q \)-th level of the LS+ hierarchy.
5.2. Lower Bound

We will proceed almost identically to Section 2.3 except that we need to bound the spectral norm of the matricised parity-tensor.

5.2.1. Candidate Moment Matrix

Given a graph \( G \) drawn from \( G_{n,1/2} \), and its \( q \)-parity-tensor \( \mathcal{A} \), let \( A := \text{Mat}(\mathcal{A}) \), and define \( B \in \mathbb{R}^\left( \binom{n}{q/2} \times \binom{n}{q/2} \right) \) and \( M \in \mathbb{R}^{[n]q/2 \times [n]q/2} \) as,

\[
B := A_0 + 2C_q (n \log^2 n)^{q/4} q^{q/2} \mathcal{W} \quad \text{(C to be determined shortly)}
\]

\( B \) is defined s. t. \( B_0 = B \)

\[
M := \frac{\text{Tr}(B)}{n^{\log q/2}}
\]

\( M \) is the candidate moment matrix that we will extend to a pseudo-expectation operator.

5.2.2. PSDness

Lemma 5.2. With high probability,

\[
\min_{\|x\|=1} x^T A_0 x \geq -2^O(q) (n \log^2 n)^{q/4} q^{q/2}.
\]

Proof. Let \( \mathcal{D} := A_0/ (q/2)! \). It is easy to see that \( \mathcal{D} \) has all \( 0, +1, -1 \) entries. By Lemma 5.3, it is easy to see that the number of sequences of multisets \( i^1, \ldots, i^p \in \binom{[n]}{q/2} \), such that \( \mathbb{E} [\mathcal{D}[i^1, i^2] \cdots \mathcal{D}[i^p, i^1]] \neq 0 \) is at most \( 2^{O(qp)} p^{pq/2} n^{pq/4+q/2} \). Thus \( \mathbb{E} [\text{Tr}(\mathcal{D}^p)] \leq 2^{O(qp)} p^{pq/2} n^{pq/4+q/2} \). Setting \( p = \Theta(\log n) \) and applying markov’s inequality completes the proof. \( \blacksquare \)

Lemma 5.3. For all \( i^1, \ldots, i^p \in \binom{[n]}{q/2} \), we have

1. \( \mathbb{E} [\mathcal{D}[i^1, i^2] \mathcal{D}[i^2, i^3] \cdots \mathcal{D}[i^p, i^1]] \geq 0 \)
2. \( \mathbb{E} [\mathcal{D}[i^1, i^2] \mathcal{D}[i^2, i^3] \cdots \mathcal{D}[i^p, i^1]] \neq 0 \Rightarrow \#(i^1, \ldots, i^p) \leq \frac{pq}{4} + \frac{q}{2} \)

Proof. The first claim follows immediately on noting that one is taking expectation of a polynomial of independent centered random variables with all coefficients positive.

For the second claim we will need some definitions. Let \( a \) (resp. \( b \)) be the number of elements in \([n]\) that appear exactly once (resp. twice) in \( i^1 \sqcup \cdots \sqcup i^p \). Let \( c \) be the number of elements in \([n]\) that appear at least thrice in \( i^1 \sqcup \cdots \sqcup i^p \). Let \( s := \#(i^1, \ldots, i^p) \). Now if \( a \leq q \), then \( s \leq pq/4 + q/2 \) since all but \( a \) elements in \([n]\) appear at least twice in \( i^1 \sqcup \cdots \sqcup i^p \) which means there are at most \((pq/2 - q)/2 + q\) distinct elements across the multisets. We will show below the fold that if \( a > q \), then \( c \geq a \). This would complete the proof since we have, \( a + b + c = s \) and \( a + 2b + 3c \leq pq/2 \), which implies that \( 2s + c - a \leq pq/2 \) which in turn implies that \( s \leq pq/4 \) (assuming \( c \geq a \)).

Thus it remains to show that if \( a > q \), then \( c \geq a \). For any unordered pair \((x, y) \in [n] \) with \( x \neq y \), say it is **covered** by the \( \ell \)th layer if \( x \in i^\ell \) or \( y \in i^\ell \) or \( x \in i^{\ell+1} \) (note that \( \ell + 1 \) is modulo \( p \)). It is easy to see that if \( \mathbb{E} [\mathcal{D}[i^1, i^2] \mathcal{D}[i^2, i^3] \cdots \mathcal{D}[i^p, i^1]] \neq 0 \), then for any pair \((x, y) \in [n] \) with \( x \neq y \), the number of \( \ell \in [p] \) that covers \((x, y) \) is even.
Call $x \in [n]$ unique if $x$ occurs exactly once in $\bigsqcup_{\ell=1}^{\ell}i^{\ell}$. Suppose $i^{\ell}$ contains a unique element $x$ for some $\ell$. For any $y \in i^{\ell-1}$, the pair $(x, y)$ is covered by the $(\ell - 1)$th layer, and since $x$ is unique, it needs to be covered by the $\ell$th layer to be covered an even number of times. Therefore $y \in i^{\ell+1}$, and we can conclude that $i^{\ell-1} = i^{\ell+1}$. If there exists an element $y \in i^{\ell-1}$ that occurs exactly twice in $i^{\ell-1} \sqcup \cdots \sqcup i^{\ell}$, for any $z \in i^{\ell-2}$, the same argument shows that $z \in i^{\ell+2}$ (it may be the case $z \in i^{\ell}$, but in that case $(y, z)$ is covered by both $\ell$th and $(\ell + 1)$th layer, so the parity does not change), therefore $i^{\ell-2} = i^{\ell+2}$. Proceeding in this manner, assuming $i^{\ell-(k-1)} = i^{\ell+(k-1)}$, $i^{\ell-k} = i^{\ell+k}$ and there exists $y \in i^{\ell-k}$ that occurs exactly twice in $i^{\ell-1} \sqcup \cdots \sqcup i^{\ell}$, we can conclude that $i^{\ell-(k+1)} = i^{\ell+(k+1)}$. When there are two multisets other than $i^{\ell}$ that contains a unique element, this inductive process must stop at some point and there exists $k < p/2$ such that

- $i^{\ell-k} = i^{\ell+k}$.
- Each element in $x \in i^{\ell-k}$ occurs at least three times in $i^{1} \sqcup \cdots \sqcup i^{\ell}$.
- Among $i^{\ell-k}, i^{\ell-(k-1)}, \ldots, i^{\ell}, \ldots, i^{\ell+(k-1)}, i^{\ell+k}, i^{\ell}$, $i^{\ell}$ is the only multiset that contains a unique element.

Let $\ell_1 < \cdots < \ell_r$ be such that $i^{\ell_k}$ contains a unique element. $a > q$ implies $r \geq 3$. Also $a \leq rq/2$. Applying the above argument for every $\ell_k$, we conclude that there exists $\ell_{k'}$ such that $\ell_k < \ell_{k'} < \ell_{k+1}$ and every $x \in \ell'$ appears at least three times in $i^{1} \sqcup \cdots \sqcup i^{\ell'}$. It implies that $c \geq rq/2 \geq a$. \hfill $\blacksquare$

**Observation 5.4.** For an appropriately chosen constant $C$, by Lemma 5.2 and properties of $\mathcal{W}$, $\mathcal{B}$ is PSD w.h.p.

**Lemma 5.5.** $\mathcal{M}$ is PSD with high probability.

**Proof.** It suffices to show that $\mathcal{B}$ is PSD w.h.p. since $\text{Tr}(\mathcal{B})$ is non-negative. Now by combining Observation 5.4 and Lemma A.7, we see that $\mathcal{B}$ is PSD w.h.p. \hfill $\blacksquare$

**5.2.3. Objective Function’s Value on $\mathcal{M}$**

**Observation 5.6.** $\text{Tr}(\mathcal{B}) = 2^{O(q)} n^{3q/4} \log^{q/2} n$ with high probability.

**Proof.** Since each entry of $\mathcal{W}$ is bounded by $2^q$, $\text{Tr}(\mathcal{B}) = \text{Tr}(\mathcal{A}_0) + 2^{O(q)} n^{3q/4} \log^{q/2} n$. We then obtain that $\text{Tr}(\mathcal{B}) = 2^{O(q)} n^{3q/4} \log^{q/2} n$ since Lastly, the claim follows by substituting $\text{Tr}(\mathcal{A}_0) = 0$ and applying Observation A.4. \hfill $\blacksquare$

**Theorem 5.7.** For any even $q \leq n$, let $\mathcal{A} \in \mathbb{R}^{[n] \times [n]}$ be the $q$-parity-tensor of a graph $G$ drawn from $G_{n, 1/2}$. There exists a constant $C$ depending only on $K_2$, such that the SoS value of the formulation from Section 5.1.1 is at least $C 2^{-O(q)} n^{q/4} / \log^{q/2} n$ with high probability.

**Proof.** $\mathcal{M}$ is PSD by Lemma 5.5 and $\text{Tr}(\mathcal{M}) = 1$ by construction. Lastly we have,

\[
\langle A, M \rangle = \frac{\langle A, B \rangle}{\text{Tr}(B)} = \frac{\langle A_0, B_0 \rangle}{\text{Tr}(B)}
\]

(by Observation A.5)

\[
= 2^{-O(q)} 2^{q} \langle \mathcal{D}, \mathcal{D} \rangle - O(n^{3q/4} q^q)
\]

(dist. of entries of $A_0$)

\[
= 2^{-O(q)} n^q \text{ w.h.p.}
\]

\[
= 2^{-O(q)} n^{q/4} \log^{q/2} n \text{ w.h.p.} \quad \text{(by Observation 5.6)}
\]
Now we construct the pseudo-expectation operator $\tilde{E}$. For any $S \in \binom{[n]}{q}$, let $\tilde{E}[x^S] = M[i, j]$ for any $i, j \in [n]^{q/2}$ such that $S = \text{multiset}(i) \sqcup \text{multiset}(j)$. It is well-defined since $M$ is degree-$q$ SoS symmetric. Finally, $\tilde{E}[x^S] = 0$ for all $S \in \binom{[n]}{q/2}$. $\text{Tr}(M) = 1$ implies $\tilde{E}[[x_i]^2] = 1$, and $\tilde{E}[\langle A^O, x^{\text{SoS}} \rangle] = \langle A, M \rangle = 2^{-\Theta(q^n)} n^{q/4} / \log^{q/2} n$ w.h.p. This completes the lower bound.

\section{A. Quotient Matrix}

\begin{definition}
We say that a matrix $A \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$ is (q-SoS) block-symmetric if for every $i, j \in [n]^{q/2}$ and $i' \in \mathcal{O}(i), j' \in \mathcal{O}(j)$, it holds that $A[i, j] = A[i', j']$.
\end{definition}

We next make an easy observation regarding the structure of a block-symmetric matrix.

\begin{observation}
For any block-symmetric $A \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$, and any $i, j \in [n]^{q/2}$,
\[ A|_{\mathcal{O}(i) \times \mathcal{O}(j)} = A[i, j] \cdot J_{|\mathcal{O}(i)| \times |\mathcal{O}(j)|} \]
\end{observation}

\begin{definition}
For a block-symmetric matrix $A \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$, we define a multiset indexed matrix called the (q-SoS) quotient matrix of $A$ and denoted by $A_Q \in \mathbb{R}^{\binom{n}{q/2} \times \binom{n}{q/2}}$ as, $A_Q[i, j] := A[i', j'] / |\mathcal{O}(i)| \cdot |\mathcal{O}(j)|$ for all $i, j \in \binom{[n]}{q/2}$, where $i'$ and $j'$ are any tuples in $\mathcal{O}(i)$ and $\mathcal{O}(j)$ respectively.
\end{definition}

We will next see some useful properties of the quotient matrix. We shall start with showing that quotient matrices preserve trace and inner products.

\begin{observation}
For any block-symmetric $A \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$, $\text{Tr}(A) = \text{Tr}(A_Q)$.
\end{observation}

\textit{Proof.}
\[ \text{Tr}(A) = \sum_{i \in [n]^{q/2}} A[i, i] = \sum_{i \in \binom{[n]}{q/2}} |\mathcal{O}(i)| \cdot A[\text{tuple}(i), \text{tuple}(i)] = \sum_{i \in \binom{[n]}{q/2}} A_Q[i, i] = \text{Tr}(A_Q). \]

\begin{observation}
For any block-symmetric $A, B \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$, $\langle A, B \rangle = \langle A_Q, B_Q \rangle$.
\end{observation}

\textit{Proof.}
We have,
\[ \langle A, B \rangle = \sum_{i,j \in [n]^{q/2}} A[i, j] \cdot B[i, j] = \sum_{i,j \in \binom{[n]}{q/2}} |\mathcal{O}(i)| \cdot |\mathcal{O}(j)| \cdot A[\text{tuple}(i), \text{tuple}(j)] \cdot B[\text{tuple}(i), \text{tuple}(j)] = \sum_{i,j \in \binom{[n]}{q/2}} A_Q[i, j] \cdot B_Q[i, j] = \langle A_Q, B_Q \rangle. \]

We next establish that an upper bound on the spectrum of the quotient matrix is an upper bound on the spectrum of the block-symmetric counterpart.

\begin{lemma}
For any block-symmetric $A \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$, $\|A\|_2 \leq \|A_Q\|_2$.
\end{lemma}
Proof. For any \( u, v \in \mathbb{R}^{[n]/2} \) s.t. \( \|u\| = \|v\| = 1 \), we have

\[
u^T A v = \sum_{i, j \in [n]/2} A[i, j] u_i v_j
= \sum_{i, j \in [n]/2} \frac{A_0[\text{multiset}(i), \text{multiset}(j)]}{\sqrt{|O(i)||O(j)|}} u_i v_j
= \sum_{i, j \in \left(\frac{n}{q/2}\right)} \frac{A_0[i, j]}{\sqrt{|O(i)||O(j)|}} \frac{\langle u_{|O(i)|}, 1 \rangle}{\sqrt{|O(i)|}} \langle v_{|O(j)|}, 1 \rangle
\]

(by Observation A.2)

\[
= a^T A_0 b \quad \text{where } a_i := \frac{\langle u_{|O(i)|}, 1 \rangle}{\sqrt{|O(i)|}}, \quad b_i := \frac{\langle v_{|O(i)|}, 1 \rangle}{\sqrt{|O(i)|}}
\]

\[
\leq \|A_0\|_2 \|a\| \cdot \|b\|
= \|A_0\|_2 \sqrt{\sum_{i \in \left(\frac{n}{q/2}\right)} \langle u_{|O(i)|}, 1 \rangle^2 / |O(i)|} \sqrt{\sum_{i \in \left(\frac{n}{q/2}\right)} \langle v_{|O(i)|}, 1 \rangle^2 / |O(i)|}
\leq \|A_0\|_2 \sqrt{\sum_{i \in \left(\frac{n}{q/2}\right)} \|u_{|O(i)|}\|^2} \sqrt{\sum_{i \in \left(\frac{n}{q/2}\right)} \|v_{|O(i)|}\|^2}
\]

(by Cauchy-Schwarz)

\[
\leq \|A_0\|_2 \|u\| \cdot \|v\| = \|A_0\|_2.
\]

Lastly we establish that a block-symmetric matrix is PSD iff its quotient matrix is PSD.

Lemma A.7. Any block-symmetric \( A \in \mathbb{R}^{[n]/2 \times [n]/2} \) is PSD iff \( A_0 \) is PSD.

Proof. This follows since we have,

\[
x^T A x = \sum_{i, j \in [n]/2} A[i, j] x_i x_j
= \sum_{i, j \in [n]/2} \frac{A_0[\text{multiset}(i), \text{multiset}(j)]}{\sqrt{|O(i)||O(j)|}} x_i x_j
= \sum_{i, j \in \left(\frac{n}{q/2}\right)} \frac{A_0[i, j]}{\sqrt{|O(i)||O(j)|}} \langle x_{|O(i)|}, 1 \rangle \langle x_{|O(j)|}, 1 \rangle
\]

(by Observation A.2)

\[
= b^T A_0 b \quad \text{where } b_i := \frac{\langle x_{|O(i)|}, 1 \rangle}{\sqrt{|O(i)|}} \text{ for } i \in \left(\frac{n}{q/2}\right).
\]

\[
\]

B. Sub-gaussian Distributions

Following the exposition of Vershynin [Ver10],

Definition B.1. Let \( X \) be a random variable. Then the following properties are equivalent. A random variable \( X \) is called a sub-gaussian random variable if it satisfies one of the equivalent properties (with parameters \( K_i > 0 \) differing from each other by at most an absolute constant factor):

1. Tails: \( \Pr[\{|X| > t\}] \leq \exp(1 - t^2/K_1^2) \) for all \( t \geq 0\);
2. Moments: $\mathbb{E} [ |X|^p ]^{1/p} \leq K_2 \sqrt{p}$ for all $p \geq 1$;

3. Super-exponential moment: $\mathbb{E} \left[ \exp(\frac{X^2}{K_2^2}) \right] \leq e$.

Moreover, if $\mathbb{E} X = 0$ then properties 1–3 are also equivalent to the following one:

4. Moment generating function: $\mathbb{E} [ e^{tX} ] \leq e^{t^2 K_2^2}$ for all $t \in \mathbb{R}$.

Definition B.2 (Sub-gaussian norm). The sub-gaussian norm of a sub-gaussian random variable $X$, denoted $\|X\|_{\psi_2}$, is defined to be the smallest $K_2$ in property 2. In other words,

$$\|X\|_{\psi_2} = \sup_{p \geq 1} p^{-1/2} (\mathbb{E} |X|^p)^{1/p}.$$ 

Rotation invariance holds for sub-gaussian random variables, although approximately:

Lemma B.3 (Rotation invariance). Consider a finite number of independent centered sub-gaussian random variables $X_i$. Then $\sum_i X_i$ is also a centered sub-gaussian random variable. Moreover,

$$\|\sum_i X_i\|_{\psi_2}^2 \leq C \sum_i \|X_i\|_{\psi_2}^2$$

where $C$ is an absolute constant.

Rotation invariance immediately yields a large deviation inequality for sums of independent sub-gaussian random variables:

Lemma B.4 (Hoeffding-type inequality). Let $X_1, \ldots, X_N$ be independent centered sub-gaussian random variables, and let $K = \max_i \|X_i\|_{\psi_2}$. Then for every $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$ and every $t \geq 0$, we have

$$\Pr \left\{ \left| \sum_{i=1}^N a_i X_i \right| \geq t \right\} \leq e \cdot \exp \left( -\frac{ct^2}{K^2 \|a\|_2^2} \right)$$

where $c > 0$ is an absolute constant.

C. Proofs for Lower Bound

Claim C.1 (Restatement of Claim 2.4). $R[i, j]$ is the number of ways to place $j$ parenses ‘(’ or ‘)’ consistently so that there are $i$ more ‘(’ than ‘)’.

Proof. We proceed by the induction on $j$. When $j = 0$, $R[0, 0] = 1$ and $R[i, 0] = 0$ for all $i \geq 1$. Assume the claim holds up to $j - 1$. By the definition $R_j = TR_{j-1}$.

- For $i = 0$, the last parenthesis must be the close parenthesis, so the definition $R[0, j] = R[1, j-1]$ still measures the number of ways to place $j$ parenses with equal number of ‘(’ and ‘)’.

- For $i = k$, the last parenthesis must be the open parenthesis, so the definition $R[k, j] = R[k-1, j-1]$ still measures the number of ways to place $j$ parenses with $k$ more ‘(’.

- For $0 < i < k$, the definition of $R$ gives $R[i, j] = R[i-1, j-1] + R[i+1, j-1]$. Since $R[i-1, j]$ corresponds to placing ‘)’ in the $j$th position and $R[i+1, j]$ corresponds to placing ‘(’ in the $j$th position, $R[i, j]$ still measures the desired quantity.

This completes the induction and proves the claim.
Lemma C.2 (Restatement of Lemma 2.7). There exist universal constants $C, c > 0$ such that with probability $1 - e^{-ct^2}$, one has $\|A_Q\|_2 \leq C 2^{O(q)} n^{q/4} / q^{q/4} + t$

Proof. Consider any $x \in \mathbb{R}^{\binom{n}{q}}$ s.t. $\|x\| = 1$.

$$x^T \overline{A}_Q x = \sum_{\ell \in \binom{n}{q}} \sum_{i,j=\ell} \overline{A}_Q[i,j] x_i x_j = \sum_{\ell_1,\ell_2} \overline{A}_Q[\ell_1,\ell_2] \sum_{i,j=\ell} x_i x_j$$

(for arbitrary $\ell_1 \cup \ell_2 = \ell$)

Now since the above is a sum of independent scaled sub-gaussians, this expression has the same distribution as a sub-gaussian random variable with squared sub-gaussian norm

$$\lesssim K^2_2 \sum_{\ell_1,\ell_2} \left( \sum_{i,j=\ell} x_i x_j \right)^2 \lesssim K^2_2 \sum_{\ell} \sum_{i,j=\ell} x_i^2 x_j^2$$

(by Cauchy-Schwarz)

$$= K^2_2 2^{2q} \|x\|^2 = K^2_2 2^{2q} \approx 2^{2q} \quad (K_2 = O(1))$$

Thus for any fixed $x$, we may use the exponential sub-gaussian tail bounds (Definition B.1) on the above quadratic form, and the rest of the claim follows by applying union bound over a sufficiently fine net of the unit sphere. 

References

[BBH+12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In Proceedings of the forty-fourth annual ACM symposium on Theory of computing, pages 307–326. ACM, 2012.

[BH13] Fernando GSL Brandao and Aram W Harrow. Quantum de finetti theorems under local measurements with applications. In Proceedings of the forty-fifth annual ACM symposium on Theory of computing, pages 861–870. ACM, 2013.

[BKS14] Boaz Barak, Jonathan A Kelner, and David Steurer. Rounding sum-of-squares relaxations. In Proceedings of the 46th Annual ACM Symposium on Theory of Computing, pages 31–40. ACM, 2014.

[BKS15] Boaz Barak, Jonathan A Kelner, and David Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, pages 143–151. ACM, 2015.

[BV09] S Charles Brubaker and Santosh S Vempala. Random tensors and planted cliques. In Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pages 406–419. Springer, 2009.

[DK08] Etienne De Klerk. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. Central European Journal of Operations Research, 16(2):111–125, 2008.

[dKLS14] Etienne de Klerk, Monique Laurent, and Zhao Sun. Convergence analysis for lasserre’s measure–based hierarchy of upper bounds for polynomial optimization. arXiv preprint arXiv:1411.6867, 2014.
[DW12] Andrew C Doherty and Stephanie Wehner. Convergence of sdp hierarchies for polynomial optimization on the hypersphere. arXiv preprint arXiv:1210.5048, 2012.

[FK00] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. Random Structures and Algorithms, 16(2):195–208, 2000.

[FK08] Alan Frieze and Ravi Kannan. A new approach to the planted clique problem. In LIPIcs-Leibniz International Proceedings in Informatics, volume 2. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2008.

[GM15] Rong Ge and Tengyu Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, page 829, 2015.

[Hås96] Johan Håstad. Clique is hard to approximate within n 1-\epsilon. In Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on, pages 627–636. IEEE, 1996.

[HSS15] Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In Proceedings of The 28th Conference on Learning Theory, pages 956–1006, 2015.

[MR14] Andrea Montanari and Emile Richard. A statistical model for tensor PCA. In Advances in Neural Information Processing Systems, pages 2897–2905, 2014.

[Nes03] Yurii Nesterov. Random walk in a simplex and quadratic optimization over convex polytopes. Technical report, UCL, 2003.

[RRS16] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random csps below the spectral threshold. Preprint, 2016.

[Sim98] Barry Simon. The classical moment problem as a self-adjoint finite difference operator. Advances in Mathematics, 137(1):82–203, 1998.

[TS14] R. Tomioka and T. Suzuki. Spectral norm of random tensors. ArXiv e-prints, July 2014.

[Ver10] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. ArXiv e-prints, November 2010.