Sum-of-Squares Certificates for Maxima of Random Tensors on the Sphere

Vijay Bhattiprolu*[1], Venkatesan Guruswami†[2], and Euiwoong Lee‡[3]

1 **Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA**
`vpb@cs.cmu.edu`

2 **Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA**
`guruswami@cmu.edu`

3 **Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA**
`euiwoonl@cs.cmu.edu`

Abstract

For an n-variate order-d tensor $\mathcal{A}$, define $\mathcal{A}_{\max} := \sup_{\|x\|_2=1}\langle \mathcal{A}, x^{\otimes d}\rangle$ to be the maximum value taken by the tensor on the unit sphere. It is known that for a random tensor with i.i.d. ±1 entries, $\mathcal{A}_{\max} \lesssim \sqrt{n \cdot d \cdot \log d}$ w.h.p. We study the problem of efficiently certifying upper bounds on $\mathcal{A}_{\max}$ via the natural relaxation from the Sum of Squares (SoS) hierarchy. Our results include:

- When $\mathcal{A}$ is a random order-q tensor, we prove that q levels of SoS certifies an upper bound B on $\mathcal{A}_{\max}$ that satisfies

$$B \quad \leq \quad \mathcal{A}_{\max} \cdot \left(\frac{n}{q^{\,1-o(1)}}\right)^{q/4-1/2} \qquad \text{w.h.p.}$$

  Our upper bound improves a result of Montanari and Richard (NIPS 2014) when q is large.
- We show the above bound is the best possible up to lower order terms, namely the optimum of the level-q SoS relaxation is at least

$$\mathcal{A}_{\max} \cdot \left(\frac{n}{q^{\,1+o(1)}}\right)^{q/4-1/2}.$$

- When $\mathcal{A}$ is a random order-d tensor, we prove that q levels of SoS certifies an upper bound B on $\mathcal{A}_{\max}$ that satisfies

$$B \;\leq\; \mathcal{A}_{\max} \cdot \left(\frac{\widetilde{O}(n)}{q}\right)^{d/4-1/2} \qquad \text{w.h.p.}$$

  For growing q, this improves upon the bound certified by constant levels of SoS. This answers in part, a question posed by Hopkins, Shi, and Steurer (COLT 2015), who gave the tight characterization for constant levels of SoS.

1998 ACM Subject Classification G.1.6 Optimization, F.2.1 Numerical Algorithms and Problems

Keywords and phrases Sum-of-Squares; Optimization over Sphere; Random Polynomials

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2017.31

* Supported by NSF grants CCF-1422045 and CCF-1526092.

† Research supported in part by NSF grants CCF-1526092 and CCF-1563742.

‡ Supported by a Samsung Fellowship, a Simons Award for Graduate Students in Theoretical Computer Science, and NSF CCF-1526092.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017).
Editors: Klaus Jansen, José D. P. Rolim, David Williamson, and Santosh S. Vempala; Article No. 31; pp. 31:1–31:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

It is a well-known fact from random matrix theory that for an $n \times n$ matrix M whose entries are i.i.d. Rademacher or standard normal random variables, the maximum value $x^T M x$ taken by the associated quadratic form on the unit sphere $\|x\|_2 = 1$, is $\Theta(\sqrt{n})$ with high probability. Further, this maximum value can be computed efficiently for any matrix, as it equals the largest eigenvalue of $(M + M^T)/2$, so one can also efficiently certify that the maximum of a random quadratic form is at most $O(\sqrt{n})$.

This paper is motivated by the problem of analogous question for tensors. Namely, given a random order-d tensor $\mathcal{A}$ who entries are i.i.d. random ± entries, we would like to certify an upper bound on the maximum value $\mathcal{A}_{\max} := \max_{\|x\|=1} \langle \mathcal{A}, x^{\otimes d} \rangle$ taken by the tensor on the unit sphere. This value is at most $O_d(\sqrt{n})$ with high probability [18]. However, for $d \geq 3$, computing $\mathcal{A}_{\max}$ for a d-tensor $\mathcal{A}$ is NP-hard, and it is likely that the problem is also very hard to approximate. Assuming the Exponential Time Hypothesis, Barak et al. [1] proved that computing $2 \to 4$ norm of a matrix, a special case of computing the norm of a 4-tensor, is hard to approximate within a factor $\exp(\log^{1/2-\epsilon}(n))$ for any $\epsilon > 0$.

Our goal is to certify an *approximate* upper bound on $\mathcal{A}_{\max}$ is not too far from the true value. Specifically, we seek an estimate $B(\mathcal{A})$ which always upper bounds $\mathcal{A}_{\max}$, and with high probability is as close to $O_d(\sqrt{n})$ as possible for a random $\mathcal{A}$.

In addition to its intrinsic interest, the problem of maximizing tensors and closely related tasks of computing tensor norms, has connections to diverse topics, such as quantum information theory [7, 2], the Small Set Expansion Hypothesis (SSEH) and the Unique Games Conjecture (UGC) (via $2 \to 4$ norm, see [1, 2]), refuting random CSPs [16], tensor decomposition [3, 10], tensor PCA [15, 12], and planted clique (via the parity tensor, see [9, 8]). Many of these applications are of considerable interest in the 2^{n^ϵ}-runtime regime.

A natural approach to tackle the above problem is through the *Sum of Squares* (SoS) semidefinite programming relaxations. There are several ways to represent a tensor $\mathcal{A} \in \mathbb{R}^{[n]^d}$ (assume d is even) in matrix form as $M \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$ so that $\langle \mathcal{A}, x^{\otimes d} \rangle = (x^{\otimes d/2})^T M x^{\otimes d/2}$ for all $x \in \mathbb{R}^n$. The largest eigenvalue $\lambda_{\max}(M)$ of any such matrix representation M serves as an (efficiently computable) upper bound on $\mathcal{A}_{\max}$. The basic SoS relaxation looks for the best matrix representation, i.e., the one minimizing $\lambda_{\max}(M)$, among all possible representations of the tensor $\mathcal{A}$. This can be expressed as a semidefinite program, and also has a natural dual view in terms of pseudo-expectations or moment matrices (see Section 2.2).

The SoS hierarchy offers a sequence of relaxations, parameterized by the *level* q, with larger q giving a (potentially) tighter relaxation. In our context, this amounts to optimizing over matrix representations of $\mathcal{A}^{q/d}$ (we assume q is divisible by $2d$); in the dual view, this involves optimizing over pseudo-expectations for polynomials of degree up to q (as opposed to degree d for the basic relaxation). The level-q relaxation can be solved in $n^{O(q)}$ time by solving the associated semidefinite program. The SoS hierarchy thus presents a trade-off between approximation guarantee and runtime, with larger levels giving more accurate estimates at the expense of higher complexity.

This work is concerned with both positive and negative results on the efficacy of the SoS hierarchy to approximately certify the maxima of random tensors. We now turn to stating our results formally.

1.1 Our Results

For an order-q tensor $\mathcal{A} \in (\mathbb{R}^n)^{\otimes d}$, the polynomial $\mathcal{A}(x)$ and its maximum on the sphere $\mathcal{A}_{\max}$ are defined as

$$\mathcal{A}(x) := \langle \mathcal{A}, x^{\otimes d} \rangle \qquad \mathcal{A}_{\max} := \sup_{\|x\|=1} \mathcal{A}(x).$$

When the entries of $\mathcal{A}$ are i.i.d. Rademacher random variables (or i.i.d. Gaussians), it is known that $\mathcal{A}_{\max} \lesssim \sqrt{n \cdot d \cdot \log d}$ (see [18]). We will also use, for a polynomial g, $g_{\max}$ to denote $\sup_{\|x\|=1} g(x)$.

SoS degree = Polynomial Degree

We study the performance of degree-q SoS on random tensors of order-q. The formal definition and basic properties of SoS relaxations are presented in Section 2.2.

▶ **Theorem 1.** *For any even $q \leq n$, let $\mathcal{A} \in (\mathbb{R}^n)^{\otimes q}$ be a q-tensor with independent, Rademacher entries. With high probability, the value B of the degree-q SoS relaxation of $\mathcal{A}_{\max}$ satisfies*

$$2^{-O(q)} \cdot \left(\frac{n}{q}\right)^{q/4-1/2} \leq \frac{B}{\mathcal{A}_{\max}} \leq 2^{O(q)} \cdot \left(\frac{n}{q}\right)^{q/4-1/2}.$$

This improves upon the $O(n^{q/4})$ upper bound by Montanari and Richard [15].

SoS Degree ≫ Polynomial Degree

▶ **Theorem 2.** *Let $\mathcal{A} \in (\mathbb{R}^n)^{\otimes d}$ be a d-tensor with independent, Rademacher entries. Then for any even q satisfying $d \leq q \leq n$, with high probability, the degree-q SoS certifies an upper bound B on $\mathcal{A}_{\max}$ where w.h.p.,*

$$\frac{B}{\mathcal{A}_{\max}} \leq \left(\frac{\widetilde{O}(n)}{q}\right)^{d/4-1/2}$$

▶ Remark. Combining our upper bounds with the work of [12] would yield improved tensor-PCA guarantees on higher levels of SoS.

▶ Remark. Raghavendra, Rao, and Schramm [16] have independently and concurrently obtained similar (but weaker) results to Theorem 2 for random degree-d polynomials. Specifically, their upper bounds appear to require the assumption that the SoS level q must be less than $n^{1/(3d^2)}$ (our result only assumes $q \leq n$). Further, they certify an upper bound that matches Theorem 2 only when $q \leq 2^{\sqrt{\log n}}$.

1.2 Related Work

Upper Bounds

Montanari and Richard [15] presented a $n^{O(d)}$-time algorithm that can certify that the optimal value of $\mathcal{A}_{\max}$ for a random d-tensor is at most $O(n^{\frac{\lceil d/2 \rceil}{2}})$ with high probability. Hopkins, Shi, and Steurer [12] improved it to $O(n^{\frac{d}{4}})$ 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 $d = 3$.

Our analysis asymptotically improves the aforementioned bound when q is growing with n, and we prove an essentially matching lower bound (but only for the case $q = d$). Secondly, we consider the case when d is fixed, 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 [12].

Raghavendra, Rao, and Schramm [16] also prove results analogous to Theorem 2 for the case of *sparse* random polynomials (a model we do not consider in this work, and which appears to pose additional technical difficulties). This implied upper bounds for refuting random instances of constraint satisfaction problems using higher levels of the SoS hierarchy, which were shown to be tight via matching SoS lower bounds in [13].

Lower Bounds

While we only give lower bounds for the case of $q = d$, subsequent to our work, Hopkins et al. [11] proved the following theorem, which gives lower bounds for the case of $q \gg d$:

▶ **Theorem 3.** *Let f be a degree-d polynomial with i.i.d. gaussian coefficients. If there is some constant $\epsilon > 0$ such that $q \geq n^{\epsilon}$, then with high probability over f, the optimum of the level-q SoS relaxation of $f_{\max}$ is at least*

$$f_{\max} \cdot \Omega_d\left((n/q^{O(1)})^{d/4-1/2}\right) .$$

Note that this almost matches our upper bounds from Theorem 2, modulo the exponent of q. For this same reason, the above result does not completely recover our lower bound in Theorem 1 for the special case of $q = d$.

Results for worst-case tensors

It is proved in [5] that the q-level SoS gives an $(O(n)/q)^{d/2-1}$ approximation to $\|\mathcal{A}\|_2$ in the case of arbitrary d-tensors and an $(O(n)/q)^{d/4-1/2}$ approximation to $\mathcal{A}_{\max}$ in the case of d-tensors with non-negative entries (for technical reasons one can only approximate $\|\mathcal{A}\|_2 = \max\{|\mathcal{A}_{\max}|, |\mathcal{A}_{\min}|\}$ in the former case).

It is interesting to note that the approximation factor in the case of non-negative tensors matches the approximation factor (upto polylogs) we achieve in the random case. Additionally, the gap given by Theorem 1 for the case of random tensors provides the best degree-q SoS gap for the problem of approximating the 2-norm of arbitrary q-tensors. Hardness results for the arbitrary tensor 2-norm problem is an important pursuit due to its connection to various problems for which subexponential algorithms are of interest.

1.3 Organization

We begin by setting some important notation concerning SoS matrices, and describe some basic preliminaries about the SoS hierarchy in Section 2. We touch upon the main technical ingredients driving our work, and give an overview of the proof of Theorem 2 and the lower bound in Theorem 1 in Section 3. We present the proof of Theorem 2 for the case of even d in Section 4, with the more tricky odd d case handled in the full version of our paper [6]. The lower bound on the value of SoS-hierarchy claimed in Theorem 1 is proved in Section 5, and the upper bound in Theorem 1 also follows based on some techniques in that section.

2 Notation and Preliminaries

Multi-index and Multiset

A multi-index is defined as a sequence $\alpha \in \mathbb{N}^n$. We use $|\alpha|$ to denote $\sum_{i=1}^{n} \alpha_i$ and $\mathbb{N}^n_d$ (resp. $\mathbb{N}^n_{\leq d}$) to denote the set of all multi-indices α with $|\alpha| = d$ (resp. $|\alpha| \leq d$). We use $\mathbf{1}$ to denote the multi-index 1^n. Thus, a homogeneous polynomial f of degree d can be expressed in terms of its coefficients as

$$f(x) \;=\; \sum_{\alpha \in \mathbb{N}^n_d} f_\alpha \cdot x^\alpha,$$

where x^α is used to denote the monomial corresponding to α. In general, with the exception of absolute-value, any scalar function/operation when applied to vectors/multi-indices, returns the vector obtained by applying the function/operation entry-wise.

2.1 Matrices

For $k \in \mathbb{N}$, we will consider $[n]^k \times [n]^k$ matrices M with real entries. All matrices considered in this paper should be taken to be symmetric (unless otherwise stated). We index entries of the matrix M as $M[I, J]$ by tuples $I, J \in [n]^k$. $\oplus$ denotes tuple-concatenation.

A tuple $I = (i_1, \ldots, i_k)$ naturally corresponds to a multi-index $\alpha(I) \in \mathbb{N}^n_k$ with $|\alpha(I)| = k$, i.e. $\alpha(I)_j = |\{\ell \mid i_\ell = j\}|$. For a tuple $I \in [n]^k$, we define $\mathcal{O}(I)$ the set of all tuples J which correspond to the same multi-index i.e., $\alpha(I) = \alpha(J)$. Thus, any multi-index $\alpha \in \mathbb{N}^n_k$, corresponds to an equivalence class in $[n]^k$. We also use $\mathcal{O}(\alpha)$ to denote the class of all tuples corresponding to α.

Note that a matrix of the form $\left(x^{\otimes k}\right)\left(x^{\otimes k}\right)^T$ has many additional symmetries, which are also present in solutions to programs given by the SoS hierarchy. To capture this, consider the following definition:

▶ **Definition 4** (SoS-Symmetry). A matrix M which satisfies $\mathsf{M}[I, J] = \mathsf{M}[K, L]$ whenever $\alpha(I) + \alpha(J) = \alpha(K) + \alpha(L)$ is referred to as SoS-symmetric.

▶ **Definition 5** (Matrix-Representation). For a homogeneous degree-t (t even) polynomial g, we say a matrix $\mathrm{M}_g \in \mathbb{R}^{[n]^{t/2} \times [n]^{t/2}}$ is a degree-t matrix representation of g if for all x, $g(x) = (x^{\otimes t/2})^T\, \mathrm{M}_g\, x^{\otimes t/2}$. (We note here that every homogeneous polynomial has a unique SoS-Symmetric matrix representation.)

Note that $\lambda_{\max}(\mathrm{M}_g)$ is an upper bound on $g_{\max}$. This prompts the following relaxation of $g_{\max}$ that is closely related to the final SoS relaxation used in our upper bounds:

▶ **Definition 6.** For a homogeneous degree-t (t even) polynomial g, define

$$\Lambda(g) \;:=\; \inf\left\{\lambda_{\max}(M_g) \;\middle|\; M_g \text{ represents } g\right\}.$$

As we will see shortly, $\Lambda(g)$ is the dual of a natural SoS relaxation of $g_{\max}$.

2.2 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 $\widetilde{\mathbf{E}} : \mathbb{R}[x]_{\leq q} \mapsto \mathbb{R}$ such that

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

The pseudo-expectation operator $\widetilde{\mathbf{E}}$ can be completely described by the ***moment matrix*** (while x is a column vector, we abuse notation and let $(1, x)$ denote the column vector $(1, x_1, \ldots, x_n)^T$)

$$\overline{\mathsf{X}} := \widetilde{\mathbf{E}}\left[(1,x)^{\otimes q/2}\ ((1,x)^{\otimes q/2})^T\right]. \tag{2.1}$$

Moreover, the condition $\widetilde{\mathbf{E}}\left[p^2\right] \geq 0$ for all $p \in \mathbb{R}[x]_{\leq q/2}$ can be shown to be equivalent to $\overline{\mathsf{X}} \succeq 0$.

Constrained Pseudoexpectations

For a system of polynomial constraints

$$C = \{f_1 = 0, \ldots, f_m = 0, g_1 \geq 0, \ldots, g_r \geq 0\},$$

we say $\widetilde{\mathbf{E}}_C$ is a pseudoexpectation operator respecting C, if in addition to the above conditions, it also satisfies

1. $\widetilde{\mathbf{E}}_C[p \cdot f_i] = 0$, $\forall i \in [m]$ and $\forall p$ such that $\deg(p \cdot f_i) \leq q$.
2. $\widetilde{\mathbf{E}}_C\left[p^2 \cdot \prod_{i \in S} g_i\right] \geq 0$, $\forall S \subseteq [r]$ and $\forall p$ such that $\deg(p^2 \cdot \prod_{i \in S} g_i) \leq q$.

It is well-known that such constrained pseudoexpectation operators can be described as solutions to semidefinite programs of size $n^{O(q)}$ [4, 14]. This hierarchy of semidefinite programs for increasing q is known as the SoS hierarchy.

Additional Facts about SoS

We shall record here some well-known facts about SoS that come in handy later.

▶ **Claim 7.** *For polynomials p_1, p_2, let $p_1 \succeq p_2$ 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 $p_1 - p_2 \succeq 0$.*

▶ **Claim 8** (Pseudo-Cauchy-Schwarz [2])**.** $\widetilde{\mathbf{E}}\left[p_1 p_2\right] \leq (\widetilde{\mathbf{E}}\left[p_1^2\right] \widetilde{\mathbf{E}}\left[p_2^2\right])^{1/2}$ *for any p_1, p_2 of degree at most $q/2$.*

SoS Relaxations for $\mathcal{A}_{\max}$

Given an order-q tensor $\mathcal{A}$, our degree-q SoS relaxation for $\mathcal{A}_{\max}$ which we will henceforth denote by $\mathsf{SoS}_q(\mathcal{A}(x))$ is given by,

$$\begin{array}{rr} \text{maximize} & \widetilde{\mathbf{E}}_C[\mathcal{A}(x)] \\ \text{subject to :} & \widetilde{\mathbf{E}}_C \text{ is a degree-}q \\ & \text{pseudoexpectation} \\ & \widetilde{\mathbf{E}}_C \text{ respects } C \equiv \{\|x\|_2^q = 1\} \end{array}$$

Assuming q is divisible by $2d$, we make an observation that is useful in our upper bounds:

$$\mathcal{A}_{\max} \quad\leq\quad \mathsf{SoS}_q(\mathcal{A}(x)) \quad\leq\quad \mathsf{SoS}_q\left(\mathcal{A}(x)^{\,q/d}\right)^{d/q} \;=\; \Lambda\left(\mathcal{A}(x)^{\,q/d}\right)^{d/q} \tag{2.2}$$

where the second inequality follows from Pseudo-Cauchy-Scwarz, and the equality follows from well known strong duality of the following programs (specifically, take $g(x) := \mathcal{A}(x)^{\,q/d}$):[1]

[1] Compared to (2.1), the primal formulation here uses a *homogeneous* moment matrix or pseudo-expectation operator, defined for polynomials of degree exactly q.

Dual

$$\Lambda(g) \ := \ \inf \left\{ \lambda_{\max}(M_g) \ \middle| \ M_g \text{ represents } g \right\}$$

Primal I

$$\begin{aligned} \text{maximize} \quad & \langle \mathsf{M}_g, \mathsf{X} \rangle \\ \text{subject to :} \quad & \mathbf{Tr}(\mathsf{X}) = 1 \\ & \mathsf{X} \text{ is SoS symmetric} \\ & \mathsf{X} \succeq 0 \end{aligned}$$

Primal II

$$\begin{aligned} \text{maximize} \quad & \widetilde{\mathbf{E}}_C[g] \\ \text{subject to :} \quad & \widetilde{\mathbf{E}}_C \text{ is a degree-}q \\ & \text{pseudoexpectation} \\ & \widetilde{\mathbf{E}}_C \text{ respects } C \equiv \left\{ \|x\|_2^q = 1 \right\} \end{aligned}$$

Figure 2.1 Duals of $\Lambda(g)$ for the degree-q homogeneous polynomial g.

Note

In the rest of the paper, we will drop the subscript C of the pseudo-expectation operator since throughout this work, we only assume the hypersphere constraint.

3 Overview of our Methods

We now give a high level view of the two broad techniques driving this work, followed by a more detailed overview of the proofs.

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 had 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 "good permutations", and moreover that there are $q^q \cdot 2^{-O(q)}$ such good permutations. On averaging the representations corresponding to these good permutations, we obtain a matrix that admits similar spectral properties to those of a matrix with i.i.d. entries, and with much lower variance (in most of the entries) compared to the naive representations.

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, $\mathbf{E}\left[p^2\right]$ 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, and we believe it to be generally useful for high degree SoS lower bounds.

3.1 Overview of Upper Bound Proofs

For even d, let $\mathcal{A} \in \mathbb{R}^{[n]^d}$ be a d-tensor with i.i.d. ±1 entries and let $A \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$ be the matrix flattening of $\mathcal{A}$, i.e., $A[I, J] = \mathcal{A}[I \oplus J]$ (recall that $\oplus$ denotes tuple concatenation). Also let $f(x) := \mathcal{A}(x) = \langle \mathcal{A}, x^{\otimes d} \rangle$. It is well known that $f_{\max} \le O(\sqrt{n \cdot d \cdot \log d})$ with high probability [18]. For such a polynomial f and any q divisible by d, in order to establish Theorem 2, by Eq. (2.2) it is sufficient to prove that with high probability,

$$\left(\Lambda\left(f^{q/d}\right)\right)^{d/q} \le \widetilde{O}\left(\frac{n}{q^{1-2/d}}\right)^{d/4} = \widetilde{O}\left(\frac{n}{q}\right)^{d/4-1/2} \cdot f_{\max}.$$

We give an overview of the proof. Let $d = 4$ for the sake of clarity of exposition. To prove an upper bound on $\Lambda\left(f^{q/4}\right)$ using degree-q SoS (assume q is a multiple of 4), we define a suitable matrix representation $M := M_{f^{q/4}} \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$ of $f^{q/4}$ and bound $\|M\|_2$. Since $\Lambda(f) \le (\|M\|_2)^{q/4}$ for any representation M, a good upper bound on $\|M\|_2$ certifies that $\Lambda(f)$ is small.

One of the intuitive reasons taking a high power gives a better bound on the spectral norm is that this creates more entries of the matrix that correspond to the same monomial, and distributing the coefficient of this monomial equally among the corresponding entries reduces variance (i.e., $\mathbf{Var}[X]$ is less than $k \cdot \mathbf{Var}[X/k]$ for $k > 1$). In this regard, the most natural representation M of $f^{q/4}$ is the *complete symmetrization.*

$$\begin{aligned} & M_c[(i_1, \dots, i_{q/2}), (i_{q/2+1}, \dots, i_q)] \\ &= \frac{1}{q!} \cdot \sum_{\pi \in \mathbb{S}_q} A^{\otimes q/4}[(i_{\pi(1)}, \dots, i_{\pi(q/2)}), (i_{\pi(q/2+1)}, \dots, i_{\pi(q)})] \\ &= \frac{1}{q!} \cdot \sum_{\pi \in \mathbb{S}_q} \prod_{j=1}^{q/4} A[(i_{\pi(2j-1)}, i_{\pi(2j)}), (i_{\pi(q/2+2j-1)}, i_{\pi(q/2+2j)})]. \end{aligned}$$

However, $\|M_c\|_2$ turns out to be much larger than $\Lambda(f)$, even when $q = 8$. One intuitive explanation is that M_c, as a $n^4 \times n^4$ matrix, contains a copy of $\mathbf{Vec}(A)\,\mathbf{Vec}(A)^T$, where $\mathbf{Vec}(A) \in \mathbb{R}^{[n]^4}$ is the vector with $\mathbf{Vec}(A)[i_1, i_2, i_3, i_4] = A[(i_1, i_2), (i_3, i_4)]$. Then $\mathbf{Vec}(A)$ is a vector that witnesses $\|M_c\|_2 \ge \Omega(n^2)$, regardless of the randomness of f. Our final representation[2] is the following *row-column independent symmetrization* that simultaneously respects the spectral structure of a random matrix A and reduces the variance. Our M is given by

$$\begin{aligned} & M[(i_1, \dots, i_{q/2}), (j_1, \dots, j_{q/2})] \\ &= \frac{1}{(q/2)!^2} \cdot \sum_{\pi, \sigma \in \mathbb{S}_{q/2}} A^{\otimes q/4}[(i_{\pi(1)}, \dots, i_{\pi(q/2)}), (j_{\sigma(1)}, \dots, j_{\sigma(q/2)})] \\ &= \frac{1}{(q/2)!^2} \cdot \sum_{\pi, \sigma \in \mathbb{S}_{q/2}} \prod_{k=1}^{q/4} A[(i_{\pi(2k-1)}, i_{\pi(2k)}), (j_{\sigma(2k-1)}, j_{\sigma(2k)})]. \end{aligned}$$

[2] The independent and concurrent work of [16] uses the same representation.

To formally show $\|M\|_2 = \tilde{O}(n/\sqrt{q})^{q/4}$ with high probability, we use the trace method to show

$$\mathbf{E}\left[\mathbf{Tr}(M^p)\right] \leq 2^{O(pq\log p)} \frac{n^{pq/4+q/2}}{q^{pq/8}},$$

where $\mathbf{E}\left[\mathbf{Tr}(M^p)\right]$ can be written as (let $I^{p+1} := I^1$)

$$\begin{aligned}
&\mathbf{E}\left[\sum_{I^1,\dots,I^p\in[n]^{q/2}} \prod_{j=1}^{p} M[I^j, I^{j+1}]\right] \\
&= \sum_{I^1,\dots,I^p} \mathbf{E}\left[\prod_{j=1}^{p}\Big(\sum_{\pi_j,\sigma_j\in\mathbb{S}_{q/2}} \prod_{k=1}^{q/4} A[(I^k_{\pi_j(2k-1)}, I^k_{\pi_j(2k)}), (I^{k+1}_{\sigma_j(2k-1)}, I^{k+1}_{\sigma_j(2k)})]\Big)\right].
\end{aligned}$$

Let $E(I^1,\dots,I^p)$ be the expectation value for $I^1,\dots,I^p$ in the right hand side. We study $E(I^1,\dots,I^p)$ for each $I^1,\dots,I^p$ by careful counting of the number of permutations on a given sequence with possibly repeated entries. For any $I^1,\dots,I^p \in [n]^{q/2}$, let $\#\left(I^1,\dots,I^p\right)$ denote the number of distinct elements of $[n]$ that occur in $I^1,\dots,I^p$, and for each $s = 1,\dots,\#\left(I^1,\dots,I^p\right)$, let $c^s \in (\{0\}\cup[q/2])^p$ denote the number of times that the jth smallest element occurs in $I^1,\dots,I^p$. When $E(I^1,\dots,I^p) \neq 0$, it means that for some permutations $\{\pi_j,\sigma_j\}_j$, every term $A[\cdot,\cdot]$ must appear even number of times. This implies that the number of distinct elements in $I^1,\dots,I^p$ is at most half the maximal possible number $pq/2$. This lemma proves the intuition via graph theoretic arguments.

▶ **Lemma 9.** *If $E(I^1,\dots,I^p) \neq 0$, $\#\left(I^1,\dots,I^p\right) \leq \frac{pq}{4} + \frac{q}{2}$.*

The number of $I^1,\dots,I^p$ that corresponds to a sequence $c^1,\dots,c^s$ is at most $\frac{n^s}{s!}\cdot\frac{((q/2)!)^p}{\prod_{\ell\in[p]} c^1_\ell!\cdot c^p_\ell!}$. Furthermore, there are at most $2^{O(pq)}p^{pq/2}$ different choices of $c^1,\dots,c^s$ that corresponds to some $I^1,\dots,I^p$. The following technical lemma bounds $E(I^1,\dots,I^p)$ by careful counting arguments.

▶ **Lemma 10.** *For any $I^1,\dots,I^p$, $E(I^1,\dots,I^p) \leq 2^{O(pq)}\frac{p^{5pq/8}}{q^{3pq/8}}\prod_{\ell\in[p]} c^1_\ell!\dots c^s_\ell!$.*

Summing over all s and multiplying all possibilities,

$$\begin{aligned}
\mathbf{E}\left[\mathbf{Tr}(M^p)\right] &\leq \sum_{s=1}^{pq/4+q/2}\left(2^{O(pq)}p^{pq/2}\right)\cdot\left(\frac{n^s}{s!}\cdot((q/2)!)^p\right)\cdot\left(2^{O(pq)}\frac{p^{5pq/8}}{q^{3pq/8}}\right) \\
&= \max_{1\leq s\leq pq/4+q/2} 2^{O(pq\log p)}\cdot n^s\cdot\frac{q^{pq/8}}{s!}.
\end{aligned}$$

When $q \leq n$, the maximum occurs when $s = pq/4+q/2$, so $\mathbf{E}\left[\mathbf{Tr}(M^p)\right] \leq 2^{O(pq\log p)}\cdot\frac{n^{pq/4+q/2}}{q^{pq/8}}$ as desired.

3.2 Overview of Lower Bound Proofs

Let $\mathcal{A}, A, f$ be as in Section 3.1. To prove the lower bound in Theorem 1, we construct a moment matrix M that is positive semidefinite, SoS-symmetric, $\mathbf{Tr}(\mathsf{M}) = 1$, and $\langle A, \mathsf{M}\rangle \geq 2^{-O(d)}\cdot\frac{n^{d/4}}{d^{d/4}}$. At a high level, our construction is $\mathsf{M} := c_1\mathsf{A} + c_2\mathsf{W}$ for some c_1, c_2, where A contains entries of A only corresponding to the multilinear indices, averaged over all SoS-symmetric positions. This gives a large inner product with A, SoS-symmetry, and nice spectral properties even though it is not positive semidefinite. The most natural way to make it positive semidefinite is adding a copy of the identity matrix, but this will again break the SoS-symmetry.

Our main technical contribution here is the construction of W that acts like a *SoS-symmetrized identity*. It has the minimum eigenvalue at least $\frac{1}{2}$, while the trace being $n^{d/2} \cdot 2^{O(d)}$, so the ratio of the average eigenvalue to the minimum eigenvalue is bounded above by $2^{O(d)}$, which allows us to prove a tight lower bound. To the best of our knowledge, no such bound was known for SoS-symmetric matrices except small values of $d = 3, 4$.

Given $I, J \in [n]^{d/2}$, we let $\mathsf{W}[I, J] := \mathbf{E}[x^{\alpha(I)+\alpha(J)}]$, where $x_1, \ldots, x_n$ are independently sampled from the *Wigner semicircle distribution*, whose probability density function is the semicircle $f(x) = \frac{2}{\pi}\sqrt{1 - x^2}$. Since $\mathbf{E}[x_1^{\ell}] = 0$ if ℓ is odd and $\mathbf{E}[x_1^{2\ell}] = \frac{1}{\ell+1}\binom{2\ell}{\ell}$, which is the ℓth Catalan number, each entry of W is bounded by $2^{O(d)}$ and $\mathbf{Tr}(\mathsf{W}) \le n^{d/2} \cdot 2^{O(d)}$. To prove a lower bound on the minimum eigenvalue, we show that for any degree-ℓ polynomial p with m variables, $\mathbf{E}[p(x_1, \ldots, x_m)^2]$ is large by induction on ℓ and m. We use another property of the Wigner semicircle distribution that if $H \in \mathbb{R}^{(d+1)\times(d+1)}$ is the univariate moment matrix of x_1 defined by $H[i, j] = \mathbf{E}[x_1^{i+j}]$ $(0 \le i, j \le d)$ and $H = (R^T)R$ is the Cholesky decomposition of H, R is an upper triangular matrix with 1's on the main diagonal. This nice Cholesky decomposition allows us to perform the induction on the number of variables while the guarantee on the minimum eigenvalue is independent of n.

4 Upper bounds for even degree tensors

For even d, let $\mathcal{A} \in \mathbb{R}^{[n]^d}$ be a d-tensor with i.i.d. ±1 entries and let $A \in \mathbb{R}^{[n]^{d/2} \times [n]^{d/2}}$ be the matrix flattening of $\mathcal{A}$, i.e., $A[I, J] = \mathcal{A}[I \oplus J]$ (recall that $\oplus$ denotes tuple concatenation). Also let $f(x) := \mathcal{A}(x) = \langle \mathcal{A}, x^{\otimes d} \rangle$. With high probability $f_{\max} = O(\sqrt{n \cdot d \cdot \log d})$. In this section, we prove that for every q divisible by d, with high probability,

$$\left(\Lambda\left(f^{q/d}\right)\right)^{d/q} \le \widetilde{O}\left(\frac{n}{q^{1-2/d}}\right)^{d/4} = \widetilde{O}\left(\frac{n}{q}\right)^{d/4-1/2} \cdot f_{\max}.$$

To prove it, we use the following matrix representation M of $f^{q/d}$, and show that $\|M\|_2 \le \tilde{O}_d\left(\left(\frac{n \log^5 n}{q^{1-2/d}}\right)^{q/4}\right)$. Given a tuple $I = (i_1, \ldots, i_q)$, and an integer d that divides q and $1 \le \ell \le q/d$, let $I_{\ell;d}$ be the d-tuple $(I_{d(\ell-1)+1}, \ldots, I_{d\ell})$ (i.e., if we divide I into q/d tuples of length d, $I_{\ell;d}$ be the ℓ-th tuple). Furthermore, given a tuple $I = (i_1, \ldots, i_q) \in [n]^q$ and a permutation $\pi \in [n]^q$, let $\pi(I)$ be another q-tuple whose ℓth coordinate is $\pi(i_\ell)$. For $I, J \in [n]^{q/2}$, $M[I, J]$ is formally given by

$$\begin{aligned} M[I, J] &= \frac{1}{q!} \cdot \sum_{\pi, \sigma \in \mathbb{S}_{q/2}} A^{\otimes q/d}[\pi(I), \sigma(J)] \\ &= \frac{1}{q!} \cdot \sum_{\pi, \sigma \in \mathbb{S}_{q/2}} \prod_{\ell=1}^{q/d} A[(\pi(I))_{\ell;d/2}, (\sigma(J))_{\ell;d/2}]. \end{aligned}$$

We perform the trace method to bound $\|M\|_2$. Let p be an even integer, that will be eventually taken as $\Theta(\log n)$. $\mathbf{Tr}(M)$ can be written as (let $I^{p+1} := I^1$)

$$\mathbf{E}\left[\sum_{I^1, \ldots, I^p \in [n]^{q/2}} \prod_{\ell=1}^{p} M[I^\ell, I^{\ell+1}]\right]$$

$$= \sum_{I^1,\ldots,I^p} \mathbf{E}\left[\prod_{\ell=1}^{p}\Big(\sum_{\pi_j,\sigma_j\in\mathbb{S}_{q/2}} \prod_{m=1}^{q/d} A[(\pi(I^\ell))_{m;d/2},(\sigma(I^{\ell+1}))_{m;d/2})]\Big)\right].$$

Let $E(I^1,\ldots,I^p) := \mathbf{E}\left[\prod_{\ell=1}^{p} M[I^\ell, I^{\ell+1}]\right]$, which is the expected value in the right hand side. To analyze $E(I^1,\ldots,I^p)$, we first introduce notions to classify $I^1,\ldots,I^p$ depending on their intersection patterns. For any $I^1,\ldots,I^p \in [n]^{q/2}$, let e_k denote the k-th smallest element in $\bigcup\limits_{\ell,\,j}\{i_j^\ell\}$. For any $c^1,\ldots,c^s \in [q/2]^p$, let

$$\mathcal{C}(c^1\ldots c^s) :=$$
$$\left\{(I^1,\ldots,I^p)\ \Big|\ \#\left(I^1,\ldots,I^p\right)=s,\ \forall k\in[s],\ell\in[p],\ e_k \text{ appears } c_\ell^k \text{ times in } I^\ell\right\}.$$

The following two observations on $c^1,\ldots,c^s$ can be easily proved.

▶ **Observation 11.** *If $\mathcal{C}(c^1,\ldots,c^s)\neq\phi$,*

$$\left|\mathcal{C}(c^1,\ldots,c^s)\right| \le \frac{n^s}{s!}\times\frac{((q/2)!)^p}{\prod\limits_{\ell\in[p]} c_\ell^1!\ldots c_\ell^s!}.$$

Moreover,

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

The following lemma bounds $E(I^1,\ldots,I^p)$ in terms of the corresponding $c_1,\ldots,c_s$.

▶ **Lemma 12.** *Consider any $c^1,\ldots,c^s\in[q/2]^p$ and $(I^1,\ldots,I^p)\in\mathcal{C}(c^1,\ldots,c^s)$. We have*

$$E(I^1,\ldots,I^p) \le 2^{O(pq)}\,\frac{p^{1/2+1/2d}}{q^{1/2-1/2d}}\prod_{\ell\in[p]} c_\ell^1!\ldots c_\ell^s!$$

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

$$\begin{aligned}
&E(I^1,\ldots,I^p)\\
&= \mathbf{E}\left[\prod_{\ell=1}^{p} M[I^\ell,I^{\ell+1}]\right]\\
&= \sum_{\pi_j,\sigma_j\in\mathbb{S}_{q/2}} \mathbf{E}\left[\prod_{\ell=1}^{p}\prod_{m=1}^{q/d} A[(\pi(I^\ell))_{m;d/2},(\pi(I^{\ell+1}))_{m;d/2}]\right]\\
&= \left(\frac{\prod_\ell\prod_s (c_\ell^s!)^2}{((q/2)!)^{2p}}\right)\cdot \sum_{(J^\ell,K^\ell\in\mathcal{O}(I^\ell))_{\ell\in[p]}} \mathbf{E}\left[\prod_{\ell=1}^{p}\prod_{m=1}^{q/d} A[J^\ell_{m;d/2},K^{\ell+1}_{m;d/2}]\right] \qquad (4.1)
\end{aligned}$$

Thus, $E(I^1,\ldots,I^p)$ is bounded by the number of choices for $J^1,\ldots,J^p,K^1,\ldots,K^p$ such that $J^\ell,K^\ell\in\mathcal{O}(I^\ell)$ for each $\ell\in[p]$, and $\mathbf{E}\left[\prod_{\ell=1}^{p}\prod_{m=1}^{q/d} A[J^\ell_{m;d/2},K^{\ell+1}_{m;d/2}]\right]$ is nonzero.

Given $J^1,\ldots,J^p$ and $K^1,\ldots,K^p$, consider the (pq/d)-tuple T where each coordinate is indexed by $(\ell,m)_{\ell\in[p],m\in[q/d]}$ and has a d-tuple $T_{\ell,m} := (J^\ell_{m;d/2})\oplus(K^{\ell+1}_{m;d/2})\in\mathbb{R}^d$ as a value. Note that $\sum_{\ell,m}\alpha(T_{\ell,m})) = (2o_1,\ldots,2o_n)$ where o_r is the number of occurences of $r\in[n]$ in $(pq/2)$-tuple $\oplus_{\ell=1}^p I^\ell$. The fact that $\mathbf{E}\left[\prod_{\ell=1}^{p}\prod_{m=1}^{q/d} A[j_{m;d/2},k_{m;d/2}]\right]\neq 0$ means that every d-tuple occurs even number of times in T.

We count the number of (pq/d)-tuples $T = (T_{\ell,m})_{\ell\in[p],m\in[q]}$ that $\sum_{\ell,m}\alpha(T_{\ell,m}) = (2o_1,\ldots,2o_n)$ and every d-tuple occurs an even number of times. Let $Q = (Q_1,\ldots,Q_{pq/2d})$, $R = (R_1,\ldots,R_{pq/2d})$ be two $(pq/2d)$-tuples of d-tuples where for every d-tuple P, the number of occurences of P is the same in Q and R, and $\sum_{\ell=1}^{pq/2d}\alpha(Q_\ell) = \sum_{\ell=1}^{pq/2d}\alpha(R_\ell) = (o_1,\ldots,o_n)$. At most $2^{pq/d}$ tuples T can be made by *interleaving* Q and R – for each (ℓ,m), choose $T_{\ell,m}$ from the first unused d-tuple in either Q or R. Furthermore, every tuple T that meets our condition can be constructed in this way.

Due to the condition $\sum_{\ell=1}^{pq/2d}\alpha(Q_\ell) = (o_1,\ldots,o_n)$, the number of choices for Q is at most the number of different ways to permute $I^1\oplus\cdots\oplus I^p$, which is at most $(pq/2)!/\prod_{m\in[s]}(\bar{c}^m)!$, where $\bar{c}^m := \sum_{\ell\in[p]} c_\ell^m$ for $m\in[s]$. For a fixed choice of Q, there are at most $(pq/2d)!$ choices of R. Therefore, the number of choices for $(J^\ell, K^\ell \in \mathcal{O}(I^\ell))_{\ell\in[p]}$ with nonzero expected value is at most

$$2^{pq/d}\cdot\frac{(pq/2)!}{\prod_{m\in[s]}(\bar{c}^m)!}\cdot(pq/2d)! = 2^{O(pq)}\cdot\frac{(pq)^{1/2+1/2d}}{\prod_{m\in[s]}(\bar{c}^m)!}.$$

Combining with Eq. (4.1),

$$E(I^1,\ldots,I^p) \le \left(2^{O(pq)}\frac{(pq)^{1/2+1/2d}}{\prod_{m\in[s]}(\bar{c}^m)!}\right)\cdot\left(\frac{\prod_\ell\prod_s(c_\ell^s!)^2}{((q/2)!)^{2p}}\right) \le 2^{O(pq)}\cdot\frac{p^{1/2+1/2d}}{q^{1/2-1/2d}}\cdot\prod_\ell\prod_s c_\ell^s!$$

as desired. ◀

▶ **Lemma 13.** *For all $I^1,\ldots,I^p\in[n]^{q/2}$, if $E(I^1,\ldots,I^p)\ne 0$, $\#\left(I^1,\ldots,I^p\right)\le\frac{pq}{4}+\frac{q}{2}$.*

Proof. Note that $E(I^1,\ldots,I^p)\ne 0$ implies that there exist $J^1,\ldots,J^p,K^1,\ldots,K^p$ such that $J^\ell,K^\ell\in\mathcal{O}(I^\ell)$ and every d-tuple occurs exactly even number of times in $((J^\ell_{m;d/2})\oplus(K^{\ell+1}_{m;d/2}))_{\ell\in[p],m\in[q/d]}$. Consider the graph $G=(V,E)$ defined by

$$V := \bigcup_{\ell\in[p]}\bigcup_{k\in[q/2]}\{I^\ell_k\}$$
$$E := \bigcup_{m\in[q/2]}\left\{\{J^1_m,K^2_m\},\{J^2_m,K^3_m\},\ldots,\{J^p_m,K^1_m\}\right\}.$$

The even multiplicity condition implies that every element in E has even multiplicity and consequently $|E|\le pq/4$. We next show that E is the union of $q/2$ paths. To this end, we construct $G^1\in\mathcal{O}(I^1),\ldots,G^\ell\in\mathcal{O}(I^\ell)$ as follows:

1. Let $G^2 := K^2$
2. For $3\le\ell\le p$ do:
   a. Since $G^\ell\in\mathcal{O}(J^\ell)$, there exists $\pi\in\mathbb{S}_{q/2}$ s.t. $\pi(J^\ell)=G^\ell$.
   b. Let $G^{\ell+1} := \pi(K^{\ell+1})$.

We observe that by construction,

$$\bigcup_{m\in[q/2]}\left\{\{J^1_m,G^2_m\},\{G^2_m,G^3_m\},\ldots,\{G^p_m,G^1_m\}\right\}$$
$$= \bigcup_{m\in[q/2]}\left\{\{J^1_m,K^2_m\},\{J^2_m,K^3_m\},\ldots,\{J^p_m,K^1_m\}\right\} = E$$

which establishes that E is a union of $q/2$ paths.

Now since E is the union of $q/2$ paths G has at most $q/2$ connected components, and one needs to add at most $q/2-1$ edges make it connected, we have $|V|\le|E|+(q/2-1)+1\le pq/4+q/2$. But $\#\left(I^1,\ldots,I^p\right)=|V|$, which completes the proof. ◀

Finally, $\mathbf{E}\left[\mathbf{Tr}(M^p)\right]$ can be bounded as follows.

$$\begin{aligned}
&\mathbf{E}\left[\mathbf{Tr}(M^p)\right] \\
&= \sum_{I^1,\ldots,I^p \in [n]^{q/2}} E(I^1,\ldots,I^p) \\
&= \sum_{s\in[pq/4+q/2]} \sum_{\#(I^1,\ldots,I^p)=s} E(I^1,\ldots,I^p) && \text{(by Lemma 13)} \\
&= \sum_{s\in[pq/4+q/2]} \sum_{c^1,\ldots,c^s\in[q/2]^p} \sum_{(I^1,\ldots,I^p)\in\mathcal{C}(c^1\ldots c^s)} E(I^1,\ldots,I^p) \\
&= \sum_{s\in[pq/4+q/2]} \sum_{c^1,\ldots,c^s\in[q/2]^p} \sum_{(I^1,\ldots,I^p)\in\mathcal{C}(c^1\ldots c^s)} E(I^1,\ldots,I^p) \\
&\leq \sum_{s\in[pq/4+q/2]} \sum_{c^1,\ldots,c^s\in[q/2]^p} \\
&\qquad \sum_{(I^1,\ldots,I^p)\in\mathcal{C}(c^1\ldots c^s)} 2^{O(pq)}\, \frac{p^{(1/2+1/2d)pq}}{q^{(1/2-1/2d)pq}} \prod_{\ell\in[p]} c^1_\ell! \ldots c^s_\ell! && \text{(by Lemma 12)} \\
&\leq \sum_{s\in[pq/4+q/2]} 2^{O(pq)}\, \frac{n^s}{s!}\, p^{(1+1/2d)pq} q^{pq/2d} && \text{(by Observation 11)} \\
&\leq \sum_{s\in[pq/4+q/2]} 2^{O(pq)}\, \frac{n^{pq/4+q/2}}{s!\, q^{pq/4+q/2-s}}\, p^{(1/2+1/2d)p1} q^{(1/2-1/2d)pq} && \text{(assuming } q \leq n\text{)} \\
&\leq \sum_{s\in[pq/4+q/2]} 2^{O(pq)}\, \frac{n^{pq/4+q/2}\, p^{(1+1/2d)pq}}{q^{(1/4-1/2d)pq}} \\
&\leq 2^{O(pq)}\, \frac{n^{pq/4+q/2}\, p^{(1+1/2d)pq}}{q^{(1/4-1/2d)pq}}.
\end{aligned}$$

Choose p to be even and let $p = \Theta(\log n)$. Applying Markov inequality shows that with high probability,

$$\left(\Lambda\left(f^{q/d}\right)\right)^{d/q} \leq \left(\|M\|_2\right)^{d/q} \leq \left(\mathbf{E}\left[\mathbf{Tr}(M^p)\right]\right)^{d/pq} = O_d\left(\frac{n^{d/4}\cdot(\log n)^{\;d+1/2}}{q^{\;d/4-1/2}}\right).$$

Thus we obtain

▶ **Theorem 14.** *For even d, let $\mathcal{A} \in \mathbb{R}^{[n]^d}$ be a d-tensor with i.i.d. ± 1 entries. Then for any even q such that $q \leq n$, we have that with probability $1 - n^{\Omega(1)}$,*

$$\frac{\mathsf{SoS}_q(\mathcal{A}(x))}{\mathcal{A}_{\max}} \leq \left(\frac{\widetilde{O}(n)}{q}\right)^{d/4-1/2}.$$

5 Proof of SoS Lower Bound in Theorem 1

For even q, let $\mathcal{A} \in \mathbb{R}^{[n]^q}$ be a q-tensor with i.i.d. ± 1 entries and let $A \in \mathbb{R}^{[n]^{q/2}\times[n]^{q/2}}$ be the matrix flattening of $\mathcal{A}$, i.e., $A[I,J] = \mathcal{A}[I \oplus J]$ (recall that $\oplus$ denotes tuple concatenation). Also let $f(x) := \mathcal{A}(x) = \langle \mathcal{A}, x^{\otimes q}\rangle$. This section proves the lower bound in Theorem 1, by constructing a moment matrix M that is positive semidefinite, SoS-symmetric, $\mathbf{Tr}(\mathsf{M}) = 1$, and $\langle A, \mathsf{M}\rangle \geq 2^{-O(q)} \cdot \frac{n^{q/4}}{q^{q/4}}$. In Section 5.1, we construct the matrix $\widehat{\mathsf{W}}$ that acts as a SoS-symmetrized identity matrix. The moment matrix M is presented in Section A.

5.1 Wigner Moment Matrix

In this section, we construct an SoS-symmetric and positive semidefinite matrix $\widehat{\mathsf{W}} \in \mathbb{R}^{\mathbb{N}^n_{q/2} \times \mathbb{N}^n_{q/2}}$ such that $\lambda_{\min}(\widehat{\mathsf{W}}) / \mathbf{Tr}\left(\widehat{\mathsf{W}}\right) \geq 1/(2^{q+1} \cdot |\mathbb{N}^n_{q/2}|)$, i.e. the ratio of the minimum eigenvalue to the average eigenvalue is at least $1/2^{q+1}$.

▶ **Theorem 15.** *For any positive integer n and any positive even integer q, there exists a matrix $\widehat{\mathsf{W}} \subseteq \mathbb{R}^{\mathbb{N}^n_{q/2} \times \mathbb{N}^n_{q/2}}$ that satisfies the following three properties: (1) $\widehat{\mathsf{W}}$ is degree-q SoS symmetric. (2) The minimum eigenvalue of $\widehat{\mathsf{W}}$ is at least $\frac{1}{2}$. (3) Each entry of $\widehat{\mathsf{W}}$ is in $[0, 2^q]$.*

Theorem 15 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 $\frac{q}{2}$, $\mathbf{E}[p^2]$ 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 $\mathbf{E}[p^2]$ is bounded away from 0. The third part constructs $\widehat{\mathsf{W}} \subseteq \mathbb{R}^{\mathbb{N}^n_{q/2} \times \mathbb{N}^n_{q/2}}$ 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 $(e_0)_0 = 1$ 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 parantheses. Given a string of parantheses '(' or ')', we call it *consistent* if any prefix has at least as many '(' as ')'. For example, ((())(is consistent, but ())((is not.

▶ **Claim 16.** *$R[i,j]$ is the number of ways to place j parantheses '(' 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 parantheses 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 parantheses 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 plaincg ')' in the jth position and $R[i+1,j]$ corresponds to placing '(' in the jth position, $R[i,j]$ still measures the desired quantity.

This completes the induction and proves the claim. ◀

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.

Let $H := (R^T)R$. Since R is upper triangular with 1's on the main diagonal, $H = (R^T)R$ gives the unique Cholesky decomposition, so H is positive definite. It is easy to see that $H[i,j] = \langle R_i, R_j \rangle$ is the total number of ways to place $i+j$ parantheses consistently with the same number of '(' and ')'. Therefore, $H[i,j] = 0$ if $i+j$ is odd, and if $i+j$ is even (let $l := \frac{i+j}{2}$), $H[i,j]$ is the lth Catalan number $C_l := \frac{1}{l+1}\binom{2l}{l}$. In particular, $H[i,j] = H[i',j']$ for all $i+j = i'+j'$. Such H is called a *Hankel matrix.*

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 $\mathbf{E}[W^i] = m_i$. 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 $H[i,j] := \mathbf{E}[W^{i+j}]$ is positive definite [17]. 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 $\mathbf{E}[W^i] = 0$ if i is odd, $\mathbf{E}[W^i] = C_{\frac{i}{2}}$ if i is even. It is known as the *Wigner semicircle distribution* with radius $R = 2$.

▶ Remark. 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 $\mathbf{E}[W^{2l}] = C_l = \frac{1}{l+1}\binom{2l}{l} \leq 2^l$ while $\mathbf{E}[G^{2l}] = (2l-1)!! \geq 2^{\Omega(l \log l)}$.

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 $x_1, \dots, x_n$ where each x_i is an independent copy of $\frac{W}{N}$ for some large number N to be determined later. Our $\widehat{\mathsf{W}}$ is later defined to be $\widehat{\mathsf{W}}[\alpha,\beta] = \mathbf{E}[x^{\alpha+\beta}] \cdot N^q$ 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 $\mathbf{E}[p(x)^2]$ for any n-variate (possibly non-homogeneous) polynomial p of degree at most k. For a multivarite polynomial $p = \sum_{\alpha \in \mathbb{N}^n_{\leq k}} p_\alpha x^\alpha$, define ℓ_2 norm of p to be $\|p\|_{\ell_2} := \sqrt{\sum_\alpha p_\alpha^2}$. For $0 \leq m \leq n$ and $0 \leq l \leq k$, let $\sigma(m,l) := \inf_p \mathbf{E}[p(x)^2]$ where the infimum is taken over polynomials p such that $\|p\|_{\ell_2} = 1$, $\deg(p) \leq l$, and p depends only on $x_1, \dots, x_m$.

▶ **Lemma 17.** *There exists $N := N(n,k)$ such that $\sigma(m,l) \geq \frac{(1-\frac{m}{2n})}{N^{2l}}$ 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 $\mathbf{E}[p^2] = 1$.

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

$$\mathbf{E}[p^2] = \mathbf{E}[(\sum_{i=0}^{l} p_i x_m^i)^2] = \sum_{0 \leq i,j \leq l} \mathbf{E}[p_i p_j]\,\mathbf{E}[x_m^{i+j}].$$

Let $\Sigma = \mathsf{diag}(1, \frac{1}{N}, \dots, \frac{1}{N^l}) \in \mathbb{R}^{(l+1)\times(l+1)}$. 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 $\{0, \dots, l\}$. Define $R_l \in \mathbb{R}^{(l+1)\times(l+1)}$ similarly from R, and r_t $(0 \leq t \leq l)$ be the tth column of $(R_l)^T$. Note $H_l = (R_l)^T R_l = \sum_{t=0}^{l} r_t r_t^T$. Let $H' = \Sigma H_l \Sigma$ such that $H'[i,j] = \mathbf{E}[x_m^{i+j}]$. Finally, let $P \in \mathbb{R}^{(l+1)\times(l+1)}$ be defined such that $P[i,j] := \mathbf{E}[p_i p_j]$. Then $\mathbf{E}[p^2]$ is equal to

$$\mathbf{Tr}(PH') = \mathbf{Tr}(P\Sigma H_l \Sigma) = \mathbf{Tr}\left(P\Sigma(\sum_{t=0}^{l} r_t r_t^T)\Sigma\right)$$
$$= \sum_{t=0}^{l} \mathbf{E}[(p_t \frac{1}{N^t} + p_{t+1}\frac{(r_t)_{t+1}}{N^{t+1}} + \cdots + p_l \frac{(r_t)_l}{N^l})^2],$$

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

$$q_t := p_t \frac{1}{N^t} + p_{t+1}\frac{(r_t)_{t+1}}{N^{t+1}} + \cdots + p_l \frac{(r_t)_l}{N^l}.$$

Since p_i 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,

$$\|q_t\|_{\ell_2} \geq \frac{1}{N^t}\Bigg(\|p_t\|_{\ell_2} - \big(\|p_{t+1}\|_{\ell_2} \frac{(r_t)_{t+1}}{N} + \cdots + \|p_l\|_{\ell_2} \frac{(r_t)_l}{N^{l-t}}\big)\Bigg) \geq \frac{1}{N^t}\Bigg(\|p_t\|_{\ell_2} - \frac{k2^k}{N}\Bigg),$$

and

$$\|q_t\|_{\ell_2}^2 \geq \frac{1}{N^{2t}}\Bigg(\|p_t\|_{\ell_2}^2 - \frac{2k2^k}{N}\Bigg).$$

Finally,

$$\begin{aligned}
\mathbf{E}[p^2] &= \sum_{t=0}^{l} \mathbf{E}[q_t^2] \\
&\geq \sum_{t=0}^{l} \sigma(m-1, l-t) \cdot \|q_t\|_{\ell_2}^2 \\
&\geq \sum_{t=0}^{l} \sigma(m-1, l-t) \cdot \frac{1}{N^{2t}}\Bigg(\|p_t\|_{\ell_2}^2 - \frac{2k2^k}{N}\Bigg) \\
&\geq \sum_{t=0}^{l} \frac{(1-\frac{m-1}{2n})}{N^{2l-2t}} \cdot \frac{1}{N^{2t}} \cdot \Bigg(\|p_t\|_{\ell_2}^2 - \frac{2k2^k}{N}\Bigg) \\
&= \frac{(1-\frac{m-1}{2n})}{N^{2l}} \cdot \sum_{t=0}^{l} \Bigg(\|p_t\|_{\ell_2}^2 - \frac{2k2^k}{N}\Bigg) \\
&\geq \frac{(1-\frac{m-1}{2n})}{N^{2l}} \cdot \big(1 - \frac{2K^2 2^k}{N}\big).
\end{aligned}$$

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

Construction of $\widehat{\mathsf{W}}$

We now prove Theorem 15. Given n and q, let $k = \frac{q}{2}$, and consider random variables $x_1, \ldots, x_n$ above. Let $\widehat{\mathsf{W}} \in \mathbb{R}^{\mathbb{N}_k^n \times \mathbb{N}_k^n}$ be such that for any $\alpha, \beta \in \mathbb{N}_k^n$, $\widehat{\mathsf{W}}[\alpha, \beta] = \mathbf{E}[x^{\alpha+\beta}] \cdot N^{2k}$. By definition, $\widehat{\mathsf{W}}$ is degree-q SoS symmetric. Since each entry of $\widehat{\mathsf{W}}$ corresponds to a monomial of degree exactly q and each x_i is drawn independently from the Wigner semicircle distribution, each entry of $\widehat{\mathsf{W}}$ is at most the $\frac{q}{2}$th Catalan number $C_{\frac{q}{2}} \leq 2^q$. For any unit vector $p = (p_S)_{S \in \mathbb{N}_k^n} \in \mathbb{R}^{\mathbb{N}_k^n}$, Lemma 17 shows $p^T \widehat{\mathsf{W}} p = \mathbf{E}[p^2] \cdot N^{2k} \geq \frac{1}{2}$ where p also represents a degree-k homogeneous polynomial $p(x_1, \ldots, x_n) = \sum_{\alpha \in \binom{[n]}{k}} p_\alpha x^\alpha$. Therefore, the minimum eigenvalue of $\widehat{\mathsf{W}}$ is at least $\frac{1}{2}$.

Due to space constraints, we defer the final construction of the moment matrix to the appendix (see Section A).

References

1 Boaz Barak, Fernando G. S. L. 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.

2 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.

3 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.

4 Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014.

5 Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Weak decoupling, polynomial folds, and approximate optimization over the sphere. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:185, 2016. URL: `https://eccc.weizmann.ac.il/report/2016/185/`.

6 Vijay Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. Certifying random polynomials over the unit sphere via sum of squares hierarchy. *arXiv preprint arXiv:1605.00903*, 2016.

7 Fernando G. S. L. 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.

8 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.

9 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.

10 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.

11 Samuel B. Hopkins, Pravesh K. Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. Personal communication, 2017.

12 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.

13 Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proceedings of the 49th ACM Symposium on Theory of Computing*, 2017. To appear.

14 Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009.

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

16 Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random CSPs below the spectral threshold. In *Proceedings of the 49th ACM Symposium on Theory of Computing*, 2017. To appear.

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

18 R. Tomioka and T. Suzuki. Spectral norm of random tensors. *ArXiv e-prints*, July 2014. `arXiv:1407.1870`.

A Constructing the Moment Matrix Realizing the Lower Bound

For even d, let $\mathcal{A} \in \mathbb{R}^{[n]^q}$ be a q-tensor with i.i.d. ±1 entries and let $A \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$ be the matrix flattening of $\mathcal{A}$, i.e., $A[I, J] = \mathcal{A}[I \oplus J]$ (recall that $\oplus$ denotes tuple concatenation). Also let $f(x) := \mathcal{A}(x) = \langle \mathcal{A}, x^{\otimes q} \rangle$. Our lower bound on $f_{\max}$ by is proved by constructing a moment matrix $\mathsf{M} \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$ that satisfies

- $\mathbf{Tr}(\mathsf{M}) = 1$.
- $\mathsf{M} \succeq 0$.
- M is SoS-symmetric.
- $\langle A, \mathsf{M}\rangle \;\geq\; 2^{-O(q)} \cdot n^{q/4}/q^{q/4}$,

where $A \in \mathbb{R}^{[n]^{q/2}\times[n]^{q/2}}$ is any matrix representation of f (SoS-symmetry of M ensures $\langle A, \mathsf{M}\rangle$ does not depend on the choice of A).

Let A be the SoS-symmetric matrix such that for any $I = (i_1, \ldots, i_{q/2})$ and $J = (j_1, \ldots, j_{q/2})$,

$$\mathsf{A}[I, J] = \begin{cases} \frac{f_{\alpha(I)+\alpha(J)}}{q!}, & \text{if } i_1, \ldots, i_{q/2}, j_1, \ldots, j_{q/2} \text{ are all distinct.} \\ 0 & \text{otherwise.} \end{cases}$$

We bound $\|\mathsf{A}\|_2$ in two steps. Let $\widehat{\mathsf{A}}_Q \in \mathbb{R}^{\mathbb{N}^n_{q/2}\times\mathbb{N}^n_{q/2}}$ be the *quotient matrix* of A defined by

$$\widehat{\mathsf{A}}_Q[\beta, \gamma] := \mathsf{A}[I, J] \cdot \sqrt{|\mathcal{O}(\beta)| \cdot |\mathcal{O}(\gamma)|},$$

where $I, J \in [n]^{q/2}$ are such that $\beta = \alpha(I), \gamma = \alpha(J)$.

▶ **Lemma 18.** *With high probability,* $\|\widehat{\mathsf{A}}_Q\|_2 \leq\; 2^{O(q)} \cdot \frac{n^{q/4}}{q^{q/4}}$.

Proof. Consider any $y \in \mathbb{R}^{\mathbb{N}^n_{q/2}}$ s.t. $\|y\| = 1$. Since

$$\begin{aligned} y^T \cdot \widehat{\mathsf{A}}_Q \cdot y &= \sum_{\beta+\gamma \leq \mathbf{1}} \widehat{A}_Q[\beta, \gamma] \cdot y_\beta \cdot y_\gamma \\ &= \sum_{\beta+\gamma \leq \mathbf{1}} y_\beta \cdot y_\gamma \sum_{\substack{\alpha(I)+\alpha(J) \\ =\beta+\gamma}} A[I, J] \cdot \frac{\sqrt{|\mathcal{O}(\beta)||\mathcal{O}(\gamma)|}}{|\mathcal{O}(\beta+\gamma)|} \\ &= \sum_{I,J\in[n]^{q/2}} A[I, J] \sum_{\substack{\beta+\gamma \leq \mathbf{1} \\ \beta+\gamma= \\ \alpha(I)+\alpha(J)}} \frac{\sqrt{|\mathcal{O}(\beta)||\mathcal{O}(\gamma)|}}{|\mathcal{O}(\beta+\gamma)|} \cdot y_\beta \cdot y_\gamma \end{aligned}$$

So $y^T \cdot \widehat{\mathsf{A}}_Q \cdot y$ is a sum of independent random variables

$$\sum_{I,J\in[n]^q} A[I, J] \cdot c_{I,J}$$

where each $A[I, J]$ is independently sampled from the Rademacher distribution and

$$c_{I,J} := \sum_{\substack{\beta+\gamma \leq \mathbf{1} \\ \beta+\gamma= \\ \alpha(I)+\alpha(J)}} \frac{\sqrt{|\mathcal{O}(\beta)||\mathcal{O}(\gamma)|}}{|\mathcal{O}(\beta+\gamma)|} \cdot y_\beta \cdot y_\gamma \,.$$

Fix any $I, J \in [n]^{q/2}$ and let $\alpha := \alpha(I) + \alpha(J)$. By Cauchy-Schwarz,

$$c^2_{I,J} \;\leq\; \left(\sum_{\beta+\gamma=\alpha} \frac{|\mathcal{O}(\beta)||\mathcal{O}(\gamma)|}{|\mathcal{O}(\alpha)|^2}\right) \cdot \Big(\sum_{\beta+\gamma=\alpha} y^2_\beta \cdot y^2_\gamma\Big) \;\leq\; \frac{2^{O(q)}}{|\mathcal{O}(\alpha)|} \cdot \sum_{\beta+\gamma=\alpha} y^2_\beta \cdot y^2_\gamma \;=:\; c^2_\alpha \,, \tag{A.1}$$

since there are at most $2^{O(q)}$ choices of β and γ with $\beta + \gamma = \alpha$, and $|\mathcal{O}(\beta)| \cdot |\mathcal{O}(\gamma)| \le |\mathcal{O}(\alpha)|$. Therefore, $y^T \cdot \widehat{\mathsf{A}}_Q \cdot y$ is the sum of independent random variables that are centred and always lie in the interval $[-1, +1]$. Furthermore, by Eq. (A.1), the total variance is

$$\sum_{I,J \in [n]^{q/2}} c^2_{I,J} \;\le\; \sum_{\alpha \in \mathbb{N}^n_q} c^2_\alpha \cdot |\mathcal{O}(\alpha)| \;\le\; 2^{O(q)} \cdot \sum_{\beta,\gamma \in \mathbb{N}^n_{q/2}} y^2_\beta \cdot y^2_\gamma \;=\; 2^{O(q)} \cdot \Big(\sum_{\beta \in \mathbb{N}^n_{q/2}} y^2_\beta \Big)^2 \;=\; 2^{O(q)}$$

The claim then follows from combining standard concentration bounds with a union bound over a sufficiently fine net of the unit sphere in $|\mathbb{N}^n_{q/2}| \le 2^{O(q)} \cdot \frac{n^{q/2}}{q^{q/2}}$ dimensions. ◀

▶ **Lemma 19.** *For any SoS-symmetric* $\mathsf{A} \in \mathbb{R}^{[n]^{q/2} \times [n]^{q/2}}$, $\|\mathsf{A}\|_2 \le \left\|\widehat{\mathsf{A}}_Q\right\|_2$.

Proof. For any $u, v \in \mathbb{R}^{[n]^{q/2}}$ *s.t.* $\|u\| = \|v\| = 1$, we have

$$\begin{aligned}
& u^T \mathsf{A} v \\
&= \sum_{I,J \in [n]^{q/2}} \mathsf{A}[I,J] u_I v_J \\
&= \sum_{I,J \in [n]^{q/2}} \frac{\widehat{\mathsf{A}}_Q[\alpha(I), \alpha(J)]}{\sqrt{|\mathcal{O}(I)|\,|\mathcal{O}(J)|}} \cdot u_I v_J \\
&= \sum_{\alpha,\beta \in \mathbb{N}^n_{q/2}} \frac{\mathsf{A}[\alpha,\beta]}{\sqrt{|\mathcal{O}(\alpha)|\,|\mathcal{O}(\beta)|}} \langle u|_{\mathcal{O}(\alpha)}, \mathbb{1}\rangle \langle v|_{\mathcal{O}(\beta)}, \mathbb{1}\rangle \\
&= a^T \widehat{\mathsf{A}}_Q\, b \qquad \text{where } a_\alpha := \frac{\langle u|_{\mathcal{O}(\alpha)}, \mathbb{1}\rangle}{\sqrt{|\mathcal{O}(\alpha)|}},\ b_\alpha := \frac{\langle v|_{\mathcal{O}(\alpha)}, \mathbb{1}\rangle}{\sqrt{|\mathcal{O}(\alpha)|}} \\
&\le \left\|\widehat{\mathsf{A}}_Q\right\|_2 \|a\| \cdot \|b\| \\
&= \left\|\widehat{\mathsf{A}}_Q\right\|_2 \sqrt{\sum_{\alpha \in \mathbb{N}^n_{q/2}} \frac{\langle u|_{\mathcal{O}(\alpha)}, \mathbb{1}\rangle^2}{|\mathcal{O}(\alpha)|}} \sqrt{\sum_{\alpha \in \mathbb{N}^n_{q/2}} \frac{\langle v|_{\mathcal{O}(\alpha)}, \mathbb{1}\rangle^2}{|\mathcal{O}(\alpha)|}} \\
&\le \left\|\widehat{\mathsf{A}}_Q\right\|_2 \sqrt{\sum_{\alpha \in \mathbb{N}^n_{q/2}} \|u|_{\mathcal{O}(\alpha)}\|^2} \sqrt{\sum_{\alpha \in \mathbb{N}^n_{q/2}} \|u|_{\mathcal{O}(\alpha)}\|^2} \qquad \text{(by Cauchy-Schwarz)} \\
&\le \left\|\widehat{\mathsf{A}}_Q\right\|_2 \|u\| \cdot \|v\| = \left\|\widehat{\mathsf{A}}_Q\right\|_2 .
\end{aligned}$$

◀

The above two lemmas imply that $\|\mathsf{A}\|_2 \le \|\widehat{\mathsf{A}}_Q\|_2 \le 2^{O(q)} \cdot \frac{n^{q/4}}{q^{q/4}}$. Our moment matrix M is defined by

$$\mathsf{M} := \frac{1}{c_1} \left(\frac{1}{c_2} \cdot \frac{q^{3q/4}}{n^{3q/4}} \mathsf{A} + \frac{\mathsf{W}}{n^{q/2}} \right),$$

where W is the direct extension of $\widehat{\mathsf{W}}$ constructed in Theorem 15 – $\mathsf{W}[I,J] := \widehat{\mathsf{W}}[\alpha(I), \alpha(J)]$ for all $I, J \in [n]^{q/2}$, and $c_1, c_2 = 2^{\Theta(q)}$ that will be determined later.

We first consider the trace of M. The trace of A is 0 by design, and the trace of W is $n^{q/2} \cdot 2^{O(q)}$. Therefore, the trace of M can be made 1 by setting c_1 appropriately. Since both A and W are SoS-symmetric, so is M. Since $\mathbf{E}[\mathsf{W}, A] = 0$ and for each $I, J \in [n]^{q/2}$ with $i_1, \ldots, i_{q/2}, j_1, \ldots, j_{q/2}$ all distinct we have $\mathbf{E}[\mathsf{A}[I,J] A[I,J]] = \frac{1}{q!}$, with high probability

$$\langle A, \mathsf{M} \rangle = \frac{1}{c_1} \cdot \langle A, \left(\frac{1}{c_2} \cdot \frac{q^{3q/4}}{n^{3q/4}} \mathsf{A} + \frac{\mathsf{W}}{n^{q/2}} \right) \rangle \ge 2^{O(-q)} \cdot \frac{q^{3q/4}}{n^{3q/4}} \cdot \frac{n^q}{q^q} = 2^{O(-q)} \cdot \frac{n^{q/4}}{q^{q/4}}.$$

It finally remains to show that M is positive semidefinite. Take an arbitrary vector $v \in \mathbb{R}^{[n]^{q/2}}$, and let

$$p = \sum_{\alpha \in \mathbb{N}^n_{q/2}} x^\alpha p_\alpha = \sum_{\alpha \in \mathbb{N}^n_{q/2}} x^\alpha \cdot \left(\sum_{I \in [n]^{q/2} : \alpha(I) = \alpha} v_I \right)$$

be the associated polynomial. If $p = 0$, SoS-symmetry of M ensures $v\mathsf{M}v^T = 0$. Normalize v so that $\|p\|_{\ell_2} = 1$. First, consider another vector $v_m \in [n]^{q/2}$ such that

$$(v_m)_I = \begin{cases} \frac{p^{\alpha(I)}}{(q/2)!}, & \text{if } i_1, \ldots, i_{q/2} \text{ are all distinct.} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\|v_m\|_2^2 \leq \sum_{\alpha \in \mathbb{N}^n_{q/2}} p_\alpha^2 / (q/2)! = \frac{1}{(q/2)!},$$

so $\|v_m\|_2 \leq \frac{2^{O(q)}}{q^{q/4}}$. Since A is SoS-symmetric, has the minimum eigenvalue at least $-2^{O(q)} \cdot \frac{n^{q/4}}{q^{q/4}}$, and has nonzero entries only on the rows and columns $(i_1, \ldots, i_{q/2})$ with all different entries,

$$v^T \mathsf{A} v = (v_m)^T \mathsf{A} (v_m) \geq 2^{-O(q)} \cdot \frac{n^{q/4}}{q^{3q/4}}.$$

We finally compute $v^T \mathsf{W} v$. Let $v_w \in [n]^{q/2}$ be the vector where for each $\alpha \in \mathbb{N}^n_{q/2}$, we choose one $I \in [n]^{q/2}$ arbitrarily and set $(v_w)_I = p_\alpha$ (all other $(v_w)_I$'s are 0). By SoS-symmetry of W,

$$v^T \mathsf{W} v = (v_w)^T \mathsf{W} (v_w) = p^T \widehat{\mathsf{W}} p \geq \frac{1}{2},$$

by Theorem 15. Therefore,

$$v^T \cdot \mathsf{M} \cdot v \quad = \quad \frac{1}{c_1} \cdot v^T \cdot \left(\frac{1}{c_2} \cdot \frac{q^{3q/4}}{n^{3q/4}} \mathsf{A} + \frac{\mathsf{W}}{n^{q/2}} \right) \cdot v \quad \geq \quad \frac{1}{c_1} \cdot \left(\frac{1}{c_2} \cdot 2^{-O(q)} \cdot \frac{n^{q/4}}{q^{3q/4}} \cdot \frac{q^{3q/4}}{n^{3q/4}} + \frac{1}{2} \cdot \frac{1}{n^{q/2}} \right) \quad \geq \quad 0,$$

by taking $c_2 = 2^{\Theta(q)}$. So M is positive semidefinite, and this finishes the proof of the lower bound in Theorem 1.

Thus we obtain,

▶ **Theorem 20** (Lower bound in Theorem 1). *For even $q \leq n$, let $\mathcal{A} \in \mathbb{R}^{[n]^q}$ be a q-tensor with i.i.d. ± 1 entries. Then with probability $1 - n^{\Omega(1)}$,*

$$\frac{\mathit{SoS}_q(\mathcal{A}(x))}{\mathcal{A}_{\max}} \quad \geq \quad \left(\frac{\Omega(n)}{q} \right)^{q/4 - 1/2}.$$

As a side note, observe that by applying Lemma 19 and the proof of Lemma 18 to the SoS-symmetric matrix representation of $f(x) = \mathcal{A}(x)$ (instead of A), we obtain a stronger SoS upper bound (by polylog factors) for the special case of $d = q$:

▶ **Theorem 21** (Upper bound in Theorem 1). *For even $q \leq n$, let $\mathcal{A} \in \mathbb{R}^{[n]^q}$ be a q-tensor with i.i.d. ± 1 entries. Then with probability $1 - n^{\Omega(1)}$,*

$$\frac{\mathit{SoS}_q(\mathcal{A}(x))}{\mathcal{A}_{\max}} \quad \leq \quad \left(\frac{O(n)}{q} \right)^{q/4 - 1/2}.$$