Tensor cumulants for statistical inference on invariant distributions

Dmitriy Kunisky*[1], Cristopher Moore†[2], and Alexander S. Wein‡[3]

[1]Department of Computer Science, Yale University
[2]Santa Fe Institute
[3]Department of Mathematics, UC Davis

April 29, 2024

Abstract

Many problems in high-dimensional statistics appear to have a *statistical-computational gap*: a range of values of the signal-to-noise ratio where inference is information-theoretically possible, but (conjecturally) computationally intractable. A canonical such problem is Tensor PCA, where we observe a tensor Y consisting of a rank-one signal plus Gaussian noise. Multiple lines of work suggest that Tensor PCA becomes computationally hard at a critical value of the signal's magnitude. In particular, below this transition, no low-degree polynomial algorithm can detect the signal with high probability; conversely, various spectral algorithms are known to succeed above this transition. We unify and extend this work by considering *tensor networks*, orthogonally invariant polynomials where multiple copies of Y are "contracted" to produce scalars, vectors, matrices, or other tensors. We define a new set of objects, *tensor cumulants*, which provide an explicit, near-orthogonal basis for invariant polynomials of a given degree. This basis lets us unify and strengthen previous results on low-degree hardness, giving a combinatorial explanation of the hardness transition and of a continuum of subexponential-time algorithms that work below it, and proving tight lower bounds against low-degree polynomials for recovering rather than just detecting the signal. It also lets us analyze a new problem of distinguishing between different tensor ensembles, such as Wigner and Wishart tensors, establishing a sharp computational threshold and giving evidence of a new statistical-computational gap in the Central Limit Theorem for random tensors. Finally, we believe these cumulants are valuable mathematical objects in their own right: they generalize the free cumulants of free probability theory from matrices to tensors, and share many of their properties, including additivity under additive free convolution.

*Email: dmitriy.kunisky@yale.edu. Partially supported by ONR Award N00014-20-1-2335 and a Simons Investigator Award to Daniel Spielman.

†Email: moore@santafe.edu. Partially supported by the National Science Foundation through grant BIGDATA-1838251.

‡Email: aswein@ucdavis.edu. Partially supported by an Alfred P. Sloan Research Fellowship and NSF CAREER Award CCF-2338091.

Contents

1 Introduction **1**
1.1 Main Results . . . 3
1.2 Main Proof Technique: Tensorial Finite Free Cumulants . . . 7
1.3 Related Work . . . 9

2 Notation **11**

3 Preliminaries **11**
3.1 Closed Tensor Networks as Invariant Polynomials . . . 11
3.2 Open Tensor Networks as Equivariant Polynomials . . . 13
3.3 Invariant Tensors, Brauer Space, and the Weingarten Function . . . 15
3.4 Low-Degree Polynomial Algorithms . . . 16

4 Tensorial Finite Free Cumulants and Invariant Bases **19**
4.1 Frobenius Pairs . . . 19
4.2 Distinct-Index Graph Moments . . . 20
4.3 Finite Free Cumulants and Additive Free Convolution . . . 23
4.4 Graph Moment Expansion of Finite Free Cumulants . . . 26
4.5 Inner Products and Approximate Orthogonality . . . 29
4.6 Advantage Bound for Invariant Detection Problems . . . 31
4.7 Bases for Vector-Valued Equivariant Polynomials . . . 33

5 Application 1: Tensor PCA **36**
5.1 Warmup: Detection with Individual Graph Moments . . . 36
5.2 Detection with General Low-Degree Polynomials: Proof of Theorem 1.6 . . . 40
5.3 Reconstruction with Low-Degree Polynomials . . . 42

6 Application 2: Distinguishing Wigner from Wishart Tensors **48**

Acknowledgments **53**

References **53**

A Characterizations of Invariants **60**

B Properties of Wigner Tensors **62**
B.1 Basic Properties . . . 62
B.2 Hardness of Computing Moments . . . 63

C Conditioning of Weingarten Matrices **68**

D Combinatorial Bounds **71**

1 Introduction

We will study statistical problems formulated over *tensors*. Here a tensor T is a p-dimensional table of real numbers, indexed as $T_{i_1,i_2,\ldots,i_p}$ with $i_a \in \{1,\ldots,n\} =: [n]$. We call p the *arity* of T, say that T is a *p-ary tensor*, and call n the *dimension* of T. A 1-ary tensor is an n-dimensional vector. A 2-ary tensor can be viewed as an $n \times n$ matrix or as an n^2-dimensional vector, and so on. A 0-ary tensor is a scalar, i.e., a real number. We say that T is *symmetric* if $T_{i_1,\ldots,i_p} = T_{\sigma(i_1),\ldots,\sigma(i_p)}$ for any permutation $\sigma \in S_p$. In this case, each entry of T is specified by a multiset in $[n]$ of size p. The vector space of symmetric tensors of dimension n and arity p is denoted $\mathrm{Sym}^p(\mathbb{R}^n)$.

Many statistical problems of broad interest are modeled with tensor observations, representing collections of estimates of degree p moments of high-dimensional random variables or p-way interaction data such as hypergraphs and generalizations thereof. Often, whether in *hypothesis testing* (distinguishing two distributions over tensors) or in *estimation* (recovering some parameter from a noisy tensor observation) problems, the distributions of tensors involved are reasonably assumed to be *orthogonally invariant*, that is, invariant to simultaneous changes of basis of all tensor "axes."

Definition 1.1 (Change of basis)**.** *For T a tensor of dimension n and Q a (usually but not necessarily orthogonal) matrix, we define the tensor $Q \cdot T$ to have entries*

$$(Q \cdot T)_{j_1,\ldots,j_p} = \sum_{i_1,\ldots,i_p} T_{i_1,\ldots,i_p} \prod_{t=1}^{p} Q_{i_t,j_t}. \tag{1}$$

Alternatively, we may view $Q \cdot T = (Q^\top)^{\otimes p} T$ where T is viewed as a vector in $\mathbb{R}^{n^p}$. This makes the action of the orthogonal group $\mathcal{O}(n)$ or the general linear group $\mathrm{GL}(n)$ on $(\mathbb{R}^n)^{\otimes p}$ a right *group action: $R \cdot (Q \cdot T) = (QR) \cdot T$, which preserves symmetry and thus is also an action on $\mathrm{Sym}^p(\mathbb{R}^n)$.*

For T a matrix one may check that $Q \cdot T = Q^\top T Q$ coincides with the usual change of basis.

Definition 1.2 (Invariance)**.** *A function $f : \mathrm{Sym}^p(\mathbb{R}^n) \to \mathbb{R}$ is* (orthogonally) invariant *if $f(Q \cdot T) = f(T)$ for all $T \in \mathrm{Sym}^p(\mathbb{R}^n)$ and all $Q \in \mathcal{O}(n)$. The law of a random T is* (orthogonally) invariant *if $Q \cdot T$ has the same law for all orthogonal $Q \in \mathcal{O}(n)$.*

Tensor networks are a graphical notation that extends linear algebra to tensors, generalizing operations like the inner product of vectors, matrix products, and traces. This gives a powerful language for discussing quantities that would be quite tricky to express in conventional notation (as we already see above for changes of basis). We will discuss general tensor networks later in Section 3.1, but we mention now that, aside from changes of basis, the following is an important class of computations that tensor networks can express.

Definition 1.3 (Graph moments)**.** *Let $T \in \mathrm{Sym}^p(\mathbb{R}^n)$ and $G = (V, E)$ a p-regular graph (not necessarily simple). Then, we define the G-*moment *of T as*

$$m_G(T) = \sum_{i \in [n]^E} \prod_{v \in V} T_{i(\partial v)}\,. \tag{2}$$

where $i(\partial v)$ denotes the multiset of indices $i(e)$ associated with the edges e incident to v. If G is the empty graph with $V = E = \varnothing$, then $m_G(T) = 1$.

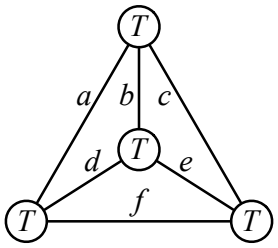


Figure 1: The graph moment $m_G(T)$ where T is a symmetric 3-ary tensor and $G = K_4$. Summing over all six indices, one on each edge, contracts the graph and yields the scalar (3).

When T is symmetric—as it always is here—we do not need to order the edges incident to a vertex.[1]

Example 1.4. *Let $G = K_4$ be the complete graph on four vertices, as shown in Figure 1. Then*

$$m_G(T) = \sum_{a,b,c,d,e,f \in [n]} T_{abc}\, T_{adf}\, T_{bde}\, T_{cef}\,, \tag{3}$$

Note that this is a homogeneous polynomial, of degree 4, of T's entries. In general, $m_G(T)$ is a homogeneous polynomial, of degree $|V(G)|$, of T's entries.

Linear combinations of the tensor networks $m_G(T)$ are a natural class of scalar functions of a tensor. In particular, as we will show in Proposition 3.13, they are themselves invariant: $m_G(Q \cdot T) = m_G(T)$ for all $Q \in \mathcal{O}(n)$. Moreover, by classical invariant theory, they span *all* invariant polynomials. As we will discuss in Section 3.1, these linear combinations should therefore be thought of as the correct "space of spectral algorithms" for computing scalar statistics from tensors, analogous to the linear spectral statistics $T \mapsto \mathrm{tr}(f(T))$ of a matrix T.

We wish to draw a connection between these algebraic facts and the substantial body of literature that has developed recently around using *low-degree polynomial algorithms* for tasks like hypothesis testing and estimation. We leave a detailed discussion to Section 3.4, but simple instances of this approach are to perform hypothesis testing between two distributions of tensors by computing and thresholding a low-degree polynomial of the observed tensor [HS17, HKP$^+$17, Hop18], or estimation by computing an estimator that consists of a vector of low-degree polynomials of the observation [SW22]. We view the degree of a polynomial as a proxy for the runtime of the associated algorithm (as, assuming we naively evaluate polynomials without taking advantage of any special structure, a degree D polynomial in n variables takes time $O(n^D)$ to evaluate), and so the degree of polynomial required for testing or estimation is taken as a measure of the problem's complexity.

The goal of this paper is to explore the consequences of the following general observation (formalized for low-degree algorithms in our Proposition 3.22): for hypothesis testing tasks over invariant tensor distributions, the best low-degree polynomials are invariant themselves, and therefore are linear combinations of graph moments $m_G(T)$. Similarly, the best low-degree estimators are linear combinations of graph moments with a single "open" edge, which evaluate to vectors rather than scalars (Proposition 3.24). In addition to giving us a set of tools for reasoning graphically about low-degree algorithms, this dramatically restricts the space of such algorithms for invariant

[1] If we considered tensors with complex entries the edges would need to be directed, since the complex inner product has $u^* v \neq v^* u$ in general. For us, T's entries will always be real, so undirected edges suffice.

problems, and therefore can also reduce the analytical difficulties of proving computational lower bounds. Indeed, the dimension of the space of low-degree invariant polynomials, corresponding for degree d to the number of p-regular multigraphs G on d vertices, does not depend on the tensor dimension n at all! Moreover, the space spanned by the $m_G(T)$ turns out to have a rich algebraic structure with important ramifications for the computational complexity of statistical problems. For example, we will see that its dimension (the number of p-regular multigraphs of a given size) has a direct connection to the tradeoff between signal strength and subexponential runtime of algorithms for hypothesis testing and estimation.

We will develop a general theory about the spaces of invariant polynomials (and *equivariant* vector-valued polynomials, to be used as estimators), extending classical invariant theory to be directly useful for the analysis of low-degree algorithms. Surprisingly, a central role will be played by a notion of *finite free cumulants* for tensors. These generalize aspects of the theory of free probability for random matrices, and yield an explicit near-orthogonal basis for the space of invariant polynomials. Using these tools, we will give new results for two examples of invariant statistical problems over tensors.

First, we will consider the well-studied problem of *tensor PCA (principal component analysis)*, recovering and unifying previous results on hypothesis testing and giving a new and tight analysis of estimation. Second, we will study a variation on the newer problem of *hypothesis testing between Wigner and Wishart tensors*, adapted to our focus on invariant distributions. Through one lens this is a relative of the task of *tensor decomposition*, and through another it is a variation on the question of distinguishing Erdős-Rényi from geometric random hypergraphs (a hypergraph version of the problem treated for matrices by [BDER16, BBN20, LMSY22, BB24]). Through yet a third it is a question about a "computational central limit theorem," a class of questions to which our results will apply more generally (a closely related tensor problem was studied by [Mik20], and similar matrix problems by [BDER16, BBH21]).

Below we first focus on these concrete applications, and then sketch some of the technical ideas before giving full details in Section 4.

1.1 Main Results

Tensor PCA Our first subject will be a model of random tensors built as follows:

$$Y = \lambda v^{\otimes p} + W\,. \tag{4}$$

Here $\lambda \geq 0$ is a signal-to-noise ratio, the *spike* v is chosen from some prior distribution on $\mathbb{R}^n$, and W is a tensor of random noise. This model is known as *tensor PCA* or the *spiked tensor model* [MR14].

In this model, hypothesis testing or *detection* entails distinguishing the spiked model above from a null model where $Y = W$ consists only of noise. Estimation or *reconstruction* entails producing a vector $\widehat{v} = \widehat{v}(Y)$ that approximates v. A reasonable request is for $\widehat{v}$ to be more correlated with v than a random guess, i.e., if $\|v\|^2, \|\widehat{v}\|^2 \approx n$, then $|\langle \widehat{v}, v\rangle| \geq \epsilon n$ for some constant $\epsilon > 0$.

Recall that we are interested in the case where the distributions involved above are orthogonally invariant. To adhere to this setting, we will assume that v is chosen uniformly from the sphere $\mathbb{S}^{n-1}(\sqrt{n}) = \{v \in \mathbb{R}^n : \|v\|^2 = n\}$ and that W is a *Wigner random tensor* with Gaussian entries. To keep Y symmetric, we will symmetrize W as follows.

Definition 1.5. *For $p \geq 1$ and $\sigma^2 > 0$, we write* $\mathrm{Wig}(p, n, \sigma^2)$ *for the law of the symmetric tensor* $W \in \mathrm{Sym}^p(\mathbb{R}^n)$ *that is given entrywise by*

$$W_{i_1,\ldots,i_p} = \frac{1}{\sqrt{p!}} \sum_{\pi \in S_p} G_{i_{\pi(1)},\ldots,i_{\pi(p)}} \tag{5}$$

where $G \in (\mathbb{R}^n)^{\otimes p}$ is an asymmetric tensor whose entries are i.i.d. as $G_{i_1,\ldots,i_p} \sim \mathcal{N}(0, \sigma^2)$. We write Wig *for* $\mathrm{Wig}(p, n, 1)$ *when these parameters are clear.*

This is the natural tensor analog of the Gaussian orthogonal ensemble (GOE), which is the matrix case $p = 2$. As there, the law $\mathrm{Wig}(p, n, \sigma^2)$ is orthogonally invariant. Up to symmetry, the entries are independent but a few have larger variance, just as in the GOE entries on the diagonal have twice the variance of the rest. See Appendix B for further details.

Using the tools we develop, we give a new proof of the following result on low-degree polynomial algorithms. We state this result informally and defer to later the precise notion of "success" of low-degree polynomials for distinguishing two distributions; see Definition 3.18 and Remark 3.19. We call $\mathbb{Q} = \mathrm{Wig}$ the law of the pure noise model, and $\mathbb{P}$ the law of $Y = \lambda v^{\otimes p} + W$ with $v \sim \mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))$ and $W \sim \mathrm{Wig}$ independently.

Theorem 1.6 (Tensor PCA detection; informal; Theorem 3.3 of [KWB22])**.** *Let $D = D(n) \in \mathbb{N}$ have $D \leq \sqrt{n/2p^2}$. There are constants $a_p, b_p > 0$ such that:*

1. *If $\lambda \leq a_p n^{-p/4} D^{-(p-2)/4}$, then polynomials of degree at most D cannot distinguish $\mathbb{P}$ from $\mathbb{Q}$.*
2. *If $\lambda \geq b_p n^{-p/4} D^{-(p-2)/4}$ and $D = \omega(1)$, then polynomials of degree at most D can distinguish $\mathbb{P}$ from $\mathbb{Q}$.*

When $p = 2$, there is no dependence on D in the thresholds for λ, and the result suggests that, on the scale $\lambda = \Theta(n^{-1/2})$, the problem transitions from being tractable in polynomial time (which is indeed achieved by various algorithms) to requiring at least time roughly $\exp(\Omega(\sqrt{n}))$. When $p \geq 3$, the dependence on D means that there are regimes of λ where the problem is tractable but only in *subexponential time* $\exp(O(n^\delta))$ for various $\delta \in (0, 1)$. In fact, these scalings for low-degree upper and lower bounds are known (see again [KWB22], which sharpened the results of [HKP$^+$17]) to hold for $D \sim n^\delta$ for *any* $\delta \in (0, 1)$. It is a technical limitation of our results that they are restricted to $D \lesssim \sqrt{n}$, as it is only in this regime that the basis of invariants we propose is approximately orthogonal.

One advantage of our analysis is that it gives an intuitive explanation for the latter subexponential scaling phenomenon. In particular, the factor of $D^{\frac{p-2}{4}}$ is directly related to the number of non-isomorphic p-regular multigraphs on D vertices, which (see Proposition 5.9) is of the order $D^{\frac{p-2}{2}D}$ up to smaller multiplicative factors; the analysis turns out to indicate that this quantity must be compared with λ^{2D}, leading to the factor of $D^{\frac{p-2}{2}D \cdot \frac{1}{2D}} = D^{\frac{p-2}{4}}$. Per our remarks above, this is also the dimension of the space of degree D invariant polynomials of a tensor.

We also give a new result for the related reconstruction problem. Since reconstruction is a more difficult task than detection, it is to be expected that degree D polynomials fail to reconstruct in the same regime of parameters that they fail to detect per Theorem 1.6. However, actually proving that low-degree polynomials fail to reconstruct—in a precise sense deferred to Section 3.4—does not follow from Theorem 1.6, and generally such reconstruction lower bounds are more difficult to

prove than their detection counterparts. The first tools for low-degree reconstruction lower bounds appeared in [SW22], and this approach can be applied to any additive Gaussian model, which includes tensor PCA. However, to use this machinery, one needs to bound certain recursively-defined quantities, and existing analyses tend to be loose by factors involving D. That is to say, we expect the existing approach of [SW22] would not obtain the precise dependence between λ and D in Theorem 1.6 (likely the exponent on D would be wrong). By using our new machinery in conjunction with the approach of [SW22], we manage to prove low-degree hardness for reconstruction with the correct relation between λ and D. To our knowledge, this is the first low-degree reconstruction result (in any model) that achieves the "correct" exponent on D.

Theorem 1.7 (Lower bound for tensor PCA reconstruction; informal)**.** *Let $D = D(n) \in \mathbb{N}$ have $D \le \sqrt{n/2p^2}$. For all odd $p \ge 3$, there is a constant $c_p > 0$ such that if $\lambda \le c_p n^{-p/4} D^{-(p-2)/4}$, then polynomials of degree at most D cannot reconstruct x from $Y \sim \mathbb{P}$ with positive correlation.*

The restriction to odd p makes estimation of the vector v itself well-defined, since for even p, $v^{\otimes p} = (-v)^{\otimes p}$. For even p one may seek to estimate $v \otimes v$ instead, but this would introduce considerable additional technicalities into our approach.

Computational central limit theorems and Wigner vs. Wishart The second problem we consider is inspired by the recent work of [Mik20], who considered a *tensor central limit theorem*: given i.i.d. tensors X_j, how quickly does the distribution of $\frac{1}{\sqrt{r}} \sum_{j=1}^{r} X_j$ converge to a Gaussian tensor as r increases? That work considered a *Wishart tensor* model, where each X_j is the rank-one tensor $x_j^{\otimes p}$ with x_j a standard Gaussian vector, with "diagonal" entries of the tensor—those indexed by tuples $(i_1, \dots, i_p)$ with a repeated entry—zeroed out.[2] The author showed that $r \gg n^{2p-1}$ suffices for *information-theoretic convergence*: past this number of summands, the Wigner and Wishart laws are close in a suitable distributional distance, and no statistic whatsoever can distinguish them with high probability.

Earlier work [BDER16] considered the matrix case $p = 2$, but considered *computational convergence*: for what scaling of r can a computationally efficient hypothesis test distinguish the Wigner and Wishart laws? Their answer was $r \ll n^3$, which coincides with the information-theoretic lower bound, $n^{2p-1} = n^3$. Here we ask: how does computational convergence behave when $p > 2$?

The distribution studied by [Mik20] is not invariant, and so is not amenable to our tools. As a substitute, we propose a family of random tensors whose entries are degree p homogeneous polynomials in Gaussian random variables, just like the $x_j^{\otimes p}$ above, which our methods can treat and which we believe should behave similarly to Wishart models.

Definition 1.8. *The* real Ginibre ensemble $\mathrm{Gin}(n, \sigma^2)$ *is the law of an (asymmetric) $n \times n$ random matrix Z whose entries are i.i.d. with law $\mathcal{N}(0, \sigma^2)$.*

We write $\mathrm{Haar} = \mathrm{Haar}(n)$ for the Haar probability measure over the orthogonal group $\mathcal{O}(n)$, omitting the n when it is clear from context. The basic idea that will be involved below is that the distributions $\mathrm{Gin}(n, 1/n)$ and $\mathrm{Haar}(n)$ behave similarly: both are orthogonally invariant (Proposition 6.2), and the entries of both are typically $O(1/\sqrt{n})$. Moreover, "invariantizing" a tensor by forming $Z \cdot T$ with $Z \sim \mathrm{Gin}(n, 1/n)$ makes its entries homogeneous of degree p in the Gaussian entries of Z, giving an object resembling $x_j^{\otimes p}$ from the model of [Mik20].

[2] This ensures that the limit tensor has centered and independent Gaussian entries, while including diagonal entries would introduce correlations.

Theorem 1.9 (Lower bound for Wigner vs. Wishart detection; informal)**.** *Suppose that μ_n are probability measures on $\mathrm{Sym}^p(\mathbb{R}^n)$ satisfying the following properties for $A \sim \mu_n$:*

1. *For all $i \in [n]^p$ having a repeated entry, $A_{i_1,\dots,i_p} = 0$ almost surely.*
2. *There is a constant $C > 0$ such that, for all $i \in [n]^p$, $|A_{i_1,\dots,i_p}| \le C$ almost surely.*
3. *$\|A\|_F^2 = n^p$ almost surely.*

Let $Z_1, \dots, Z_r \sim \mathrm{Gin}(n, 1/n)$ be i.i.d. and $A_1, \dots, A_r \sim \mu_n$ be i.i.d., and write $\mathbb{P} = \mathbb{P}_{n,r}$ for the law of $r^{-1/2} \sum_{j=1}^r Z_j \cdot A_j$. There is a constant $a_{p,C} > 0$ such that the following holds. Suppose that $D = D(n) \le \sqrt{n/2p^2}$ is given and $r = r(n)$ satisfies

$$r \ge a_{p,C} \cdot \left\{ \begin{array}{ll} n^p & \text{if } p \text{ is odd,} \\ n^{3p/2} & \text{if } p \text{ is even} \end{array} \right\}. \tag{6}$$

Then, polynomials of degree at most D cannot distinguish $\mathbb{P}_{n,r(n)}$ from Wig*.*

We emphasize one simple example: consider A a deterministic tensor, with

$$A_{i_1,\dots,i_p} = c \cdot \mathbf{1}\{\text{no entry is repeated in } i\}, \tag{7}$$

where $c = c(p, n)$ is chosen (close but not equal to 1) so that $\|A\|_F^2 = n^p$. If A were simply the all-ones tensor $1^{\otimes p}$, then $Z \cdot A$ would have the law of $x_j^{\otimes p}$ from the above Wishart model. Thus, we believe it is reasonable to think of A as an invariant surrogate for the law of $x_j^{\otimes p}$ with positions having repeated indices zeroed out. Note also that, as in the latter model, in our model we have $\mathbb{E} Z_j \cdot A_j = 0$ even when p is even, so there is no need to center these terms.

The result suggests that computational central limit theorem convergence (i.e., with respect to polynomial-time algorithms) in this Wishart-like model occurs once $r \gg n^{(2+\mathbf{1}\{p \text{ even}\})p/2}$. For the special case of the above choice of A, we may bolster this proposal with a matching upper bound.

Theorem 1.10 (Upper bound for Wigner vs. Wishart detection; informal)**.** *Let A be as in* (7)*, and μ_n be the Dirac delta mass on A. Then, in the setting of Theorem 1.9, if $r = r(n)$ satisfies*

$$r \ll \left\{ \begin{array}{ll} n^p & \text{if } p \text{ is odd,} \\ n^{3p/2} & \text{if } p \text{ is even} \end{array} \right\}, \tag{8}$$

then there is a polynomial of degree $D = 3$ if p is even and $D = 4$ if p is odd that can distinguish $\mathbb{P}_{n,r(n)}$ from Wig*.*

As in the case of tensor PCA, our diagrammatic approach gives a clear intuitive explanation of the structure of this threshold. Here, the unusual-seeming dependence on the parity of p arises because the parity of p controls the smallest possible size of a p-regular multigraph on more than two vertices, which is 3 when p is even but 4 when p is odd. Our calculations will demonstrate that in this setting the smallest graphs (or disjoint unions thereof) correspond to the most powerful polynomials for hypothesis testing, so the above phenomenon accounts for detection being proportionally easier (as reflected in a larger threshold for r) when p is even.

We note also that the computational threshold $r \gg n^{(2+\mathbf{1}\{p \text{ even}\})p/2}$ we establish coincides with the information-theoretic lower bound $r \gg n^{2p-1}$ of [Mik20] when $p = 2$, but is strictly lower by a polynomial factor once $p \ge 3$. In this sense, our result gives initial evidence that testing Wigner vs. Wishart tensors may exhibit a statistical-computational gap not present in the matrix case (provided that the information-theoretic lower bound of [Mik20] can be matched by an upper bound, which to the best of our knowledge is not yet established).

1.2 Main Proof Technique: Tensorial Finite Free Cumulants

The following remarkable objects are at the heart of our approach to analyzing low-degree algorithms for invariant problems.

Theorem 1.11 (Finite free cumulants of a tensor)**.** *For any $0 \le D \le \sqrt{n/2p^2}$, there is a collection of polynomials $\kappa_G(T)$ of degree at most D in the indices of a tensor T, indexed by (non-isomorphic) p-regular multigraphs on at most D vertices, such that the following hold:*

1. *(Empty Graph) $\kappa_\varnothing(T) = 1$.*
2. *(Invariance) The κ_G are invariant polynomials: $\kappa_G(Q \cdot T) = \kappa_G(T)$ for all $Q \in \mathcal{O}(n)$.*
3. *(Basis) The κ_G are a basis for the invariant polynomials of degree at most D.*
4. *(Approximate Orthonormality) The Gram matrix $M_{G,H} := \mathbb{E}_{T \sim \mathrm{Wig}}[\kappa_G(T)\kappa_H(T)]$, restricted to those G and H none of whose connected components consist of two vertices connected by p parallel edges, satisfies*
$$\frac{1}{2} < \lambda_{\min}(M) \le \lambda_{\max}(M) \le 2. \tag{9}$$
5. *(Wigner Expectations) For any G that is not empty and not a single two-vertex connected component of the kind described above,*
$$\mathop{\mathbb{E}}_{T \sim \mathrm{Wig}} \kappa_G(T) = 0. \tag{10}$$
6. *(Additive Free Convolution) For any $S, T \in \mathrm{Sym}^p(\mathbb{R}^n)$ and G connected,*
$$\mathop{\mathbb{E}}_{Q \sim \mathrm{Haar}} \kappa_G(S + Q \cdot T) = \kappa_G(S) + \kappa_G(T). \tag{11}$$

Remark 1.12. *When $D > \sqrt{n/2p^2}$, the finite free cumulants exist and satisfy Conditions 1–3 and 5–6, but Condition 4 breaks down. More dramatically, once $D \gg \sqrt{n}$ the Gram matrix of the finite free cumulants no longer has a bounded condition number, so they are no longer approximately orthogonal. See Appendix C for further details.*

Of course, by basic linear algebra over the space of polynomials endowed with inner product $\langle f, g \rangle = \mathbb{E}_{T \sim \mathrm{Wig}}[f(T)g(T)]$, there must exist polynomials satisfying Claims 1–5, and indeed if this were all we were interested in we could demand exact orthonormality in Claim 4. What is surprising is, first, that we can identify such polynomials explicitly (analogous to, say, the Boolean Fourier basis or the Hermite basis, as opposed to arbitrary abstract collections of orthogonal polynomials), and second, that those same polynomials happen to satisfy the exact algebraic Condition 6.

We will see that these objects are useful tools for the analysis of low-degree polynomial algorithms in models having both invariance and additive structure. Consider the case of hypothesis testing. The analysis of low-degree polynomials, given an approximate basis of this kind, boils down to evaluating expectations of κ_G under the alternative hypothesis—the distribution of $Y = \lambda v^{\otimes p} + W$ in tensor PCA, or of $r^{-1/2} \sum_{i=1}^r Z_i \cdot A_i$ in a computational central limit theorem. But, Claim 6

above gives a powerful tool to handle such situations, as is especially apparent for tensor PCA: combining Claims 5 and 6, we see that

$$\mathop{\mathbb{E}}_{v,W} \kappa_G(\lambda v^{\otimes p} + W) = \mathop{\mathbb{E}}_{v,W,Q} \kappa_G(\lambda v^{\otimes p} + Q \cdot W) = \mathop{\mathbb{E}}_{v} \kappa_G(\lambda v^{\otimes p}) + \mathop{\mathbb{E}}_{W} \kappa_G(W) = \mathop{\mathbb{E}}_{v} \kappa_G(\lambda v^{\otimes p}), \tag{12}$$

and the last expression is simple to compute in closed form (Proposition 5.8). To give some intuition, the result is equal to leading order to $\mathbb{E}_v m_G(\lambda v^{\otimes p})$ (generally, the κ_G are built by a series of "adjustments" to the corresponding m_G), which, even inside the expectation, is easily computed by hand as $\lambda^{|V(G)|}\|v\|^{2|E(G)|}$ (Proposition 5.5).

As we have mentioned, there is a deeper mathematical meaning to the κ_G: as Claim 6 hints at, the κ_G are really the correct generalization to tensors of *finite free cumulants* from the free probability theory of random matrices. In this regard, our κ_G are the third step in a chain of generalizations in the literature, which we review below.

First, the classical cumulants are polynomials $\kappa_t^{(1)}$ in the moments of a random variable with the property that, if A and B are independent scalar random variables, then the cumulant of their convolution $A + B$ is the sum of their cumulants: $\kappa(A + B) = \kappa(A) + \kappa(B)$. For instance, the first cumulant is the expectation, and the second cumulant is the variance, for which these additivity rules are familiar.

Second, the theory of *free probability* considers the less straightforward question of how addition affects the spectra of *matrix-valued* random variables A and B. To make this as tractable as the scalar case, it is not enough to assume that A and B are independent—the typical spectrum of $A+B$ still depends on the extent to which A and B commute. Free probability restricts its attention to *freely independent* pairs A, B whose frames of eigenvectors are "maximally uncorrelated" with each other. For our purposes, we may take this to mean that we are interested in the typical spectrum of $A + Q^\top BQ = A + Q \cdot B$ for $Q \sim$ Haar drawn independently of A and B. The *finite free cumulants* $\kappa_t^{(2)}$ then satisfy this additivity property in expectation, $\mathbb{E}_Q \kappa_t^{(2)}(A+Q\cdot B) = \kappa_t^{(2)}(A)+\kappa_t^{(2)}(B)$, just as in Claim 6. They are a relatively recent innovation in free probability [AP18, AGVP23], but a notion of their limit as $n \to \infty$, called just the *free cumulants*, was crucial to the earlier origins of the theory (see, e.g., the reference [MS17]).

Given that they satisfy Claim 6, it should now be clear that our κ_G generalize this notion of finite free cumulant to tensors. Even for matrices, it seems not to have been noticed previously that finite free cumulants lead two parallel lives: on the one hand, they are statistics of random matrices that behave well under freely independent summations; on the other, as our work shows, they yield a natural, explicit, near-orthogonal basis of invariant functions under the inner product induced by the Wigner distribution, which for matrices is the Gaussian orthogonal ensemble. Also interesting is that there is no clear notion of limiting free cumulants for tensors, because there is also no clear notion of eigenvalues or empirical spectral distribution. Still, *finite* free cumulants are sensible and useful objects.

The astute reader will notice that Claim 6 is restricted to connected graphs—a simple issue that we address in Proposition 4.27—and that Claims 4 and 5 make exceptions for a particular type of connected component, namely a two-vertex multigraph, which we call a *Frobenius pair* (see Section 4.1). The associated cumulant may be viewed as a formal analog of the variance (the second classical cumulant), and like the variance is always non-negative. To actually use the finite free cumulants as a basis, we will have to define and work carefully with a suitable re-centering that accounts for this, which will be one of our main tasks in Section 4.

We also have not said anything about the application of these ideas to estimation, which requires vector-valued functions of a tensor. This turns out to admit a parallel theory concerning *equivariant functions*, ones $f : \mathrm{Sym}^p(\mathbb{R}^n) \to \mathbb{R}^n$ satisfying $f(Q \cdot T) = Q \cdot f(T)$, which are spanned by constructions like $m_G(T)$ but where G has an "open" edge corresponding to the vector output. These modified graph moments also are spanned by a basis of objects similar to the κ_G. Our lower bound for reconstruction is obtained by using this basis in conjunction with the existing strategy of [SW22].

1.3 Related Work

Eigenvalues of tensors There are several reasonable different definitions of eigenvalues and eigenvalue-eigenvector pairs for tensors [Qi05, Qi07, QCC18]. It is non-trivial to compute these eigenvalues [CDN14], and, under the common notion of "E-eigenvalues," typical tensors have exponentially many eigenvalues in their dimension, per the analysis of the critical points of spherical spin glass models [ABAČ13, Sub17]. Most importantly, there seems to be no natural analog for tensors of the spectral decomposition for matrices, making it seemingly impossible in general to reduce the evaluation of tensor invariants to computations with eigenvalues. This is in contrast to the matrix setting where the invariants are just power sums of the eigenvalues, $\mathrm{tr}(T^\ell) = \sum_{i=1}^n \lambda_i^\ell$, and polynomials thereof.

Random tensor theory Aspects of random matrix theory that have been generalized to the tensor setting[3] include the derivation of the maximum eigenvalue of a Gaussian random tensor by a "tensorial power method" [Evn21], a definition of a resolvent, a spectral density, and Wigner's semicircle law for Gaussian random symmetric tensors [Gur20], universality of this generalization of Wigner's law [Gur14] (with applications to spherical spin glass models [BGS13]), and Harish-Chandra–Itzykson–Zuber integrals [CGL23b, CGL23a]. See also the book [Gur17] for many details. We will both elaborate on some of these results (see, e.g., our Theorem 5.3 on the scaling of graph moments of Gaussian random tensors) and will show generalizations to the tensor setting of aspects of free probability theory, including the notion of free cumulants, properties of additive free convolution, and Voiculescu's theorem on the asymptotic freeness of matrices under random rotation of their eigenvectors.

Very recently, and independently of our work, another approach to free probability for tensors based on Gurau's tensor resolvent was proposed in [Bon24]. This line of work differs from our results in focusing on a coarser collection of invariants which is, in our notation, the sum of the moments $m_G(T)$ over all connected G on a given number of vertices.

Tensor PCA Starting with the work of [MR14], the tensor PCA problem is now well-studied. A variety of different polynomial-time algorithms are known to achieve the scaling $\lambda \sim n^{-p/4}$ [MR14, HSS15, HSSS16, ADGM17, Has20, BCRT20]. While information-theoretically it is possible to succeed for smaller λ [JLM20], we expect that no poly-time algorithm can achieve this due to lower bounds against low-degree polynomials [HKP$^+$17] and a reduction from a variant of the planted clique problem [BB20]. Furthermore, various algorithms are known to achieve a particular

[3]Some works, such as [AHH12], also use the term "random tensor theory" to describe their results, though they really concern random matrices built out of random tensors, say by taking sample covariances of vectorizations of these tensors.

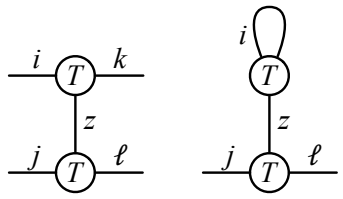


Figure 2: On the left, a tensor network defining a 4-index tensor from two copies of a 3-ary tensor; we sum over all n values of the internal index z. It can be viewed, among other things, as an n^2-dimensional matrix $M_{(i,j),(k,\ell)}$. This matrix is used in a spectral algorithm for tensor PCA in [HSS15]. On the right, a partial trace where we also sum over i, giving an n-dimensional matrix $M_{j\ell}$ used by [HSSS16] for another spectral algorithm.

tradeoff between the SNR λ and the (super-polynomial but sub-exponential) runtime, when $\lambda \ll n^{-p/4}$ [RRS17a, BGG$^+$16, BGL17, WAM19].[4] The low-degree result of [KWB22] that we recalled in Theorem 1.6 suggests that this tradeoff cannot be improved. Most closely related to our results, there is a line of work considering the application of tensor invariants to tensor PCA, using the "random tensor theory" machinery mentioned above [Gur20, OTR22, OR20, Oue22].

Algorithms using tensor networks and distinct indices Our notion of "tensorial free cumulants" arose from seeking an invariant version of the "distinct-index" tensor networks used in the recent work [Wei23] on tensor decomposition. In the matrix setting, if A is the adjacency matrix of a graph, the distinct-index trace

$$A^{(\ell)} = \sum_{\substack{i_1,i_2,\ldots,i_\ell \\ \text{distinct}}} A_{i_1,i_2} A_{i_2,i_3} \cdots A_{i_{\ell-1},i_\ell}$$

counts self-avoiding walks. This was used previously by [Mas14] in the context of community detection in the stochastic block model, and by [DHS20] as a variance reduction technique for estimators and test statistics in matrix PCA. Other previous uses of tensor networks in tensor PCA include [HSS15], which uses the n^2-dimensional matrix $M_{(i,j),(k,\ell)} = \sum_z T_{ikz} T_{j\ell z}$ shown in Figure 2 on the left, and [HSSS16] which uses its partial trace on the right. Algorithmic applications of more elaborate tensor networks can be found in [MW19, DdL$^+$22]. The recent work of [Sem24] uses similar ideas of optimizing over invariant and equivariant polynomials as well, but, working in the matrix setting, does not use the graphical tensor network notation.

The *graph matrices* widely used to analyze the *pseudocalibration* construction in lower bounds against the sum-of-squares hierarchy also feature a restriction of similar summations to distinct indices, and through this give an orthogonal basis for the polynomials invariant under a group action (in that case the action of the symmetric group). See [BHK$^+$19, GJJ$^+$20, PR22, JPRX23] for applications of pseudocalibration, and Remark 2.3 of [AMP16] for discussion of the orthogonality properties of graph matrices. Finally, the combinatorial structure of a nearly-orthogonal basis for a special class of tensor networks arising in [GJJ$^+$20] was studied by [JP21].

[4] Notably, unlike in the matrix case, once $p \geq 3$ the tensor analog of power iteration—which may be viewed as computing a tree-shaped tensor network—performs suboptimally [MR14, WZ24].

Cumulants in low-degree lower bounds The work of [SW22] also treats reconstruction using low-degree polynomials, and also encounters a quantity that is referred to as a "cumulant." Those quantities are simply the classical (joint) cumulants of scalar random variables. In contrast, our cumulants are firstly the *free* cumulants of free probability, and secondly are polynomial functions of a matrix or tensor, rather than statistics of an entire probability distribution.

2 Notation

We write $\mathbb{1}$ for the identity matrix, J for the all-ones matrix, and 1 for the all-ones vector (when it is used in linear-algebraic context). We write $n^{\underline{b}} := n!/(n-b)!$ for the falling factorial, and use the less conventional $n^{\underline{b}} := n!!/(n-2b)!!$ for the falling double factorial.

3 Preliminaries

3.1 Closed Tensor Networks as Invariant Polynomials

First, let us motivate the family of linear combinations of tensor networks as a natural class of algorithms. For computations over symmetric matrices T, an important role is played by *linear spectral statistics*, the quantities $F(T) = \sum_{i=1}^{n} f(\lambda_i(T))$ for some function $f : \mathbb{R} \to \mathbb{R}$.

The space of all linear spectral statistics is a natural limit of the same space where f are polynomials. That space admits a more algebraic interpretation: it is the same as the space of all symmetric polynomials of the eigenvalues of T, which is also the space of all invariant polynomials of a matrix T: if $F(QTQ^\top) = F(T)$ for all $Q \in \mathcal{O}(n)$ for a polynomial F, then F must be a symmetric polynomial of $(\lambda_1(T), \dots, \lambda_n(T))$.

In the matrix case, when $f(\lambda)$ is a polynomial, then we have $F(T) = \mathrm{tr}(f(T))$. These are linear combinations of $\mathrm{tr}(T^k)$, and one may check that $\mathrm{tr}(T^k) = m_{C_k}(T)$, where C_k is the k-cycle. In summary, then, the linear spectral statistics for f a polynomial—a natural class of "spectral algorithms" over matrices—have two coinciding interpretations: on the one hand, they are all invariant functions of a matrix T; on the other, they are linear combinations of the $m_G(T)$ where T is a 2-regular multigraph.

These interpretations do not depend on the notion of "eigenvalue." While there is no one clear definition of eigenvalues for tensors, this algebraic interpretation of linear spectral statistics does admit a tensorial generalization. In particular, we have:

Definition 3.1 (p-regular multigraphs)**.** *Write $\mathcal{G}_{d,p}$ for the set of (non-isomorphic) p-regular multigraphs on d unlabelled vertices.*

Theorem 3.2. *Let R be the ring of invariant polynomials $f : \mathrm{Sym}^p(\mathbb{R}^n) \to \mathbb{R}$. Then R is generated as a vector space by the polynomials $m_G(T)$ over $G \in \bigsqcup_{d \geq 0} \mathcal{G}_{d,p}$. In particular, any invariant homogeneous polynomial $f(T)$ of degree d is a linear combination $\sum_i \alpha_i m_{G_i}(T)$ for $G_i \in \mathcal{G}_{d,p}$.*

In this sense, we propose that the space of linear combinations of $m_G(T)$ should be viewed as the "right" tensorial generalization of linear spectral statistics. (The analogy between $m_G(T)$ for T a tensor and $m_{C_k}(T) = \mathrm{tr}(T^k)$ for T a matrix, the latter of which is, up to renormalization, the kth moment of the empirical spectral distribution of T, is also why we call the $m_G(T)$ "moments.")

Theorem 3.2 is a classical result in invariant theory, namely the "First Fundamental Theorem" for the orthogonal group [Wey46]; see [GW98, §4.3] for a modern treatment, following [ABP73,

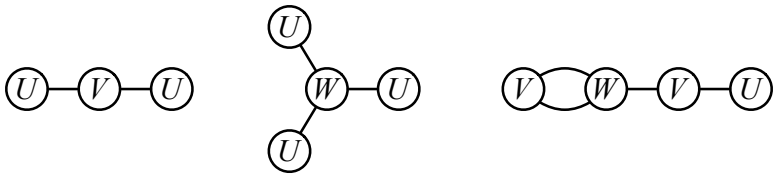


Figure 3: Three examples of mixed moments of tensors. Here U, V, and W have arity 1, 2, and 3 respectively. From left to right, these yield the bilinear form $\langle U, VU\rangle$, the trilinear form $\langle W, U^{\otimes 3}\rangle$, and the quantity in (14).

pp. 285–6].[5] However, these proofs may be opaque to some readers, and lack graphical panache. We provide a graphical proof in the style of our other calculations in Appendix A for the reader's appreciation.

We can also consider mixed moments of multiple tensors, possibly with different arities:

Definition 3.3. *Let $\mathcal{T}$ be a set of symmetric tensors, and let $G = (V, E)$ be a multigraph associated with a function $T : V \to \mathcal{T}$ that labels each vertex $v \in V$ with a tensor $T(v) \in \mathcal{T}$ of arity $\deg v$. Then define the mixed moment $m_G(\mathcal{T})$ as the contraction*

$$m_G(\mathcal{T}) = \sum_{i \in [n]^E} \prod_{v \in V} T(v)_{i(\partial v)}\,. \tag{13}$$

Example 3.4. *Let U, V, and W be tensors of arity 1, 2, and 3 respectively. The rightmost multigraph G in Figure 3 yields*

$$m_G(U, V, W) = \sum_{ijk\ell} V_{ij}\, W_{ijk}\, V_{k\ell}\, U_\ell\,, \tag{14}$$

which is quadratic in V and multilinear in W and U.

Example 3.5. *Let U and V be matrices of the same dimension. Then, the mixed connected graph moments of these matrices are the mixed traces, $\operatorname{tr}(U^{k_1} V^{\ell_1} \cdots U^{k_m} V^{k_m})$ for $k_i, \ell_j \geq 1$. These are the quantities to which the concept of* freeness *in free probability theory pertains; we will see later that graph moments of tensors admit a generalization of parts of this theory.*

A version of Theorem 3.2 also holds for this setting:

Theorem 3.6. *Let R be the ring of invariant polynomials $f : \operatorname{Sym}^{p_1}(\mathbb{R}^n) \times \cdots \times \operatorname{Sym}^{p_m}(\mathbb{R}^n)$, in the sense that $f(T_1, \ldots, T_m) = f(Q \cdot T_1, \ldots, Q \cdot T_m)$ for all $Q \in \mathcal{O}(n)$. Then, R is generated as a vector space by the $m_G(T_1, \ldots, T_m)$: any invariant homogeneous polynomial $f(T_1, \ldots, T_m)$ of degree d_i in T_i for each $i \in [m]$ is a linear combination of mixed moments $m_G(T_1, \ldots, T_m)$ where each T_i corresponds to d_i vertices of G.*

We again give a proof in Appendix A, which is essentially identical to that of Theorem 3.2. This is a non-trivial generalization even in the case $d_i = 2$ of matrices, due to [Pro76], and we give a simple and unified proof for either setting.

[5] These results are usually stated for the action of $\mathcal{O}(n)$ on all "axes" of a general tensor $T \in (\mathbb{R}_n)^{\otimes p}$, but by Proposition A.1 they also immediately apply to invariants of symmetric tensors.

3.2 Open Tensor Networks as Equivariant Polynomials

We can also consider vector-, matrix-, or tensor-valued functions $f(T)$ whose outputs transform properly when T is transformed. This will include, for instance, an optimal algorithm for reconstructing the spike in tensor PCA.

Definition 3.7. *Let f be a function from p-ary tensors to ℓ-ary tensors. We say that f is* equivariant *if for any T and any orthogonal matrix $Q \in \mathcal{O}(n)$,*

$$f(Q \cdot T) = Q \cdot f(T)\,,$$

and similarly if f is a function from multiple tensors to ℓ-ary tensors.

Example 3.8. *Let Q be a matrix. If v is a left eigenvector of T, then $Q \cdot v = vQ$ is a left eigenvector of $Q \cdot T = Q^\top T Q$. Thus the output of a spectral algorithm that returns the dominant eigenvector of a matrix is equivariant.*

Given the machinery we have set up, it is easy to generalize Theorem 3.2 to show that the equivariant polynomials are linear combinations of "open graph" moments where the graph has ℓ open edges; see [Wey46, p.64] and [ABP73, p.286]. These are also known as partial contractions.

Definition 3.9. *An t-*open multigraph *G is a triple (V, E, E') where each edge $e \in E \cup E'$ is a multiset $e \subseteq V$ with $|e| = 2$ if $e \in E$ and $|e| = 1$ if $e \in E'$, and where $|E'| = t$. We say E and E' are the sets of closed and open edges respectively. The degree of a vertex $v \in V$ is the total number of times it appears in $E \cup E'$, including self-loops in E where it occurs twice. We say G is p-regular if every vertex has degree p.*

Definition 3.10. *Let $T \in \mathrm{Sym}^p(\mathbb{R}^n)$ and let $G = (V, E, E')$ be a p-regular open multigraph with $|E'| = \ell$ open edges. Let $i \in [n]^{E \cup E'}$ denote a labeling of the edges with indices in $[n]$. Let $i(E)$ and $i(E')$ denote the labels of the closed and open edges respectively, and as in Definition 1.3 let $i(\partial v)$ denote the multiset of indices $i(e)$ associated with the edges e incident to v. Then, the associated* open graph moment *$m_G(T)$ is the following ℓ-ary tensor,*

$$m_G(T)_{i(E')} = \sum_{i(E) \in [n]^E} \prod_{v \in V} T_{i(\partial v)}\,.$$

For a set of tensors $\mathcal{T}$, we similarly define

$$m_G(\mathcal{T})_{i(E')} = \sum_{i(E) \in [n]^E} \prod_{v \in V} T(v)_{i(\partial v)}\,.$$

where for each $v \in V$, $T(v) \in \mathcal{T}$ has arity $\deg v$.

In words, these partial contractions are graph moments where we sum only over the indices on closed edges, and leave the indices on open edges as unbound variables.

Example 3.11. *We show three 3-regular open multigraphs in Figure 4. The first two have one open edge each, and the third has two open edges. Thus their partial contractions $m_G(T)_{i(E')}$ result in a vector or matrix. For the first graph we have the partial trace $m_G(T)_i = \sum_j T_{ijj}$, and for the third one we have the matrix*

$$m_G(T)_{ii'} = \sum_{jk\ell mn} T_{ijk}\, T_{j\ell m}\, T_{i'\ell n}\, T_{kmn}\,. \tag{15}$$

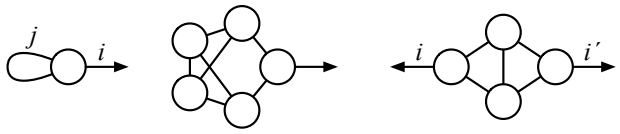


Figure 4: Three open multigraphs whose partial contractions $m_G(T)$ yield vectors or matrices, i.e., 1-ary or 2-ary tensors. The first is the vector $m_G(T)_i = \sum_j T_{ijj}$. The third is the matrix $m_G(T)_{ii'}$ given by (15).

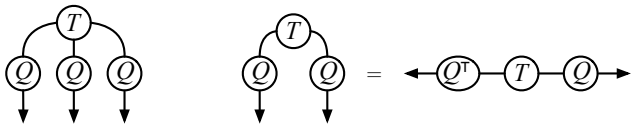


Figure 5: On the left, conjugating a 3-ary tensor T by an orthogonal matrix Q. Each index of T undergoes the same orthogonal basis change. On the right, for a 2-ary tensor, i.e., a matrix, this coincides with the usual notion of conjugation $Q^\top TQ$. Arrows indicate that T's indices are contracted with the left index i of Q_{ij}. Reversing an arrow converts Q to its transpose $Q^\top$.

This is the graph moment $G_{K_4}(T)$ given by (3) except that i and i' are open indices rather than being set equal and summed over. Taking the trace of this matrix closes this loop and yields $G_{K_4}(T)$.

We may already start to glean some of the benefits of this graphical language. For example, it lets us understand more clearly the change of basis from Definition 1.1.

Remark 3.12. *As shown in Figure 5, $Q \cdot T$ consists of placing a copy of Q on each "outward" edge of T. Since Q is not symmetric, we place arrows on the edges to indicate which of its indices is contracted with an index of T; reversing this arrow swaps Q's indices and thus converts it to its transpose $Q^\top$. In the matrix case $p = 2$, $Q \cdot T$ corresponds to standard matrix conjugation, since $Q^\top TQ$ can also be written $T \cdot Q^{\otimes 2}$:*

$$\left(Q^\top TQ\right)_{jj'} = \sum_{i,i'} Q^\top_{ji}\, T_{ii'}\, Q_{i'j'} = \sum_{ii'} T_{ii'}\, Q_{ij}\, Q_{i'j'} = \left(T \cdot Q^{\otimes 2}\right)_{jj'} .$$

Using this, the proof that graph moments are invariant functions becomes straightforward.

Proposition 3.13. *For any $T \in \mathrm{Sym}^p(\mathbb{R}^n)$, any multigraph G, and any $Q \in \mathcal{O}(n)$, $m_G(Q \cdot T) = m_G(T)$. The same holds for mixed moments $m_G(\mathcal{T})$ as in Definition 3.3.*

Proof. Since reversing the direction of an edge changes Q to $Q^\top$, placing outgoing Q's on the edges of each vertex creates a pair $QQ^\top$ on each edge. Since Q is orthogonal, $Q^\top = Q^{-1}$ and $QQ^\top = \mathbb{1}$. □

Finally, analogous results hold for open graph moments as did for closed ones. First, just as closed graph moments are invariant, open graph moments are equivariant:

Proposition 3.14. *For any $T \in \mathrm{Sym}^p(\mathbb{R}^n)$, any open multigraph G, and any $Q \in \mathcal{O}(n)$, $m_G(Q \cdot T) = Q \cdot m_G(T)$. The same holds for mixed moments $m_G(\mathcal{T})$ as in Definition 3.10.*

Proof. As in Proposition 3.13, placing an outgoing Q on the edges of each vertex causes a canceling pair $QQ^\top = \mathbb{1}$ on each closed edge, and an outgoing Q on each open edge. The latter transform $m_G(T)$ to $Q \cdot m_G(T)$. □

And, as for closed graph moments and invariant polynomials, the open graph moments span all equivariant polynomials.

Theorem 3.15. *Let p be an equivariant function from p-ary tensors to ℓ-ary tensors. If f is a homogeneous polynomial of degree d, then $f(T)$ is a linear combination $\sum_i \alpha_i m_{G_i}(T)$ where the G_i are p-regular open multigraphs with d vertices and ℓ open edges.*

More generally, let $f(T_1, \ldots, T_m)$ be an equivariant function that yields ℓ-ary tensors. If p is a homogeneous polynomial of degree d_i in each T_i, then is a linear combination of mixed moments $m_G(T_1, \ldots, T_m)$ where G has ℓ open edges and each T_i corresponds to d_i vertices of G.

Once again, we defer a proof to Appendix A.

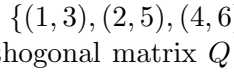
3.3 Invariant Tensors, Brauer Space, and the Weingarten Function

An important role in our calculations will be played by the subspace of vectors in $(\mathbb{R}^n)^{\otimes \ell}$ that is invariant under the action of $\mathcal{O}(n)$ on this space, i.e., the vectors w such that $Q^{\otimes \ell} w = w$ for all $Q \in \mathcal{O}(n)$ (not to be confused with our other discussion of invariant polynomials). In our case we will have $\ell = pd$ for working with $G \in \mathcal{G}_{d,p}$. This is sometimes called the *Brauer space* after Brauer's early work on its structure.

In particular, his classic results in the representation theory of Lie groups [Bra37] show that this subspace is spanned by vectors of the following form. Assume ℓ is even; otherwise, since $-\mathbb{1} \in \mathcal{O}(n)$, only the zero vector is invariant. Then, for each perfect matching $\mu = \{(t_1, t'_1), \ldots, (t_{\ell/2}, t'_{\ell/2})\}$ of $[\ell]$, define

$$w(\mu)_{i_1,\ldots,i_\ell} = \prod_{(t,t') \in \mu} \delta_{i_t, i_{t'}} \tag{16}$$

where δ is the Kronecker delta $\delta_{i,i'} = 1$ if $i = i'$ and 0 otherwise. In tensor network notation, we can regard these *matching vectors* as ℓ-ary tensors consisting of a set of "cups," each of which ensures that two indices are identical. Alternatively, one may view them as tensor powers of the identity matrix, though viewed as a tensor rather than a matrix, i.e. $(\sum_{i=1}^n e_i \otimes e_i)^{\otimes \ell/2}$, subject to various permutations of the vector axes. Note in particular that the $w(\mu)$, unlike the tensors discussed earlier, are *not* symmetric tensors; indeed, their symmetrizations are all the same, not depending on the matching μ.

For instance, if $\ell = 6$ and $\mu = \{(1,3), (2,5), (4,6)\}$ then $w(\mu) =$  . These vectors are invariant since, if we apply an orthogonal matrix Q to both indices of a cup, Q and $Q^\top$ meet in the middle and cancel (just as we will discuss in Proposition 3.13). Brauer's result cited above is that these are *all* of the invariant tensors.

Definition 3.16. *For a finite set X, write $\mathcal{M}(X)$ for the set of perfect matchings of X.*

Proposition 3.17. *The subspace of $\mathcal{O}(n)$-invariant tensors in $(\mathbb{R}^n)^{\otimes \ell}$ is spanned by the $w(\mu)$ over $\mu \in \mathcal{M}([\ell])$.*

We will also need to work with the orthogonal projection to this invariant subspace. On the one hand, by general representation theory, this object is just the averaging operator over all rotations,

$$\Pi_\ell := \mathop{\mathbb{E}}_{Q} Q^{\otimes \ell}, \tag{17}$$

viewed as an operator on $(\mathbb{R}^n)^{\otimes \ell}$ (equivalently, as a large matrix formed as the Kronecker power). In particular, Π_ℓ contains as its entries the moments of the Haar measure over $\mathcal{O}(n)$. This gives an elegant connection between the probability theory of the Haar measure and the representation theory of $\mathcal{O}(n)$.

On the other hand, by elementary linear algebra, Π_ℓ is given by some linear combination of rank one matrices,

$$\Pi_\ell = \sum_{\mu,\nu} \mathrm{Wg}_{\mu,\nu} w(\mu) \otimes w(\nu). \tag{18}$$

Indeed, the matrix Wg can be taken to be the Moore–Penrose pseudoinverse of the Gram matrix of the $w(\mu)$. Its entries as a function of μ, ν are known as the *(orthogonal) Weingarten function*.

In particular, Π_ℓ therefore admits a diagrammatic description, as a linear combination of operator "switching" one matching to another (the outer products of two matching vectors). For instance, Π_2 is the "cupcap"

$$\Pi_2 = \mathop{\mathbb{E}}_{Q} Q \otimes Q = \frac{1}{n}$$ 

namely the outer product of $w(\{(1,2)\}) =$  with itself. The normalization $1/n$ comes from the fact that

$$\Big|$$ $$\Big|^2 = \mathrm{tr}$$  $$= \bigcirc = n\,,$$

corresponding to the trace of the identity matrix equaling n. Π_4 is already more complicated:

$$\Pi_4 = \frac{1}{(n+2)n(n-1)} \Big[(n+1)$$  $$\quad -$$  $$\quad -$$ 

$$-$$  $$+ (n+1)$$  $$\quad -$$ 

$$-$$  $$\quad -$$  $$+ (n+1)$$  $$\Big]$$

Fortunately, there is a combinatorial description of the Weingarten function that aids in computing these coefficients, as we discuss in Appendix C.

3.4 Low-Degree Polynomial Algorithms

The idea to consider polynomials as algorithms for statistical problems with degree as a measure of complexity first arose from [BHK+19, HS17, HKP+17] and the framework has been subsequently refined over the years, e.g., by [Hop18, KWB22]; see, for instance, [BEAH+22] for a modern account. Originally this line of work considered simple "planted versus null" hypothesis testing problems, but later extensions treated other styles of questions such as estimation [SW22], optimization [GJW24],

and refutation [KVWX23], as well as more complex testing problems [RSWY23]. The low-degree polynomial framework has by now found numerous applications and is a leading approach to understand computational complexity of statistical problems.

We will next specify what it means for a polynomial to solve a detection or reconstruction problem.

Definition 3.18 (Low-degree advantage)**.**

$$\mathrm{Adv}_{\leq D}(\mathbb{Q},\mathbb{P}) := \sup_{\substack{f\in\mathbb{R}[Y]_{\leq D}\\ \mathbb{E}_{Y\sim\mathbb{Q}}f(Y)^2\neq 0}} \frac{\mathbb{E}_{Y\sim\mathbb{P}}f(Y)}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}f(Y)^2}}. \tag{19}$$

Remark 3.19. *This notion has been standard since the work of [HS17, HKP$^+$17, Hop18], where its boundedness, i.e.,* $\mathrm{Adv}_{\leq D}(\mathbb{Q},\mathbb{P}) = O(1)$*, indicates hardness at degree* D*. The value* Adv *is also called the* norm of the low-degree likelihood ratio *[Hop18] but the terminology* Adv *avoids defining likelihood ratios and low-degree projections. While boundedness of* Adv *does not completely rule out low-degree polynomials as a test statistic, it means that we cannot use Chebyshev's inequality to show that one succeeds. If* $\mathrm{Adv} = \omega(1)$*, this may suggest success of low-degree tests, but a number of recent examples show that this is not necessarily the case [BEAH$^+$22, COGHK$^+$22, DMW23, DDL23].*

In light of this, one should really define "success" for low-degree tests not in terms of Adv *but by the notion of "strong separation" first coined in [BEAH$^+$22]. This way, "success" legitimately implies a high-probability distinguisher, and furthermore "success" can be precluded by showing* $\mathrm{Adv} = O(1)$ *(or even by bounding a conditional variant of* Adv*). There is also a related notion of "weak separation"; again, see [BEAH$^+$22].*

Our Theorem 1.6 (or rather, the formal version Theorem 5.7) will focus on showing Adv *is either* $O(1)$ *or* $\omega(1)$*, but we expect that our proof of* $\mathrm{Adv} = \omega(1)$ *could be strengthened to a proof of strong separation.*

For reconstruction tasks, success is naturally measured in terms of mean squared error.

Definition 3.20 (Low-degree minimum mean squared error [SW22])**.**

$$\mathrm{MMSE}_{\leq D}(\mathbb{P}) := \inf_{f\in\mathbb{R}[Y]^n_{\leq D}} \mathop{\mathbb{E}}_{(x,Y)\sim\mathbb{P}} \|f(Y)-x\|^2. \tag{20}$$

Similar to Fact 1.1 of [SW22], this can be equivalently formulated in terms of "correlation":

$$\mathrm{MMSE}_{\leq D}(\mathbb{P}) = \mathbb{E}\|x\|^2 - \mathrm{Corr}_{\leq D}(\mathbb{P})^2 \tag{21}$$

where Corr is defined below.

Definition 3.21 (Low-degree correlation)**.**

$$\mathrm{Corr}_{\leq D}(\mathbb{P}) := \sup_{\substack{f\in\mathbb{R}[Y]^n_{\leq D}\\ \mathbb{E}_{Y\sim\mathbb{P}}\|f(Y)\|^2\neq 0}} \frac{\mathbb{E}_{(x,Y)\sim\mathbb{P}}\langle x, f(Y)\rangle}{\sqrt{\mathbb{E}_{Y\sim\mathbb{P}}\|f(Y)\|^2}}. \tag{22}$$

In tensor PCA we have $\|x\|^2 = n$, so to rule out non-trivial reconstruction we will aim to show $\mathrm{Corr}_{\leq D}(\mathbb{P})^2 = o(n)$.

Finally, to relate low-degree polynomial algorithms to our study of invariant polynomials, we show that both the advantage and correlation may be restricted to invariant polynomials without changing their value, provided the underlying distributions have certain invariance properties.

Proposition 3.22. *Suppose that both $\mathbb{Q}$ and $\mathbb{P}$ are invariant distributions. Then,*

$$\mathrm{Adv}_{\leq D}(\mathbb{Q},\mathbb{P}) = \sup_{\substack{f\in\mathbb{R}[Y]_{\leq D}\\ f \text{ invariant}\\ \mathbb{E}_{Y\sim\mathbb{Q}}f(Y)^2\neq 0}} \frac{\mathbb{E}_{Y\sim\mathbb{P}}f(Y)}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}f(Y)^2}}. \tag{23}$$

Proof. Consider $\bar{f}(Y) := \mathbb{E}_{Q\sim\mathrm{Haar}(n)}f(Q\cdot Y)$. Then, $\bar{f}$ is invariant. And, we have

$$\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}\bar{f}(Y) = \mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}(n)}\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}f(Q\cdot Y) = \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}f(Y), \tag{24}$$

by the invariance of $\mathbb{P}$. The same holds for $\mathbb{Q}$ with f replaced by f^2, and the result follows. □

Remark 3.23. *By the same operations on f, one may also show that, outside of the low-degree framework but under the same assumptions on $\mathbb{P}$ and $\mathbb{Q}$, if there is a test statistic achieving given Type I and II error probabilities, then there is an invariant test statistic achieving the same.*

Proposition 3.24. *Suppose that $\mathbb{P}$ is a distribution over $\mathbb{R}^n\times\mathrm{Sym}^p(\mathbb{R}^n)$ such that $(x,Y)\sim\mathbb{P}$ has the same law as $(Qx,Q\cdot Y)$ for any $Q\in\mathcal{O}(n)$ (that is, so that $\mathbb{P}$ viewed as a distribution over $x\otimes Y$ is invariant). Then,*

$$\mathrm{Corr}_{\leq D}(\mathbb{P}) = \sup_{\substack{f\in\mathbb{R}[Y]^n_{\leq D}\\ f \text{ equivariant}\\ \mathbb{E}_{Y\sim\mathbb{P}}\|f(Y)\|^2\neq 0}} \frac{\mathbb{E}_{(x,Y)\sim\mathbb{P}}\langle x,f(Y)\rangle}{\sqrt{\mathbb{E}_{Y\sim\mathbb{P}}\|f(Y)\|^2}}. \tag{25}$$

Proof. Consider $\bar{f}(Y) := \mathbb{E}_{Q\sim\mathrm{Haar}(n)}Q^\top f(Q\cdot Y)$. $\bar{f}$ is equivariant, since

$$\bar{f}(R\cdot Y) = \mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}(n)}Q^\top f(QR\cdot Y) = \mathop{\mathbb{E}}_{Q}RQ^\top f(Q\cdot Y) = R\bar{f}(Y). \tag{26}$$

And we have

$$\begin{aligned}\mathop{\mathbb{E}}_{(x,Y)\sim\mathbb{P}}\langle x,\bar{f}(Y)\rangle &= \mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}(n)}\mathop{\mathbb{E}}_{(x,Y)\sim\mathbb{P}}\langle x,Q^\top f(Q\cdot Y)\rangle\\ &= \mathop{\mathbb{E}}_{Q}\mathop{\mathbb{E}}_{(x,Y)\sim\mathbb{P}}\langle Qx,f(Q\cdot Y)\rangle\\ &= \mathop{\mathbb{E}}_{(x,Y)\sim\mathbb{P}}\langle x,f(Y)\rangle,\end{aligned}$$

and similarly for the denominator,

$$\begin{aligned}\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}\|\bar{f}(Y)\|^2 &= \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}\|\mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}(n)}Q^\top f(Q\cdot Y)\|^2\\ &\leq \mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}(n)}\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}\|Q^\top f(Q\cdot Y)\|^2\\ &= \mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}(n)}\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}\|f(Q\cdot Y)\|^2\\ &= \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}\|f(Y)\|^2\end{aligned}$$

where we have used Jensen's inequality followed by the invariance of the marginal distribution of $\mathbb{P}$ on Y. The result then follows. We note here that, because of the inequality in the last calculation, $\bar{f}$ can actually have *strictly larger* objective function in the correlation than f; on the other hand, the equivariant polynomials are included in the original maximization for the correlation, so the stated result is in fact an exact formula and not just a bound for $\mathrm{Corr}_{\leq D}(\mathbb{P})$. □

Remark 3.25. *When p is even, the reconstruction problem has an ambiguity since v is only identifiable up to a sign. In this case we can define $g(Y)$ as returning $v \otimes v$, or the entire rank-1 tensor $v^{\otimes p}$. More generally if we define the accuracy of g as $\langle g(Y), v^{\otimes p} \rangle$, the same equivariance argument applies.*

4 Tensorial Finite Free Cumulants and Invariant Bases

4.1 Frobenius Pairs

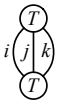
A special role in our techniques will be played by the following multigraph, which is called the "melon" in high energy physics.[6]

Definition 4.1. *The* Frobenius pair *F of degree p is the p-regular multigraph on two vertices with p parallel edges connecting the vertices.*

We call this the Frobenius pair because the associated graph moment is the Frobenius norm:

$$m_F(T) = \|T\|_F^2 = \sum_{i_1,\ldots,i_p \in [n]} T^2_{i_1,\ldots,i_p}. \tag{27}$$

For instance, for a 3-ary tensor we have



$$= \sum_{i,j,k} T^2_{ijk}\,.$$

For 2-ary tensors F is a 2-cycle, and we recover the matrix case

$$= \|T\|_F^2 = \mathrm{tr}(T^\top T)\,.$$

For multigraphs that include multiple Frobenius pairs as connected components, we use "Frobenii" as a plural and urge others to do the same.

As we have mentioned in the Introduction, the issue that Frobenii raise is that they are the only connected graph moment whose terms are not centered for generic (or, say, Gaussian as in the Wigner law) tensors T. This will create an obstruction to orthogonality for our initial definition of finite free cumulants, which we will need to circumvent with an additional centering operation.

[6] A related recursive structure gives the class of "melonic graphs" (the melon among them) that play an important role in other diagrammatic computations; see, e.g., [Evn21] or Chapter 4 of [Gur17].

4.2 Distinct-Index Graph Moments

In order to define tensor cumulants, we first follow recent work [Wei23] that modifies graph moments by requiring that the indices on the edges be distinct.

Definition 4.2. *Given a p-ary symmetric tensor T and a p-regular multigraph G with $|E| = b$ edges, define*

$$m_G^!(T) = \sum_{\substack{i \in [n]^E \\ i_1, \ldots, i_b \text{ distinct}}} \prod_{v \in V} T_{i(\partial v)}\,, \tag{28}$$

where the ! connotes distinctness.

Clearly $m_G^!(T) = 0$ unless $n \geq b$. As with ordinary graph moments, we define $m_G^!(T) = 1$ if G is empty.

Remark 4.3. *If T is a random tensor whose law is invariant under conjugation by orthogonal matrices, or even just conjugation by permutation matrices, then every term in the sum arising from $\mathbb{E}\, m_G^!(T)$ is equal. In that case, $\mathbb{E}\, m_G^!(T)$ is just the number of distinct tuples of indices, i.e., $n^{\underline{b}} = n(n-1)(n-2)\cdots(n-b+1)$, times the expectation of any one of them.*

Requiring distinct indices has an important consequence: except in a Frobenius pair, vertices cannot have matching edge neighborhoods. To deal with Frobenii we also define a centered version of $m_G^!$ as follows.

Definition 4.4. *Let T be a p-ary symmetric tensor and let $G = (V, E)$ be a p-regular multigraph with $|E| = b$ edges. Let U be the set of vertices forming Frobenius pairs, and for a given pair $\pi = \{u, v\} \subseteq U$ let $i(\pi) = i(\partial u) = i(\partial v)$ be the p indices appearing on the edges of π. Then define*

$$m_G^c(T) = \sum_{\substack{i \in [n]^E \\ i_1, \ldots, i_b \text{ distinct}}} \left[\prod_{v \in V \setminus U} T_{i(\partial v)} \prod_{\pi \subseteq U} \left(T_{i(\pi)}^2 - 1 \right) \right]. \tag{29}$$

This definition centers the Frobenii so that m_G^c has zero expectation in the Wigner model for all G. Recall that we abbreviate $\mathrm{Wig}(p, n, \sigma^2)$ as Wig when $\sigma^2 = 1$ and p and n are clear.

Proposition 4.5. *Let G be a nonempty p-regular multigraph, and let $W \sim$ Wig. Then*

$$\mathop{\mathbb{E}}_{W} m_G^c(W) = 0\,.$$

Proof. We will show that the summand in (29) has zero expectation for each fixed set of indices $i \in [n]^E$. If G has a connected component which is not a Frobenius pair, then the $W_{i(\partial v)}$ in the first product are all distinct and independent for different $v \in V \setminus W$. They are also independent from the $W_{i(\pi)}$, which are independent from each other for different Frobenii p.

Thus the expectation is a product of expectations. For any $i \in [n]^p$ we have $\mathbb{E}\, W_i = 0$, so the first product has zero expectation if $V \setminus U$ is nonempty. For the second products, the p indices in $i(\pi)$ are distinct, and therefore $\mathbb{E}\, W_{i(\pi)}^2 = 1$. Thus the second product has zero expectation if G has any Frobenii. □

Note what has happened here. As in Proposition 5.2, the only nonzero contributions to $\mathbb{E}_W m_G(W)$ correspond to even colorings of G's edges, where vertex neighborhoods $i(\partial v)$ are repeated. By requiring all indices to be distinct, we force the vertex neighborhoods to be distinct, except on Frobenius pairs where we center the normal distribution.

More generally, in planted models $Y = T + W$ where W is Wigner noise, the expectation of $m_G^c(Y)$ is simply the distinct-index moment of the signal T. This addresses one of the drawbacks of ordinary graph moments, namely that their expectations in the spiked model involves mixed terms with the signal at some vertices and the noise at others.

Proposition 4.6. *For any p-regular multigraph G and $T \in \mathrm{Sym}^p(\mathbb{R}^n)$ we have*

$$\mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} m_G^c(T + W) = m_G^!(T)\,. \tag{30}$$

Proof. Fix $i \in [n]^E$. Then as in Proposition 4.5, the entries $Y_{i(\partial v)}$ at distinct vertices v which are not part of Frobenius pairs are independent, and have expectation $T_{i(\partial v)}$. Similarly, the terms $Y_{i(\pi)}^2 - 1$ on distinct Frobenius pairs are independent, and have expectation

$$\mathbb{E}[(T+W)_{i(\pi)}^2 - 1] = T_{i(\pi)}^2 + 2T_{i(\pi)}\,\mathbb{E}\,W_{i(\pi)} + \mathbb{E}[W_{i(\pi)}^2] - 1 = T_{i(\pi)}^2\,,$$

which for a pair $\pi = (u, v)$ is the same as $T_{i(\partial u)}T_{i(\partial v)}$. Since the expectation of a product of independent variables is the product of their expectations, we are left with $\prod_{v \in V} T_{i(\partial v)} = m_G^!(T)$. □

Helpfully, these centered distinct-index moments have zero covariance in the Wigner model if their graphs are nonisomorphic. That is, they are orthogonal with respect to the inner product $\langle f, g\rangle = \mathbb{E}_{T \sim \mathrm{Wig}} f(T)\, g(T)$.

Proposition 4.7. *Let G, H be p-regular multigraphs. If they are not isomorphic, then*

$$\mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} m_G^c(W) m_H^c(W) = 0\,.$$

Proof. With the restriction that the edge indices in each graph are distinct, we have a sum of products like that in (29), but where $i(G)$ and $i(H)$ are each tuples of distinct indices. (This is not to be confused with $m_{G \sqcup H}^c$, in which case all $|E_G| + |E_H|$ indices would be distinct.)

Fix $i \in [n]^{E_G \sqcup E_H}$. We will show that if i makes a nonzero contribution to $\mathbb{E} m_G^c(W) m_H^c(W)$, then G and H are isomorphic. First, if $v \in V_G$ is not part of a Frobenius pair, then $T_{i(\partial v)}$ appears exactly once in G, and similarly if $v \in V_H$ is not in a Frobenius pair. Thus each such vertex in G must be matched with one in H, with corresponding indices on their edges. This establishes an isomorphism between the components of G with those of H that do not consist of Frobenii.

It remains to prove that G and H have the same number of Frobenii and are hence isomorphic. Suppose without loss of generality that G has more Frobenii than H does. Then for some Frobenius pair $\pi = (u, v)$ in G, the indices $i(\pi)$ do not appear in H. Since the entries of T are independent, this contributes $\mathbb{E}[T_{i(\pi)}^2 - 1] = 0$ to the product, in which case i's contribution to $\mathbb{E} m_G^c(W) m_H^c(W)$ is zero.

Thus if any i makes a nonzero contribution to the inner product, $G \cong H$. □

We can also compute the variances of the centered moments by counting a slightly nontraditional kind of graph automorphism. These automorphisms will appear again in the inner products of cumulants.

Definition 4.8. *Let $G = (V, E)$ be a multigraph. Give each edge $e \in E$ an arbitrary direction, and write it as an ordered pair (e_1, e_2). Then let an* edge automorphism *be a one-to-one mapping $\phi : E \to E$ such that, for all $e, e' \in E$, and for each $j, j' \in \{1, 2\}$, $\phi(e)_j = \phi(e')_{j'}$ if and only if $e_j = e'_{j'}$. In other words, ϕ preserves whether e and e' share their "heads" or their "tails", or whether one's head is the other's tail. Note that $\phi(e) = e$ means that $\phi(e)$ is either (e_1, e_2) or (e_2, e_1), where the latter reverses e's direction. Finally, let* $\mathrm{eAut}(G)$ *be the group of such mappings.*

Clearly an edge automorphism determines a traditional vertex automorphism, i.e., a mapping $\psi : V \to V$ such that $(\psi(u), \psi(v)) \in E$ if and only if $(u, v) \in E$. For simple graphs, the equivalence between the two types of automorphism is one-to-one. This more elaborate definition includes permuting parallel edges in a multigraph, as well as reversing the direction of a self-loop without moving its endpoint. In terms of the group $\mathrm{Aut}(G)$ of vertex automorphisms, we have

$$|\mathrm{eAut}(G)| = 2^{\#\{\text{self-loops in } G\}} \times \prod_{\substack{\text{bundles of } t \\ \text{parallel edges}}} t! \times |\mathrm{Aut}(G)| \,. \tag{31}$$

Proposition 4.9. *Let G be a p-regular multigraph with b edges. Then*

$$\mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} m_G^c(W)^2 = n^{\underline{b}} \, |\mathrm{eAut}(G)| \,. \tag{32}$$

Proof. First suppose that G has no Frobenii or self-loops. The nonzero contributions to the second moment consist of even colorings of the disjoint union of two copies of G, but now where the colors within each copy of G are distinct. The only such colorings are those like Figure 9, where each edge e on the left copy of G has a counterpart $\phi(e)$ on the right with the same index, and where each vertex v on the left has a counterpart $\psi(v)$ on the right with the same neighborhood. This describes a vertex automorphism and a permutation of each bundle of parallel edges. Along with the $n!/(n-b)!$ distinct labelings of the left copy, we obtain

$$\mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} m_G^c(W)^2 = n^{\underline{b}} \prod_{\substack{\text{bundles of } t \\ \text{parallel edges}}} t! \times |\mathrm{Aut}(G)| \,.$$

Now suppose that some vertex v has t self-loops. Their indices can be permuted, but we already have the factor $t!$ since these count as a bundle of parallel edges. The other effect is that the entry $T_{i(v)}$ has variance larger than 1 since some indices are repeated. Specifically, if $i(v)$ contains t pairs of repeated indices, by Proposition B.1 we have $\mathbb{E}\, T_{i(v)}^2 = 2^t$. Taking the product over all vertices gives the factor $2^{\#\{\text{self-loops in } G\}}$.

Finally we consider the possibility that G contains one or more Frobenii. As in Proposition 4.7, for each Frobenius pair π in the left copy of G, its indices $i(\pi)$ must appear in a counterpart pair on the right; otherwise $T_{i(\pi)}$ is independent of the other entries, and centering gives a factor of $\mathbb{E}[T_{i(\pi)}^2 - 1] = 0$. The p parallel edges of π and its counterpart can be matched in $p!$ ways, and for each matching they contribute a factor of

$$\mathbb{E}\left[\left(T_{i(\pi)}^2 - 1\right)^2\right] = 2 \,,$$

where we use the fact that $T_{i(\pi)} \sim \mathcal{N}(0, 1)$. Finally, if we have t Frobenius pairs, they can be matched with each other in $t!$ ways, and their total contribution is $(2p!)^t t!$. But this is also the number of edge automorphisms of the disjoint union of t Frobenii, so (32) continues to hold. □

4.3 Finite Free Cumulants and Additive Free Convolution

The orthogonality of the distinct-index graph moments suggests that they are a good basis for the space of polynomial functions of T with respect to the Gaussian inner product. However, they are not invariant with respect to orthogonal basis changes. This motivates the following definition, where we symmetrize them with random orthogonal rotations.

Definition 4.10. *Let T be a symmetric p-ary tensor and let G be a p-regular multigraph. Define the* free cumulant $\kappa_G(T)$ *as*

$$\kappa_G(T) = \mathop{\mathbb{E}}_{Q}\left[m^!_G(Q \cdot T)\right], \tag{33}$$

and the centered free cumulant $\kappa^c_G(T)$ *as*

$$\kappa^c_G(T) = \mathop{\mathbb{E}}_{Q}\left[m^c_G(Q \cdot T)\right], \tag{34}$$

where Q is a Haar-random orthogonal matrix. If G is the empty graph then $\kappa_G = \kappa^c_G = 1$.

Remark 4.11. *Recall that $m^!_G(T)$ and $m^c_G(T)$ are both summations over assignments of distinct indices to the edges of G. If there are b edges, then there are $n^{\underline{b}}$ such assignments. After symmetrizing T to $Q \cdot T$ and taking the expectation over $Q \sim \mathrm{Haar}(n)$, by symmetry each of these terms will be equal. If we view the original summation as splitting each edge in half and connecting either end to the matrix $e_i \otimes e_i$, where i is the label that that edge gets in that term of the summation, then the symmetrization yields an expectation where the $e_1, \ldots, e_n$ are replaced by a Haar-random orthonormal basis of $\mathbb{R}^n$. In notation, we may fix $i \in [n]^E$ an arbitrary distinct index labelling of the edges, in which case we have*

$$\kappa_G(T) = n^{\underline{b}} \mathop{\mathbb{E}}_{Q \sim \mathrm{Haar}} \prod_{v \in V} (Q \cdot T)_{i(\partial v)}, \tag{35}$$

$$\kappa^c_G(T) = n^{\underline{b}} \mathop{\mathbb{E}}_{Q \sim \mathrm{Haar}} \prod_{v \in V(G \setminus \mathrm{Frob}(G))} (Q \cdot T)_{i(\partial v)} \prod_{F \in \mathrm{Frob}(G)} ((Q \cdot T)^2_{i(F)} - 1). \tag{36}$$

These functions κ_G are manifestly invariant, and κ_G coincides with κ^c_G unless G includes one or more Frobenii. We call them free cumulants because, if $p = 2$ and G is the cycle of size t, they coincide—up to an appropriate scaling factor—with the classic free cumulant κ_t of free probability in the limit $n \to \infty$ (e.g. [MFC+19, Eq.27]). If G is a disjoint union of two or more cycles, the $n \to \infty$ limit yields the so-called *higher-order free cumulants* [CMSS06] which [Sem24] used recently for orthogonally-invariant denoising problems in matrices.

One of the most useful properties of the free cumulants is additivity [NŚ11]. In classical statistics, cumulants are polynomials κ in the moments of a random variable with the property that, if A and B are independent random variables, then the cumulant of their convolution $A + B$ is the sum of their cumulants: $\kappa(A + B) = \kappa(A) + \kappa(B)$. For instance, the first, second, and third cumulants are the expectation $\kappa_1(A) = \mathbb{E}[A]$, the variance $\kappa_2(A) = \mathbb{E}[A^2] - \mathbb{E}[A]^2$, and $\kappa_3(A) = \mathbb{E}[A^3] - 3\,\mathbb{E}[A^2]\,\mathbb{E}[A] + 2\,\mathbb{E}[A]^3$. The tth cumulant κ_t is homogeneous in the sense that it is a linear combination of products of moments with total order t.

In the nonabelian setting of free probability (e.g. [MS17]), A and B are matrix-valued random variables in $\mathbb{R}^{n \times n}$, and a central question is to determine the typical spectrum of $A + Q^\top B Q$ where

$Q \in \mathcal{O}(n)$ is a Haar-random orthogonal matrix. (One can also study complex-valued matrices where R is unitary.) While $A + Q^\top BQ$ is a random matrix even for fixed A and B, in the limit $n \to \infty$ its empirical spectral distribution converges to a particular probability measure under mild conditions provided the same convergence holds for A and B. Moreover, this limiting measure is a function of the limiting spectral distributions of A and B individually. If we call these distributions α and β respectively, then the limiting law of the spectrum of $A + Q^\top BQ$ is written $\alpha \boxplus \beta$ and called the *free convolution* of α and β. The free cumulants are like the classical cumulants in that they are polynomials in the moments of a measure, but they are additive with respect to this free convolution operation rather than ordinary convolution (which is the operation on distributions induced by addition of independent scalar random variables), i.e.,

$$\kappa_t(\alpha \boxplus \beta) = \kappa_t(\alpha) + \kappa_t(\beta)\,. \tag{37}$$

The analogy with classical cumulants is that, since conjugating by Q randomizes B's eigenvectors, A and $Q^\top BQ$ are as independent as two matrices can be given their spectra. We say that they are *freely independent*.

We call the κ_G free cumulants of a tensor both because they coincide with matrix free cumulants in the limit $n \to \infty$ when G is a cycle, and because they satisfy an exact additivity property even for finite n.[7] The latter is closely related to the additivity property for distinct index moments we gave in Proposition 4.6.

Proposition 4.12. *Let $G = (V, E)$ be a connected p-regular graph with $b = |E|$ edges, and let A and B be symmetric p-ary tensors. Let $Q \in \mathcal{O}(n)$ be a Haar-random orthogonal matrix. Then*

$$\mathop{\mathbb{E}}_{Q} \kappa_G(A + Q \cdot B) = \kappa_G(A) + \kappa_G(B)\,. \tag{38}$$

More generally, if G is not necessarily connected,

$$\mathop{\mathbb{E}}_{Q} \kappa_G(A + Q \cdot B) = \sum_{G_A \sqcup G_B = G} \frac{n^{\underline{b}}}{n^{\underline{b_A}} n^{\underline{b_B}}}\, \kappa_{G_A}(A)\, \kappa_{G_B}(B). \tag{39}$$

Here the sum is over pairs of graphs $G_A = (V_A, E_A)$ and $G_B = (V_B, E_B)$ such that $G = G_A \sqcup G_B$, and where $b_A = |E_A|$ and $b_B = |E_B|$ with $b_A + b_B = b$. In particular, G_A and G_B are disjoint unions of connected components of G.

Proof. We start with connected graphs. We will prove a partial additivity for the distinct-index moments,

$$\mathop{\mathbb{E}}_{Q} m^!_G(A + Q \cdot B) = m^!_G(A) + \kappa_G(B)\,. \tag{40}$$

Symmetrizing and replacing A with $R \cdot A$ for a Haar-random R, and using the fact that QR is also Haar-random, yields (38).

First we give a more diagrammatic picture of $m^!_G$. Extracting a given product of entries in (28) corresponds to placing a basis vector e_i on each edge i—or rather two copies of e_i, pointing in each direction—and contracting these vectors with the tensors at their endpoints as in Figure 6.

[7] While it seems technical to verify, it is natural to conjecture that in the case $p = 2$, our κ_{C_k} for C_k the k-cycle should be closely related to the finite free cumulants defined by [AP18] in terms of characteristic polynomials.

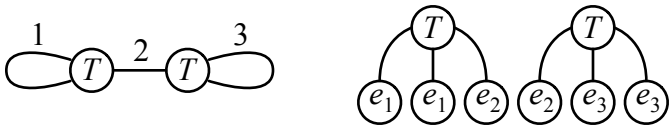


Figure 6: A given term in (28) corresponds to placing basis vectors e_i on each edge i, and contracting them with i's endpoints. That is, $m!_G(T)$ is the sum of the diagram on the right over all distinct tuples e_1, e_2, e_3 in the orthogonal basis.

Expanding $m^!_G(A+B)$ gives $2^{|V|}$ terms, where A appears on some vertices and B appears on the others. (Except for the distinct-index requirement, these cross-terms are mixed graph moments as in Definition 3.3.) Call these two sets of vertices V_A and V_B respectively.

Now consider the effect of replacing B with $Q \cdot B$. This applies the projection operator $\Pi_\ell = \mathbb{E}\, Q^{\otimes \ell}$ defined in (108) with $\ell = d|V_B|$, which is a linear combination of outer products of matching vectors. However, unlike in the proof of Theorem 3.2, these matchings only apply to the half-edges incident to vertices in V_B. Since G is connected, if V_A and V_B are both nonempty, there is at least one edge i in the cut between them. But in each matching contributing to Π_ℓ, the copy of e_i incident to V_B is matched with e_j for some other edge j incident to V_B as shown in Figure 7. Since $\langle e_i, e_j \rangle = 0$, any such term is zero.

This implies that the only nonzero contributions to $m^!_G(A + Q \cdot B)$ come from the terms where V_A or V_B is empty, i.e., where the vertices are either all labeled with A or all labeled with $Q \cdot B$. But these terms are exactly $m^!_G(A)$ and $\kappa_G(B)$ as stated.

For graphs with multiple connected components, the same argument implies that each component is labeled entirely with A or with B. Call the unions of these components G_A and G_B respectively; then each such partition contributes $\kappa_{G_A}(A)\,\kappa_{G_B}(B)$. Since all the indices in G are distinct, not just those within G_A or G_B, we need to divide the prefactor $n^{\underline{b}}$ appearing in κ_G (see Remark 4.11) by the analogous prefactors $n^{\underline{b_A}}$ and $n^{\underline{b_B}}$ for κ_{G_A} and κ_{G_B}. This yields (39). □

In fact, this additivity phenomenon is more general and a version of it is enjoyed both by the centered κ^c_G and further generalizations thereof. To present this version, it is convenient to work over a more general family of centered cumulants.

Definition 4.13. *Write* $\mathrm{conn}(G)$ *for the set of connected components of* G. *Given* $x \in \mathbb{R}^{\mathrm{conn}(G)}$, *define*

$$m^c_G(T;x) := \sum_{\substack{i\in[n]^E \\ i_1,\dots,i_b \text{ distinct}}} \prod_{C\in\mathrm{conn}(G)} \left(\prod_{v\in C} T_{i(\partial v)} - x_C \right),$$

$$\kappa^c_G(T;x) := \mathop{\mathbb{E}}_Q m^c_G(Q\cdot T;x).$$

Our definition from earlier is $m^c_G(T) = m^c_G(T;x)$ where $x_C = -\mathbf{1}\{C \text{ is a Frobenius pair}\} := -\mathbf{1}_{\mathrm{Frob}}$.

Proposition 4.14 (General additivity after centering)**.** *Suppose* $x = x_A + x_B$. *Then,*

$$\mathop{\mathbb{E}}_Q \kappa^c_G(A + Q\cdot B;x) = \sum_{G_A \sqcup G_B = G} \frac{n^{\underline{b}}}{n^{\underline{b_A}} n^{\underline{b_B}}} \kappa^c_{G_A}(A;x_A)\kappa^c_{G_B}(B;x_B). \tag{41}$$

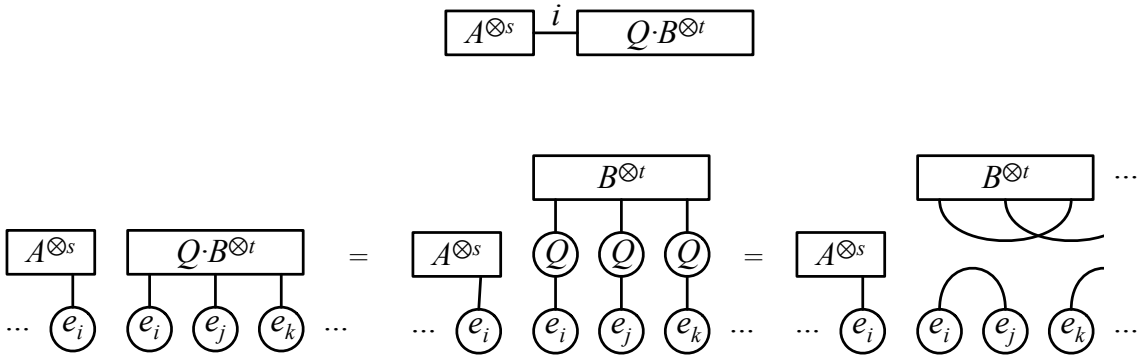


Figure 7: Illustrating the proof of Proposition 4.12. As in (40), we consider $m_G^!(A + Q \cdot B)$. Expanding the tensor product creates $2^{|V|}$ cross-terms, where vertices in V_A and V_B are labeled with A and $Q \cdot B$ respectively. Applying Q to the half-edges of V_B creates a linear combination of matchings. But if any edge i crosses from V_A to V_B, this matching creates an inner product between its vector e_i with some other e_j, contributing $\langle e_i, e_j \rangle$. Thus only partitions where V_A and V_B are unions of connected components of G contribute to $m_G^!(A + Q \cdot B)$. In particular, if G is connected then either $V_A = V$ and $V_B = \varnothing$ or vice versa.

The proof is essentially identical to that of Proposition 4.12.

Proposition 4.12 is recovered as the special case $x = x_A = x_B = 0$, which is formally pleasing but not useful for analyzing low-degree polynomials, because one of the free cumulants is the symmetrized distinct index Frobenius pair $\sum_{i \text{ distinct}} (Q \cdot T)_i^2$, which is non-zero for any non-zero tensor. But, by choosing x, x_A, and x_B prudently, we may arrange for even the expected centered free cumulants with G including Frobenii to vanish exactly.

In particular, when we are using this with $B \sim \mathrm{Wig}$, then it is most natural to take $x_B = x = -\mathbf{1}_{\mathrm{Frob}}$ and $x_A = 0$, because of the following.

Proposition 4.15. *For any G, $\mathbb{E}_{W \sim \mathrm{Wig}} \kappa_G^c(W) = 0$.*

Proof. By orthogonal invariance of the law Wig, we have $\mathbb{E}_{W \sim \mathrm{Wig}} \kappa_G^c(W) = \mathbb{E}_{W \sim \mathrm{Wig}} m_G^c(W)$, and the result then follows from Proposition 4.5. □

However, for other applications other ways of "distributing" the correction x are also useful, as we will see in Section 6 on computational central limit theorems and distinguishing Wigner from Wishart tensors.

4.4 Graph Moment Expansion of Finite Free Cumulants

Since the κ_G are invariant, by Theorem 3.2 they must admit an expansion in the graph moments m_H. Below we give an explicit such expansion, which will also prove useful below.

Definition 4.16 (Graph Weingarten function)**.** *Say that a perfect matching μ of $[dp]$* realizes *a p-regular multigraph G on d unlabelled vertices if, when μ is viewed as a graph on $[dp]$ and the sets of vertices $\{1, \ldots, p\}, \{p+1, \ldots, 2p\}$ and so forth are each identified into a single vertex, then the*

resulting graph is isomorphic to G. For two p-regular graphs G, H on a vertices, we write

$$\mathrm{Wg}_{G,H} := \sum_{\substack{\mu \text{ realizes } G \\ \nu \text{ realizes } H}} \mathrm{Wg}_{\mu,\nu}. \tag{42}$$

As a convention, if G and H have different numbers of vertices, we set $\mathrm{Wg}_{G,H} := 0$.

To attend to the normalizations that will appear in the calculations below, the following will be useful.

Proposition 4.17. *The number of matchings realizing $G \in \mathcal{G}_{d,p}$ is $p!^d d!/|\mathrm{eAut}(G)|$.*

Proof. We may expand the claimed quantity as

$$\frac{p!^d d!}{|\mathrm{eAut}(G)|} = \frac{d!}{|\mathrm{Aut}(G)|} \cdot \frac{p!^d}{2^{\#\{\text{self-loops in } G\}} \prod_{\text{bundles of } t \text{ parallel edges in } G} t!}. \tag{43}$$

Here the first term is the number of permutations of vertices that map any matching realizing G to another different matching realizing G, and the second term is the same for permutations of half-edges around each vertex. □

Lemma 4.18. *For any p-regular multigraph G with d vertices and b edges,*

$$\kappa_G(T) = \frac{n^{\underline{b}}}{p!^d d!} \cdot |\mathrm{eAut}(G)| \cdot \sum_H \mathrm{Wg}_{G,H} m_H(T). \tag{44}$$

The sum may be taken over all p-regular multigraphs H, but only those with d vertices will contribute.

Proof. The graph moments may be defined in terms of matching vectors: if μ is a matching realizing G, then $m_G(T) = \langle T^{\otimes d}, w(\mu) \rangle$. To be more explicit, for $i \in [n]^{pd}$, call i *compatible with* μ if i is constant on each pair matched under μ. Then, $w(\mu) = \sum_{i \text{ compatible with } \mu} e_i$, which recovers Definition 1.3 of $m_G(T)$.

Let us describe a variant of this for the distinct index moments. We say that i is *distinctly compatible with* μ if i is compatible with μ and if i is distinct on distinct pairs in μ: that is, the level sets of i are exactly the pairs in μ. Then if we define

$$w^!(\mu) := \sum_{\substack{i \text{ distinctly} \\ \text{compatible with } \mu}} e_i \,,$$

then by the same token we have

$$m^!_G(T) = \langle T^{\otimes d}, w^!(\mu) \rangle.$$

Moreover, in the symmetrization we may take the adjoint of the application of Q and find

$$\begin{aligned}
\kappa_G(T) &= \mathop{\mathbb{E}}_{Q} m_G^!(Q \cdot T) \\
&= \mathop{\mathbb{E}}_{Q} \langle Q \cdot T^{\otimes d}, w^!(\mu) \rangle \\
&= \mathop{\mathbb{E}}_{Q} \langle T^{\otimes d}, Q^\top \cdot w^!(\mu) \rangle \\
&= \mathop{\mathbb{E}}_{Q} \langle T^{\otimes d}, Q^{\otimes dp} w^!(\mu) \rangle \\
&= \sum_{\substack{i \text{ distinctly} \\ \text{compatible with } \mu}} \mathop{\mathbb{E}}_{Q} \langle T^{\otimes d}, Q^{\otimes dp} e_i \rangle
\end{aligned}$$

and using that every term is equal and does not depend on i (as may be seen by multiplying Q by a permutation matrix),

$$\begin{aligned}
&= n^{\underline{b}} \mathop{\mathbb{E}}_{Q} \langle T^{\otimes d}, Q^{\otimes dp} e_\mu \rangle \\
&= n^{\underline{b}} \mathop{\mathbb{E}}_{Q} \langle T^{\otimes d}, \Pi_d e_\mu \rangle
\end{aligned}$$

where we write $e_\mu = e_i$ for some choice of $i = i(\mu)$ compatible with μ. Next, substituting in the Weingarten function expression for Π_d (see Appendix C) gives

$$= n^{\underline{b}} \sum_{\rho,\nu} \mathrm{Wg}_{\rho,\nu} \langle w(\rho), e_\mu \rangle \langle T^{\otimes d}, w(\nu) \rangle$$

and we have $\langle w(\rho), e_\mu \rangle = \mathbf{1}\{\rho = \mu\}$, so this reduces to

$$\begin{aligned}
&= n^{\underline{b}} \sum_{\nu} \mathrm{Wg}_{\mu,\nu} \langle T^{\otimes d}, w(\nu) \rangle \\
&= n^{\underline{b}} \sum_{\nu} \mathrm{Wg}_{\mu,\nu} \, m_{G(\nu)}(T),
\end{aligned}$$

where, as before, $G(\nu)$ is the graph resulting from identifying groups of p vertices suitably in ν. We notice that a given $m_H(T)$ appears many times in this formula, once for each ν such that $H = G(\nu)$, i.e., for each ν that realizes H. We can give a proper expansion of $\kappa_G(T)$ in graph moments by combining these terms,

$$= n^{\underline{b}} \sum_{H} \left(\sum_{\nu \text{ realizes } H} \mathrm{Wg}_{\mu,\nu} \right) m_H(T) \,.$$

This expansion does not depend on the choice of μ realizing G. Indeed, this independence even holds over the inner summations giving the coefficient of each $m_H(T)$: choosing a different μ does not affect the terms in this summation, since $\mathrm{Wg}_{\mu,\nu} = \mathrm{Wg}_{\pi\mu,\pi\nu}$ for a permutation π of the indices

being matched. By Proposition 4.17, we may then rewrite

$$\begin{aligned}\kappa_G(T) &= \frac{n^{\underline{b}}}{\#\{\mu : \mu \text{ realizes } G\}} \sum_H \mathrm{Wg}_{G,H} m_H(T) \\ &= \frac{n^{\underline{b}}}{p!^d d!} \cdot |\mathrm{eAut}(G)| \cdot \sum_H \mathrm{Wg}_{G,H} m_H(T),\end{aligned}$$

as claimed. □

Lastly, let us derive the simple relationship between the κ_G^c and the κ_G.

Proposition 4.19. *For any p-regular multigraph G on d vertices and b edges,*

$$\kappa_G^c(T) = \sum_{S \subseteq \mathrm{Frob}(G)} (-1)^{|S|} (n - b + p|S|)^{\underline{p|S|}} \kappa_{G \setminus S}(T). \tag{45}$$

Proof. Expanding the product in the definition of $m_G^c(T)$, we have

$$m_G^c(T) = \sum_{S \subseteq \mathrm{Frob}(G)} (-1)^{|S|} (n - b + p|S|)^{\underline{p|S|}} m_{G \setminus S}(T), \tag{46}$$

since there are $p|S|$ indices in the summation forced to be distinct from the others but otherwise playing no rule in the summands. Projecting each side to invariant polynomials (by averaging over Q the value applied to $Q \cdot T$) gives the result. □

4.5 Inner Products and Approximate Orthogonality

We now turn to computing the inner products $\mathbb{E}_{W \sim \mathrm{Wig}} \kappa_G^c(W) \kappa_H^c(W)$, which will show that the κ_G^c form an approximately orthogonal basis for the invariant polynomials. We first establish a simple preliminary.

Proposition 4.20. *Suppose $|V(G)| \geq |V(H)|$. Then, $\mathbb{E}_{W \sim \mathrm{Wig}} m_G^c(W) m_H(W) = 0$ unless G and H are isomorphic, and*

$$\mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} m_G^c(W) m_G(W) = \mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} m_G^c(W)^2 = n^{\underline{b}} \cdot |\mathrm{eAut}(G)|. \tag{47}$$

The proof is a simple extension of the argument for Propositions 4.7 and 4.9 on the values of $\mathbb{E}_{W \sim \mathrm{Wig}} m_G^c(W) m_H^c(W)$, as changing m_H^c to m_H does not change the value of these inner products.

The following is the main calculation of the inner products we are interested in. In fact, these inner products are very close to the values of the graph Weingarten function.

Lemma 4.21. *For $G, H \in \mathcal{G}_{d,p}$, if $|V(G)| \neq |V(H)|$, then $\mathbb{E}_{W \sim \mathrm{Wig}} \kappa_G^c(W) \kappa_H^c(W) = 0$. If $|V(G)| = |V(H)| = d$, then*

$$\mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} \kappa_G^c(W) \kappa_H^c(W) = \frac{(n^{\underline{b}})^2}{p!^d d!} \cdot |\mathrm{eAut}(G)| \cdot |\mathrm{eAut}(H)| \cdot \mathrm{Wg}_{G,H}. \tag{48}$$

Proof. Suppose without loss of generality that $|V(G)| \geq |V(H)|$. We have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} \kappa_G^c(W)\kappa_H^c(W) \\
&= \mathop{\mathbb{E}}_{\substack{W\sim\mathrm{Wig}\\U,V\sim\mathrm{Haar}}} m_G^c(U\cdot W)m_H^c(V\cdot W) \\
&= \mathop{\mathbb{E}}_{\substack{W\sim\mathrm{Wig}\\V\sim\mathrm{Haar}}} m_G^c(W)m_H^c(V\cdot W) \\
&= \mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} m_G^c(W)\kappa_H^c(W)
\end{aligned}$$

and using Proposition 4.19 followed by Lemma 4.18, we find

$$\begin{aligned}
&= \sum_{S\subseteq\mathrm{Frob}(H)} (-1)^{|S|}(n-b+p|S|)^{\underline{p|S|}} \mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} m_G^c(W)\kappa_{H\setminus S}^c(W) \\
&= \sum_{S\subseteq\mathrm{Frob}(G)} (-1)^{|S|}(n-b+p|S|)^{\underline{p|S|}} \frac{n^{\underline{b}}}{p!^d d!}|\mathrm{eAut}(H\setminus S)| \sum_K \mathrm{Wg}_{H\setminus S,K} \mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} m_G^c(W)m_K(W)
\end{aligned}$$

By Proposition 4.20, we immediately find that all terms are zero unless $|V(G)| = |V(H)|$. If this holds, then all terms with $S \neq \varnothing$ are zero, and among those the only non-zero term has $K = G$, and we are left with simply

$$\begin{aligned}
&= \frac{n^{\underline{b}}}{p!^d d!}|\mathrm{eAut}(H)| \sum_K \mathrm{Wg}_{G,H} \mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} m_G^c(W)m_G(W) \\
&= \frac{(n^{\underline{b}})^2}{p!^d d!} \cdot |\mathrm{eAut}(G)| \cdot |\mathrm{eAut}(H)| \cdot \mathrm{Wg}_{G,H},
\end{aligned}$$

completing the proof. □

Finally, we may construct a nearly *orthonormal* basis by normalizing these polynomials as follows.

Definition 4.22. *For each G a p-regular multigraph, define*

$$\widehat{\kappa_G^c}(T) := \frac{n^{b/2}}{n^{\underline{b}}\sqrt{|\mathrm{eAut}(G)|}}\kappa_G^c(T). \tag{49}$$

For the sake of simplicity, we do not try to exactly normalize these polynomials to have norm 1. On the other hand, they obey the following conditioning property.

Lemma 4.23. *If $|V(G)| \neq |V(H)|$, then $\mathbb{E}_{W\sim\mathrm{Wig}}\widehat{\kappa_G^c}(W)\widehat{\kappa_H^c}(W) = 0$. Let $M \in \mathbb{R}^{\mathcal{G}_{d,p}\times\mathcal{G}_{d,p}}$ have entries $M_{G,H} := \mathbb{E}_{W\sim\mathrm{Wig}}\widehat{\kappa_G^c}(W)\widehat{\kappa_H^c}(W)$. Suppose that $d \leq \sqrt{n/2p^2}$. Then,*

$$\frac{1}{2} \leq \lambda_{\min}(M) \leq \lambda_{\max}(M) \leq 2. \tag{50}$$

Proof. By Lemma 4.21,

$$M_{G,H} = n^b \cdot \frac{\sqrt{|\mathrm{eAut}(G)| \cdot |\mathrm{eAut}(H)|}}{p!^d d!} \mathrm{Wg}_{G,H}. \tag{51}$$

The result then follows by the result of Corollary C.4 on the conditioning of the graph Weingarten function. □

Roughly, the idea of the normalization in Definition 4.22 is that $\mathrm{Wg} \asymp n^{-\ell/2}\mathbb{1}$, so if we want the diagonal of M in the proof above to be roughly constant, we must make the diagonal entries normalized by $\#\{\mu \text{ realizes } G\}$ for each G, which is given in terms of $|\mathrm{eAut}(G)|$ in Proposition 4.17. See also Remark C.3 for an argument showing that the upper bound on d is necessary for such a condition to hold.

4.6 Advantage Bound for Invariant Detection Problems

Using the tools developed above, we derive the following general bound on the advantage between the Wigner tensor law and any invariant distribution, which for the purposes of applications to hypothesis testing is the final prize of our work with finite free cumulants.

Corollary 4.24. *Let $\mathbb{P}$ be any probability measure on symmetric tensors that is orthogonally invariant. Suppose that $D \le \sqrt{n/2p^2}$. Then,*

$$\mathrm{Adv}_{\le D}(\mathrm{Wig}, \mathbb{P})^2 \asymp \sum_{d=0}^{D} \frac{1}{n^{pd/2}} \sum_{G \in \mathcal{G}_{d,p}} \frac{1}{|\mathrm{eAut}(G)|} \left(\mathop{\mathbb{E}}_{Y \sim \mathbb{P}} \kappa_G^c(Y) \right)^2, \tag{52}$$

where the sum is over p-regular multigraphs up to isomorphism.

Concretely, the constant in an upper bound may be taken to be 2, and the constant in a lower bound to be $\frac{1}{2}\exp(-\frac{1}{8}) \ge \frac{1}{3}$.

Proof. Recall that, by Proposition 3.22, under these assumptions we have the formula for the advantage

$$\mathrm{Adv}_{\le D}(\mathrm{Wig}, \mathbb{P}) = \sup_{\substack{f \in \mathbb{R}[Y]_{\le D} \\ f \text{ invariant} \\ \mathbb{E}_{W \sim \mathrm{Wig}} f(W)^2 \neq 0}} \frac{\mathbb{E}_{Y \sim \mathbb{P}} f(Y)}{\sqrt{\mathbb{E}_{W \sim \mathrm{Wig}} f(W)^2}}. \tag{53}$$

By Lemma 4.23, the $\widehat{\kappa_G^c}$ over $G \in \mathcal{G}_{0,p} \sqcup \cdots \sqcup \mathcal{G}_{d,p}$ are a basis for the invariant polynomials in $\mathbb{R}[Y]_{\le D}$. Thus, consider f in the optimization stated, and write the expansion

$$f(T) = \sum_{G \in \mathcal{G}_{0,p} \sqcup \cdots \sqcup \mathcal{G}_{d,p}} \alpha_G \cdot \widehat{\kappa_G^c}(T). \tag{54}$$

Again by Lemma 4.23, we have

$$\frac{1}{2}\|\alpha\|^2 = \frac{1}{2} \sum_{G \in \mathcal{G}_{0,p} \sqcup \cdots \sqcup \mathcal{G}_{d,p}} \alpha_G^2 \le \mathbb{E}_{W \sim \mathrm{Wig}} f(W)^2 \le 2 \sum_{G \in \mathcal{G}_{0,p} \sqcup \cdots \sqcup \mathcal{G}_{d,p}} \alpha_G^2 = 2\|\alpha\|^2. \tag{55}$$

Define

$$\beta_G := \mathop{\mathbb{E}}_{Y\sim\mathbb{P}} \widehat{\kappa^c_G}(Y). \tag{56}$$

Then, we have by linearity

$$\mathop{\mathbb{E}}_{Y\sim\mathbb{P}} f(Y) = \langle \alpha, \beta\rangle, \tag{57}$$

and thus by basic linear algebra, with lower and upper bound constants $\frac{1}{2}$ and 2 respectively,

$$\begin{aligned}
\mathrm{Adv}_{\le D}(\mathrm{Wig}, \mathbb{P})^2 &\asymp \left(\sup_{\|\alpha\|\neq 0} \frac{\langle \alpha, \beta\rangle}{\|\alpha\|}\right)^2 \\
&= \|\beta\|^2 \\
&= \sum_G \left(\mathop{\mathbb{E}}_{Y\sim\mathbb{P}} \widehat{\kappa^c_G}(Y)\right)^2
\end{aligned}$$

and substituting in the normalization of $\widehat{\kappa^c_G}$,

$$\begin{aligned}
&= \sum_G \frac{n^{b_G}}{(n^{\underline{b_G}})^2} \frac{1}{|\mathrm{eAut}(G)|} \left(\mathop{\mathbb{E}}_{Y\sim\mathbb{P}} \kappa^c_G(Y)\right)^2 \\
&= \sum_{d=0}^{D} \frac{n^{pd/2}}{(n^{\underline{pd/2}})^2} \sum_{G\in\mathcal{G}_{d,p}} \frac{1}{|\mathrm{eAut}(G)|} \left(\mathop{\mathbb{E}}_{Y\sim\mathbb{P}} \kappa^c_G(Y)\right)^2
\end{aligned}$$

and finally by Proposition D.1 on the falling factorial,

$$\asymp \sum_{d=0}^{D} \frac{1}{n^{pd/2}} \sum_{G\in\mathcal{G}_{d,p}} \frac{1}{|\mathrm{eAut}(G)|} \left(\mathop{\mathbb{E}}_{Y\sim\mathbb{P}} \kappa^c_G(Y)\right)^2,$$

as stated, with lower and upper bound constants $\exp(-\frac{1}{8})$ and 1, respectively, in the last step. □

At a high level, this bound is making the following reasonable claim: the $\mathbb{E}\kappa^c_G(Y)$ are quantities that are zero when $Y \sim \mathrm{Wig}$, and their magnitude for a given distribution—a kind of "cumulant distance" between a distribution and the Wigner law—controls the difficulty of hypothesis testing against a Wigner null hypothesis. Let us give some more intuition about what "critical scaling" of the cumulants—around the threshold of computational hardness—this implies.

As we will see, in the models we will consider all cumulants $\mathbb{E}\kappa^c_G(T)$ with $|V(G)| = d$ will have roughly the same size. The number of p-regular multigraphs on d vertices, as we will see later in Proposition 5.9, grows roughly as $|G_{d,p}| \approx d^{O(d)}$ for fixed d. Thus, in order for the advantage to be bounded, it will suffice to have $|\mathbb{E}\kappa^c_G(T)| \lesssim n^{pd/4}/\mathrm{poly}(d)$, as then the sum will be a truncation of a convergent geometric series. This is sensible: recall that the cumulant is a sum of $n^{pd/2}$ centered terms (as this is the number of labellings of the edges of G by $[n]$), so $n^{pd/4}$ is the "usual" value of the cumulants of a law $\mathbb{P}$ whose entries are on the same scale as Wig.

4.7 Bases for Vector-Valued Equivariant Polynomials

Finally, we develop some parallel theory for the case of equivariant polynomials and open multigraphs. For the sake of simplicity, we only work with 1-open multigraphs, i.e., ones with a single open edge. Similar results should hold for general t-open multigraphs, albeit with more elaborate combinatorics. We will be brisk in our presentation, as the ideas are similar to the case of closed multigraphs.

We slightly abuse notation and reuse the notations $m_G, m_G^!, m_G^c, \kappa_G$, κ_G^c, and $\mathcal{G}_{d,p}$. However, to remind the reader that we are working with vector-valued functions and open multigraphs, we replace "G" with "$G \to$", a visual indication of the one "loose" or "dangling" edge on a 1-open multigraph. We also use:

Definition 4.25 (1-open p-regular multigraphs)**.** *Write $\mathcal{G}_{d,p\to}$ for the set of (non-isomorphic) 1-open p-regular multigraphs on d unlabelled vertices. Equivalently, these may be viewed as multigraphs on $d+1$ vertices, all of which have degree p except for one, which has degree $p-1$.*

Remark 4.26 (Arity parity)**.** *In order for $\mathcal{G}_{d,p\to}$ to not be empty, pd must be odd, and in particular p, the tensor arity, must be odd.*

For $G \in \mathcal{G}_{d,p\to}$, we write $m_{G\to} : \mathrm{Sym}^p(\mathbb{R}^n) \to \mathbb{R}^n$ for the ordinary 1-open graph moments previously defined in Definition 3.10. We also write $m_{G\to}^!(T) \in \mathbb{R}^n$ for the same quantities with all indices in the summation involved restricted to be distinct: letting e_0 be the open edge of G,

$$m_{G\to}^!(T)_i := \sum_{\substack{j\in[n]^E \\ j(e_0)=i \\ j_1,\dots,j_b \text{ distinct}}} \prod_{v\in V} T_{j(\partial v)}\,. \tag{58}$$

Note that this restriction means that the summation defining $m_{G\to}^!(T)_i$ depends on i, since all other indices in this summation must be different than i. We similarly define the centered version of these summations, where our handling of Frobenius pairs only affects connected components other than the one containing e_0:

$$m_{G\to}^c(T)_i := \sum_{\substack{j\in[n]^E \\ j(e_0)=i \\ j_1,\dots,j_b \text{ distinct}}} \prod_{v\in V(G\setminus \mathrm{Frob}(G))} T_{j(\partial v)} \prod_{F\in \mathrm{Frob}(G)} (T^2_{j(F)} - 1). \tag{59}$$

We likewise define the symmetrized versions $\kappa_{G\to}$ and $\kappa_{G\to}^c$. Note here that, as we used previously in Proposition 3.24 on the low-degree correlation, the "right" mapping from a general function to an equivariant one is $f(T) \mapsto \mathbb{E}_{Q\sim\mathrm{Haar}}\, Q^\top f(Q\cdot T)$. Indeed, 1-open graph moments, being equivariant already, are easily checked to be unchanged by this operation. For the other two quantities, we define analogs of the closed cumulants, which we call *1-open finite free cumulants*:

$$\kappa_{G\to}(T) := \mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}} Q^\top m_{G\to}^!(Q\cdot T),$$
$$\kappa_{G\to}^c(T) := \mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}} Q^\top m_{G\to}^c(Q\cdot T).$$

We now derive analogs of our results for closed graph moments and the associated finite free cumulants for these quantities. First, we give the analogous additivity property. A general statement like Proposition 4.14 holds, but we will not need it, so for the sake of simplicity we restrict our attention to the following special case.

Proposition 4.27. *For any $G \in \mathcal{G}_{d,p\to}$ and $A, B \in \mathrm{Sym}^p(\mathbb{R}^n)$,*

$$\mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}} \kappa_{G\to}(A + Q\cdot B) = \sum_{\substack{G=G_A\sqcup G_B \\ G_A \text{ open} \\ G_B \text{ closed}}} \frac{n^{\underline{b-1}}}{n^{\underline{b_A-1}} n^{\underline{b_B}}} \kappa_{G_A\to}(A)\kappa_{G_B}(B), \tag{60}$$

$$\mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}} \kappa^c_{G\to}(A + Q\cdot B) = \sum_{\substack{G=G_A\sqcup G_B \\ G_A \text{ open} \\ G_B \text{ closed}}} \frac{n^{\underline{b-1}}}{n^{\underline{b_A-1}} n^{\underline{b_B}}} \kappa^c_{G_A\to}(A)\kappa_{G_B}(B). \tag{61}$$

Again, the proof is a simple variation on that of Proposition 4.12.

Next, we describe the expansion of each of the $\kappa_{G\to}$ and $\kappa^c_{G\to}$ in open graph moments.

Definition 4.28. *For $G \in \mathcal{G}_{d,p\to}$, write* $\mathrm{chop}(G)$ *for the (closed) graph formed by deleting the open edge of G. This is a graph where every vertex has degree p, except for one vertex (the one that used to be the only endpoint of the open edge in G) that has degree $p-1$.*

We will need some generalizations of the ideas discussed previously involving the Weingarten function. First, note that the notion of a matching *realizing* a graph is sensible for a graph with any fixed degree sequence: for a graph G with degree sequence $p_1, \ldots, p_d$, we may similarly say that $\mu \in \mathcal{M}([p_1 + \cdots + p_d])$ realizes G if G is isomorphic to the graph formed by identifying $\{1, \ldots, p_1\}$, $\{p_1 + 1, \ldots, p_1 + p_2\}$, and so forth in μ. Accordingly, the graph Weingarten function of Definition 4.16 may be generalized to graphs with arbitrary specified degree sequence, and in particular $\mathrm{Wg}_{\mathrm{chop}(G),\mathrm{chop}(H)}$ is defined for $G, H \in \mathcal{G}_{d,p\to}$. Lastly, the notion of edge automorphisms from Definition 4.8 is still valid for graphs that are not p-regular, and we may again make sense of $\mathrm{eAut}(\mathrm{chop}(G))$ for $G \in \mathcal{G}_{d,p\to}$. With these tools in hand, we may formulate:

Proposition 4.29. *For $G \in \mathcal{G}_{d,p\to}$, the number of matchings realizing* $\mathrm{chop}(G)$ *is $p!^{d-1}(p-1)!(d-1)!/|\mathrm{eAut}(\mathrm{chop}(G))|$.*

Proof. The proof is the same as for Proposition 4.17, except that permutations of vertices must preserve the unique vertex of degree $p-1$, and permutations of the half-edges around this vertex contribute $(p-1)!$ rather than $p!$. □

Lemma 4.30. *For any $G \in \mathcal{G}_{d,p\to}$ with d vertices and b edges (including the open edge),*

$$\kappa_{G\to}(T) = \frac{n^{\underline{b-1}}}{p!^{d-1}(p-1)!(d-1)!} \cdot |\mathrm{eAut}(\mathrm{chop}(G))| \cdot \sum_H \mathrm{Wg}_{\mathrm{chop}(G),\mathrm{chop}(H)} m_{H\to}(T). \tag{62}$$

Proof. As in Lemma 4.18, we think in terms of matchings. We will use the notations $w(\mu)$ and $w^!(\mu)$ used there for tensors associated to matchings μ.

If μ is a matching of $[pd-1] = [2(b-1)]$ realizing $\mathrm{chop}(G)$, then $m_{G\to}(T)_i = \langle T^{\otimes d}, w(\mu)\otimes e_i\rangle$, and $m^!_{G\to}(T)_i = \langle T^{\otimes d}, w^!(\mu)\otimes e_i\rangle$. Following the argument from Lemma 4.18, we find

$$\begin{aligned}\kappa_{G\to}(T)_i &= \sum_{\substack{j \text{ distinctly} \\ \text{compatible with } \mu}} \mathop{\mathbb{E}}_{Q\sim\mathrm{Haar}} \langle T^{\otimes d}, (Q^{\otimes pd-1} e_j)\otimes e_i\rangle \\ &= n^{\underline{b-1}} \langle T^{\otimes d}, (\Pi_{pd-1} e_\mu)\otimes e_i\rangle\end{aligned}$$

and by the Weingarten formula,

$$= n^{\underline{b-1}} \sum_{\nu} \mathrm{Wg}_{\mu,\nu} \langle T^{\otimes d}, w(\nu) \otimes e_i \rangle$$
$$= n^{\underline{b-1}} \sum_{\nu} \mathrm{Wg}_{\mu,\nu} m_{G(\nu)\to}(T)_i,$$

and the rest of the calculations proceed identically to Lemma 4.18. □

The following is the analog of Proposition 4.19 for 1-open multigraphs, and follows from exactly the same simple expansion.

Proposition 4.31. *For any p-regular open multigraph G on d vertices and b edges,*

$$\kappa^c_{G\to}(T) = \sum_{S \subseteq \mathrm{Frob}(G)} (-1)^{|S|} (n - b + p|S|)^{\underline{p|S|}} \kappa_{G\setminus S\to}(T). \tag{63}$$

Proposition 4.32. *Suppose that G and H are 1-open p-regular multigraphs with $|V(G)| \geq |V(H)|$. Then, $\mathbb{E}_{W\sim\mathrm{Wig}} \langle m^c_{G\to}(W), m_{H\to}(W) \rangle = 0$ unless G and H are isomorphic, and*

$$\mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} \langle m^c_{G\to}(W), m_{G\to}(W) \rangle = \mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} \langle m^c_{G\to}(W), m^c_{G\to}(W) \rangle = n^{\underline{b}} \cdot |\mathrm{eAut}(\mathrm{chop}(G))|. \tag{64}$$

Lemma 4.33. *Suppose that G and H are 1-open p-regular multigraphs. If $|V(G)| \neq |V(H)|$, then $\mathbb{E}_{W\sim\mathrm{Wig}} \langle \kappa^c_{G\to}(W), \kappa^c_{H\to}(W) \rangle = 0$. If $|V(G)| = |V(H)| = d$, then*

$$\mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} \langle \kappa^c_{G\to}(W), \kappa^c_{H\to}(W) \rangle$$
$$= \frac{n^{\underline{b-1}} n^{\underline{b}}}{p!^{d-1}(p-1)!(d-1)!} \cdot |\mathrm{eAut}(\mathrm{chop}(G))| \cdot |\mathrm{eAut}(\mathrm{chop}(H))| \cdot \mathrm{Wg}_{\mathrm{chop}(G),\mathrm{chop}(H)}.$$

If the overall scaling factor of n^{2b-1} appears unusual, the reader may consider, at an intuitive level, that while the inner product of two closed graph moments each on b edges has $2b$ edges, the inner product of two 1-open graph moments each on b edges only has $2b-1$ edges, since the two open edges "fuse" to become one.

Proof. We again proceed similarly to the closed case in Lemma 4.21. Suppose without loss of generality that $|V(G)| \geq |V(H)|$. We have:

$$\mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} \langle \kappa^c_G(W), \kappa^c_H(W) \rangle$$
$$= \mathop{\mathbb{E}}_{\substack{W\sim\mathrm{Wig} \\ Q,R\sim\mathrm{Haar}}} \langle Q^\top m^c_G(Q \cdot W), R^\top m^c_H(R \cdot W) \rangle$$
$$= \mathop{\mathbb{E}}_{\substack{W\sim\mathrm{Wig} \\ Q,R\sim\mathrm{Haar}}} \langle m^c_G(Q \cdot Q^\top \cdot W), QR^\top m^c_H(R \cdot Q^\top \cdot W) \rangle$$
$$= \mathop{\mathbb{E}}_{\substack{W\sim\mathrm{Wig} \\ Q\sim\mathrm{Haar}}} \langle m^c_G(W), Q^\top m^c_H(Q \cdot W) \rangle$$
$$= \mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} \langle m^c_G(W), \kappa^c_H(W) \rangle$$

and, using Lemma 4.30 and Proposition 4.31,

$$= \sum_{S \subseteq \mathrm{Frob}(H)} (-1)^{|S|} (n - b_H + p|S|)^{\underline{p|S|}} \frac{n^{\underline{b_H - 1}}}{p!^{d-1}(p-1)(d-1)!}$$
$$|\mathrm{eAut}(\mathrm{chop}(H))| \cdot \sum_K \mathrm{Wg}_{\mathrm{chop}(H \setminus S), \mathrm{chop}(K)} \mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} \langle m^c_{G\to}(W), m_{K\to}(W) \rangle$$

Here, by Proposition 4.32 all terms are zero unless $|V(G)| = |V(H)|$. In that case, only the term $K = G$ and $S = \varnothing$ contributes, and we are left with

$$= \frac{n^{\underline{b-1}} n^{\underline{b}}}{p!^{d-1}(p-1)!(d-1)!} \cdot |\mathrm{eAut}(\mathrm{chop}(G))| \cdot |\mathrm{eAut}(\mathrm{chop}(H))| \cdot \mathrm{Wg}_{\mathrm{chop}(G), \mathrm{chop}(H)},$$

completing the calculation. □

Lastly, as in the closed case, we construct a nearly orthonormal basis with respect to this inner product.

Definition 4.34. *For each G a p-regular 1-open multigraph having b edges, define*

$$\widehat{\kappa^c_{G\to}}(T) := \frac{n^{b/2}}{n^{\underline{b}} \cdot \sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}} \kappa^c_{G\to}(T). \tag{65}$$

Lemma 4.35. *If $|V(G)| \neq |V(H)|$, then $\mathbb{E}_{W \sim \mathrm{Wig}} \langle \widehat{\kappa^c_{G\to}}(W), \widehat{\kappa^c_{H\to}}(W) \rangle = 0$. Let $M \in \mathbb{R}^{\mathcal{G}_{d,p\to} \times \mathcal{G}_{d,p\to}}$ have entries*

$$M_{G,H} := \mathbb{E}_{W \sim \mathrm{Wig}} \langle \widehat{\kappa^c_{G\to}}(W), \widehat{\kappa^c_{H\to}}(W) \rangle. \tag{66}$$

Suppose that $d \leq \sqrt{n/2p^2}$. Then,

$$\frac{1}{2} \leq \lambda_{\min}(M) \leq \lambda_{\max}(M) \leq 2. \tag{67}$$

In particular, the $\kappa^c_{G\to}(T)$ are linearly independent as vectors of polynomials.

Proof. By Lemma 4.33, we have

$$M_{G,H} = n^{b-1} \cdot \frac{\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))| \cdot |\mathrm{eAut}(\mathrm{chop}(H))|}}{p!^{d-1}(p-1)!(d-1)!} \mathrm{Wg}_{\mathrm{chop}(G), \mathrm{chop}(H)}. \tag{68}$$

The result then follows by Corollary C.5, where we observe that $b = (pd+1)/2$ for $G \in \mathcal{G}_{d,p\to}$. □

5 Application 1: Tensor PCA

5.1 Warmup: Detection with Individual Graph Moments

Before proceeding to the analysis of general low-degree polynomials, let us consider the simpler question of whether individual graph moments $m_G(T)$ can distinguish between the two distributions associated with tensor PCA. In order to do this, we will need to understand how these moments behave under both the null and planted distributions.

The question for the null model is perhaps of independent interest, and we will see that it holds some surprises already. Recall that, for $k = 2$ and $G = C_\ell$ a cycle of length ℓ, classical random matrix theory tells us that

$$\mathop{\mathbb{E}}_{T\sim\mathrm{Wig}(2,n,1)}[m_G(T)] = \mathop{\mathbb{E}}_{T\sim\mathrm{GOE}(n)}[\mathrm{tr}(T^\ell)] = \left(1 + O\left(\frac{1}{n}\right)\right)\mathbf{1}\{\ell \text{ even}\}\,\mathrm{Cat}(\ell/2)\,n^{\ell/2+1}\,, \tag{69}$$

where $\mathrm{Cat}(t) = \frac{1}{t+1}\binom{2t}{t}$ is the t'th Catalan number. What is the tensor analog of this calculation?

Definition 5.1. *Let $G = (V, E)$ be a p-regular graph. For C a finite set of colors, let $\sigma : E \to C$ be an edge coloring. We call σ an* even *coloring if, among the $\sigma(\partial v)$ over $v \in V$, viewed as multisets of colors, each multiset occurs an even number of times. We write $c_{\max}(G)$ for the largest possible number of colors in an even edge coloring of G. For σ an even edge coloring, suppose that the distinct colorings of neighborhoods (i.e., the multisets $\sigma(\partial v)$ over vertices v) occurring in σ are $N_1, \dots, N_m$, where N_i occurs $f_i \in 2\mathbb{N}$ times. Also, write $c_j(N_i)$ for the number of times that $j \in C$ occurs in N_i. We call the* weight *of σ the quantity*

$$w(\sigma) := \prod_{i=1}^{m}(f_i - 1)!! \prod_{j\in C} c_j(N_i)!^{f_i/2}. \tag{70}$$

Proposition 5.2. *Define*

$$w_{\max}(G) := \sum_{\substack{\sigma \text{ even edge coloring of } G \\ |\sigma(E)|=c_{\max}(G)}} w(\sigma), \tag{71}$$

where the sum is over non-isomorphic edge colorings with the given number of colors. Then,

$$\mathop{\mathbb{E}}_{T\sim\mathrm{Wig}(p,n,1)}[m_G(T)] = \left(1 + O_G\left(\frac{1}{n}\right)\right) w_{\max}(G)\,n^{c_{\max}(G)}. \tag{72}$$

Actually, as we show in Propositions B.13 and B.14, any maximum even edge coloring σ has $w(\sigma) = 1$ (that is, no vertex is adjacent to several edges of the same color, and every colored neighborhood occurs exactly twice). Therefore, $w_{\max}(G)$ is purely a counting problem of a particular type of coloring.

Proof. We expand the definition of the tensor moment and use linearity:

$$\mathop{\mathbb{E}}_{T\sim\mathrm{Wig}(p,n,1)} m_G(T) = \sum_{i\in[n]^E} \mathbb{E}\prod_{v\in V} T_{i(\partial v)} \tag{73}$$

$$= \sum_{\sigma \text{ even edge coloring of } G} n^{\underline{|\sigma(E)|}}\, w(\sigma) \tag{74}$$

$$= \left(1 + O_G\left(\frac{1}{n}\right)\right) \sum_{\sigma \text{ even edge coloring of } G} n^{|\sigma(E)|} w(\sigma), \tag{75}$$

and extracting the leading order contribution completes the proof. Here we sum over all non-isomorphic even edge colorings σ; labeling the colors of such colorings by distinct values in $[n]$ enumerates all $i \in [n]^E$ that contribute to the sum. □

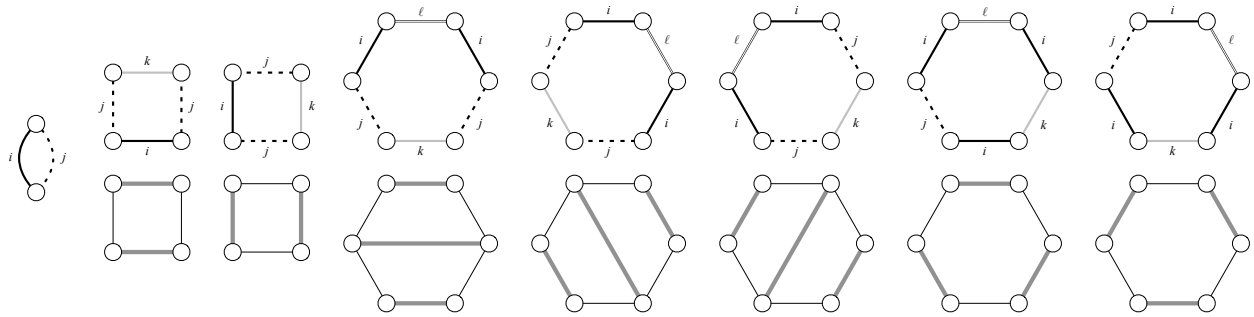


Figure 8: Examples of legal edge colorings for the matrix case $p = 2$. In the cycle of length 2 there is one legal coloring with two distinct colors (up to permutations of the colors) recovering the fact that $\operatorname{tr} A^2 = \sum_{ij} A_{ij}A_{ji} = n^2$. There are two legal colorings of the cycle of length 4: the one on the left, for instance, corresponds to the term $\sum_{ijk} A_{ij}A_{jk}A_{kj}A_{ji}$. There are five legal colorings of the cycle of length 6. Each coloring corresponds to the matching of vertices shown below it, where we show the pairs of vertices with the same neighborhoods: these matchings are the non-crossing partitions familiar in combinatorial free probability [NS06]. This recovers the leading behavior $\operatorname{tr} A^\ell = \mathrm{Cat}(\ell/2)n^{\ell/2+1}$ where $\mathrm{Cat}(t) = 1, 2, 5, \ldots$ is the tth Catalan number.

As an aside, we mention the following ancillary result. Despite this characterization, unlike in the matrix case, and perhaps surprisingly, it is difficult in general to compute even the scaling of a given moment of a Wigner tensor for large n.

Theorem 5.3. *It is NP-hard to decide whether $c_{\max}(G) \geq c$ given $c \geq 0$ and a regular graph G.*

We give a proof in Appendix B.

In particular, we may understand the scaling of *variances* of $m_G(T)$ as follows.

Corollary 5.4. *For any G, for $W \sim \mathrm{Wig}$, $\operatorname{Var} m_G(W) \geq n^{\underline{|E(G)|}}$.*

Proof. First, note that, in general, $m_G(T)m_H(T) = m_{G \sqcup H}(T)$, where $\sqcup$ denotes disjoint union of graphs. Thus

$$\mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} m_G(W)^2 = \mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} m_{G \sqcup G}(W) = \sum_{\sigma \text{ even coloring of } G \sqcup G} n^{\underline{|\sigma(E)|}}\, w(\sigma). \tag{76}$$

On the other hand, we have

$$\left(\mathop{\mathbb{E}}_{W \sim \mathrm{Wig}} m_G(W)\right)^2 = \left(\sum_{\sigma \text{ even coloring of } G} n^{\underline{|\sigma(E)|}}\, w(\sigma)\right)^2 \tag{77}$$

which is at most the sum of the terms in (76) where each colored vertex neighborhood occurs at least twice within one of the two copies of G. There is another coloring in the summation in (76), which is not of this kind: it is the coloring that colors every edge in one copy of G with a different color, and then repeats the same coloring on the other copy of G. This is an even coloring which uses $|E(G)|$ colors, and which is not cancelled in computing $\operatorname{Var} m_G(W) = \mathbb{E}_{W \sim \mathrm{Wig}} m_G(W)^2 - (\mathbb{E}_{W \sim \mathrm{Wig}} m_G(W))^2$. See Figure 9 for an illustration. □

For the planted model, we are interested in the special case of *rank one* tensors. On these, evaluating the graph moments is easy:

Proposition 5.5. *Let $v^{\otimes p}$ be the pth tensor power of v: that is, $v^{\otimes p}_{i_1,\dots,i_p} = v_{i_1} v_{i_2} \cdots v_{i_p}$. Then if $G = (V, E)$, we have $m_G(v^{\otimes p}) = \|v\|^{2|E|} = \|v\|^{p|V|}$.*

Proof. The summation over the index associated to each edge gives a factor of $\langle v, v\rangle = \|v\|^2$ on each edge, for a total of $\|v\|^{2|E|}$. Since G is p-regular, $|E| = p|V|/2$. □

We now sketch an argument of [OTR22] that no single graph moment can solve the detection problem for tensor PCA below the conjectured threshold.

Proposition 5.6. *Let G be a p-regular graph. If $\lambda \ll n^{-p/4}$, then*

$$\left| \mathop{\mathbb{E}}_{Y \sim \mathbb{P}} m_G(Y) - \mathop{\mathbb{E}}_{Y \sim \mathbb{Q}} m_G(Y) \right| \ll \sqrt{\mathop{\mathrm{Var}}_{Y \sim \mathbb{Q}} m_G(Y)}. \tag{78}$$

Proof Sketch. Suppose G has d vertices and b edges. First we consider the expected difference in m_G between the two models. If $Y = \lambda v^{\otimes p} + W$, then expanding $m_G(Y)$ creates 2^d terms, including cross-terms where the spike $\lambda v^{\otimes p}$ appears at some vertices and the noise W appears at others. For some graphs these cross-terms can be neglected [OTR22]. In this proof sketch we look only at the contribution of the all-spike term. By Proposition 5.5 this is $\lambda^d n^b$, since we get a factor of λ at each vertex and a factor of $|v|^2 = n$ on each edge.

On the other hand, by Corollary 5.4, we have $\mathrm{Var}\, m_G(Y) \gtrsim n^b$. Comparing the spike's contribution to the expectation with the standard deviation in the null model, we see that m_G fails as a test statistic whenever

$$\lambda^a n^b \ll n^{b/2}. \tag{79}$$

But since G is p-regular, we have $b = pa/2$. Then (79) becomes

$$\lambda^a n^{pa/2} \ll n^{pa/4},$$

and solving for λ gives $\lambda \ll n^{-p/4}$ as stated. □

This argument is tight for some graphs [OTR22]. However, while graph moments are fascinating, for general G they are not ideal quantities to prove upper and lower bounds on problems such as tensor PCA, for three reasons:

1. By Theorem 5.3, even computing how the expectations of $m_G(W)$ scales with n in the null model $W \sim \mathrm{Wig}$ is computationally hard.

2. As discussed above, the expectation of $m_G(Y)$ in spiked models involves cross-terms where the spike appears at some vertices and the noise appears at others. These mixed moments are often nonzero, making precise calculations combinatorially challenging.

3. Different graph moments m_G and m_H can be correlated in the Wigner model, even when G and H are non-isomorphic. This raises the possibility that, while the variance of any one graph moment is too large to beat the threshold, it might be possible to construct linear combinations of small graphs that cancel out much of each others' variance. Indeed, Proposition 5.6 applies to graphs of any size, even where a grows with n, but there are subexponential-time

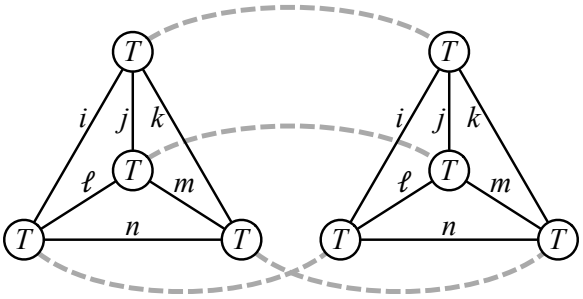


Figure 9: Even colorings of the disjoint union of two copies of K_4 where each vertex is matched to its counterpart. The colors or indices i, j, k, ℓ, m, n in one copy must match their counterparts, but within each copy these indices can range freely over $[n]$. Thus they contribute n^6 to the variance of T^{K_4} when T is a Wigner random tensor.

> algorithms of growing degree that succeed even when $\lambda \ll n^{-p/4}$ [BGG+16, BGL17, WAM19] (see also [RRS17b] for analogous algorithms for random XOR-SAT). Low-degree algorithms matching this performance must somehow involve linear combinations of many graph moments.

We therefore proceed to full lower bounds against low-degree polynomials, now leaning on the machinery we have developed based on finite free cumulants, which will allow us to circumvent all of these difficulties.

5.2 Detection with General Low-Degree Polynomials: Proof of Theorem 1.6

Before proceeding, let us state the formal version of Theorem 1.6 that we will prove here. Recall we are in the setting where $\mathbb{Q} = \mathrm{Wig}$ and $\mathbb{P}$ the spiked tensor model $Y = \lambda v^{\otimes p} + W$.

Theorem 5.7 (Low-degree analysis for tensor PCA detection)**.** *Let $D = D(n) \in \mathbb{N}$ have $D \le \sqrt{n/2p^2}$. There are constants $a_p, b_p > 0$ such that:*

1. *If $\lambda \le a_p n^{-p/4} D^{-(p-2)/4}$, then $\mathrm{Adv}_{\le D}(\mathbb{Q} = \mathrm{Wig}, \mathbb{P}) = O(1)$.*
2. *If $\lambda \ge b_p n^{-p/4} D^{-(p-2)/4}$ and $D = \omega(1)$, then $\mathrm{Adv}_{\le D}(\mathbb{Q} = \mathrm{Wig}, \mathbb{P}) = \omega(1)$.*

First we compute the expectation of a graph cumulant in the spiked model. In fact, the cumulant of $v^{\otimes p}$ is the same as its ordinary graph moment $m_G(v^{\otimes p})$ given by Proposition 5.5 with a correction term.

Proposition 5.8. *Let $v \in \mathbb{R}^n$. If $G = (V, E)$ with $|V| = a$ and $|E| = b$, then*

$$\kappa_G(v^{\otimes p}) = \frac{n^{\underline{b}}}{(n + 2b - 2)^{\underline{b}}} \|v\|^{2b}.$$

Proof. Recall that $\kappa_G(T) = \mathbb{E}_Q \, m_G^!(Q \cdot T)$, and that $Q \cdot v^{\otimes p} = (Q \cdot v)^{\otimes p}$. Since Q is Haar-random, $u = Q \cdot v$ is uniformly chosen from the sphere of radius $\|v\|$. As in Proposition 5.5, we have a copy of u on the outgoing half-edges of each vertex. But rather than the inner product $\langle u, u \rangle = \|v\|^2$ on

each edge, we have $\langle u, e\rangle\langle e, u\rangle = \langle e, u\rangle^2$ where e is the basis vector associated with that edge. Since all $n^{\underline{b}}$ tuples of distinct indices will contribute the same expectation, we can fix an orthogonal basis $e_1, \dots, e_b$ and write

$$\kappa_G(v^{\otimes p}) = n^{\underline{b}} \mathop{\mathbb{E}}_u \prod_{i=1}^{b} \langle e_i, u\rangle^2 . \tag{80}$$

We can write this product as an inner product of tensors,

$$\mathop{\mathbb{E}}_u \prod_{i=1}^{b} \langle e_i, u\rangle^2 = \left\langle \left(\bigotimes_{i=1}^{b} e_i \right)^{\otimes 2} , \mathop{\mathbb{E}}_u u^{\otimes 2b} \right\rangle . \tag{81}$$

Since $\mathbb{E}_u\, u^{\otimes 2b}$ is fixed under conjugation by any $Q \in \mathcal{O}(n)$, it lies in the trivial subspace of $\mathbb{R}^{\otimes 2b}$. As in the proof of Theorem 3.2, this implies that it is a linear combination of matchings of the $2b$ half-edges, each of which corresponds to a "rewiring" of G à la the configuration model. Moreover, by symmetry, all of these matchings have equal weight. Using the representation theory of the orthogonal group (e.g. [MR11]) we can obtain

$$\mathop{\mathbb{E}}_u u^{\otimes 2b} = \frac{\|u\|^{2b}}{n(n+2)(n+4)\cdots(n+2b-2)} \sum_{\mu:\ \text{matching of } [2b]} w_\mu . \tag{82}$$

This is analogous to Isserlis's or Wick's theorem for the Gaussian measure [Iss18, Wic50] where the denominator would simply be n^b.

However, since the e_i are orthogonal, any matching μ that partners one with another yields a zero inner product. Thus the only nonzero contribution to (81) comes from the matching given by the original wiring of G, where for each $i \in [b]$ the two copies of e_i in $(\bigotimes_i e_i)^{\otimes 2}$ are matched with each other. That matching contributes $\prod_i \langle e_i, e_i\rangle = 1$ to (81), and combining this with (80) and (82) gives

$$\kappa_G(v^{\otimes p}) = \|v\|^{2b} \frac{n(n-1)(n-2)\cdots(n-b+1)}{n(n+2)(n+4)\cdots(n+2b-2)} = \|v\|^{2b} \frac{n^{\underline{b}}}{(n+2b-2)^{\underline{b}}} .$$

This correction factor is roughly $\mathrm{e}^{-(3/2)b^2/n}$, and in any case is $1 - O(b^2/n)$. □

Proposition 5.9. *The number of (non-isomorphic, unlabelled) p-regular multigraphs on d vertices is asymptotically as $d \to \infty$*

$$|\mathcal{G}_{d,p}| \sim \frac{(pd-1)!!}{d!p!^d} \sim \sqrt{\frac{p}{\pi}} \left(\frac{p^{p/2}}{p!} e^{-\frac{p-2}{2}} d^{\frac{p-2}{2}} \right)^d . \tag{83}$$

We also have the concrete upper bound:

$$|\mathcal{G}_{d,p}| \le \frac{(pd-1)!!}{d!p!^d} \le \left(e^{p+1} p^{-p/2} d^{\frac{p-2}{2}} \right)^d . \tag{84}$$

Proof. The asymptotics follow from the proof of [Bol82], repeated without the restriction to simple graphs, and Stirling's approximation. The initial upper bound follows since every p-regular multigraph on d vertices corresponds to at least $d!p!^d$ perfect matchings of $[pd]$, as in the configuration model. The second bound follows by a non-asymptotic version of Stirling's approximation. □

Proof of Theorem 5.7. We begin by computing the centered cumulants under $\mathbb{P}$:

$$\begin{aligned}
\mathop{\mathbb{E}}_{T\sim\mathbb{P}} \kappa_G^c(T) &= \mathop{\mathbb{E}}_{\substack{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))\\ W\sim\mathrm{Wig}}} \kappa_G^c(\lambda v^{\otimes p} + W) \\
&= \mathop{\mathbb{E}}_{W\sim\mathrm{Wig}} \kappa_G^c(W) + \mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))} \kappa_G(\lambda v^{\otimes p}) \\
&= \mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))} \kappa_G(\lambda v^{\otimes p}) && \text{(Proposition 4.15)} \\
&\asymp \lambda^d n^{pd/2}\,. && \text{(Proposition 5.8)}
\end{aligned}$$

Now, substituting this into Corollary 4.24,

$$\begin{aligned}
\mathrm{Adv}_{\le D}(\mathrm{Wig},\mathbb{P}) &\asymp \sum_{d=0}^{D} \frac{1}{n^{pd/2}} \cdot \lambda^{2d} n^{pd} \sum_{|V(G)|=d} \frac{1}{|\mathrm{eAut}(G)|} \\
&\le \sum_{d=0}^{D} d! \left(\lambda^2 n^{p/2}\right)^d \#\{p\text{-regular multigraphs on } d \text{ vertices}\}
\end{aligned}$$

and bounding the number of p-regular multigraphs by Proposition 5.9, we have

$$\begin{aligned}
&\le \sum_{d=0}^{D} d! \left(\lambda^2 n^{p/2}\right)^d \frac{(pd-1)!!}{d!} \\
&\le \sum_{d=0}^{D} \left(\lambda^2 n^{p/2}\right)^d \frac{p d^{pd/2}}{(d/e)^d} \\
&\le \sum_{d=0}^{D} \left(\lambda^2 \cdot e^{p+1} p^{-p/2} d^{(p-2)/2} n^{p/2}\right)^d,
\end{aligned}$$

and the result follows since, under the stated assumptions and by a suitable choice of the constant a_p, the quantity being raised to a power will be, say, smaller than $1/2$. For the lower bound, we may use our asymptotic formula for the advantage and use that a positive proportion of d-regular multigraphs are simple and almost all have trivial automorphism group (see, e.g., [Bol82]). □

We elaborate on a remark from the Introduction: the factor of $d^{\frac{p-2}{2}d}$ in the final summation governs the tradeoff between the power of detection algorithms and their subexponential runtime budget. In our calculation, this quantity has a clear and direct source: it is just the asymptotic value of the number of non-isomorphic p-regular multigraphs (Proposition 5.9), and therefore also the dimension of the space of degree d invariant polynomials.

5.3 Reconstruction with Low-Degree Polynomials

We first state the formal result we will show.

Theorem 5.10 (Low-degree lower bound for tensor PCA reconstruction)**.** *Let $D = D(n) \in \mathbb{N}$ have $D \le \sqrt{n/2p^2}$. For all odd $p \ge 3$, there is a constant $c_p > 0$ such that if $\lambda \le c_p n^{-p/4} D^{-(p-2)/4}$, then $\mathrm{Corr}_{\le D}(\mathbb{P})^2 = O(\sqrt{n})$, and therefore $\mathrm{MMSE}_{\le D}(\mathbb{P}) = n - O(\sqrt{n})$.*

Remark 5.11 (Better-than-random reconstruction)**.** *The $O(\sqrt{n})$ scaling of the correlation even in the computationally hard regime is perhaps surprising, since a uniformly random estimator $\widehat{v} \in \mathbb{S}^{n-1}(\sqrt{n})$ achieves squared correlation with the signal $v \in \mathbb{S}^{n-1}(\sqrt{n})$ of $\langle v, \widehat{v}\rangle^2/\|\widehat{v}\|^2 = O(1)$ with high probability. Indeed, estimation slightly better than this is always possible: consider the estimator $\widehat{v} = \widehat{v}(Y)$ formed by the open graph moment $m_{G\to}(Y)$ where G is the graph on one vertex with $(p-1)/2$ self-loops and one open edge (for p odd). This is a linear function, so $m_{G\to}(\lambda v^{\otimes p} + W) = \lambda\|v\|^{p-1}v + m_{G\to}(W) = \lambda n^{(p-1)/2}v + m_{G\to}(W)$. Let us write $g := m_{G\to}(W)$. This is independent of v and distributed roughly as $\mathcal{N}(0, n^{(p-1)/2}I)$. Thus, proceeding heuristically, the squared correlation is typically*

$$\frac{\langle v, \widehat{v}\rangle^2}{\|\widehat{v}\|^2} \approx \frac{\lambda^2 n^{p+1}}{\lambda^2 n^p + n^{(p+1)/2}}\,. \tag{85}$$

When $\lambda \ll n^{-p/4}$, the first term in the denominator is smaller than the second, so this is roughly $\lambda^2 n^{(p+1)/2}$, which when $\lambda \sim n^{-p/4}$ becomes as large as $n^{1/2}$, as claimed. Indeed, once $\lambda \gg n^{-(p-1)/4}$, this expression is $\Omega(n)$, so this simple estimator achieves linear correlation.

We begin with a number of preliminaries. We will use the following analog of Proposition 5.9 for counting open multigraphs.

Proposition 5.12. *The number of (non-isomorphic, unlabelled) 1-open p-regular multigraphs on d vertices is bounded by*

$$|\mathcal{G}_{d,p\to}| \le \frac{(pd-2)!!}{(d-1)!p!^{d-1}(p-1)!} \le \left((2e)^{p+1}(p-1)^{-(p-1)/2}d^{\frac{p-2}{2}}\right)^d. \tag{86}$$

Proof. The initial upper bound follows first by associating a 1-open multigraph to a closed multigraph with all vertices having degree p except for one having degree $p-1$. The total number of edges in this graph is $\frac{pd-1}{2}$, and the remaining bounds follow as in Proposition 5.9. □

Proposition 5.13. *For any 1-open p-regular multigraph G and any $T \in \mathrm{Sym}^p(\mathbb{R}^n)$,*

$$\mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))} \kappa^c_{G\to}(\lambda v^{\otimes p} + T) = \sum_{\substack{G=G_A\sqcup G_B\\ G_A \text{ open}\\ G_B \text{ closed}}} \frac{n^{\underline{b-1}}}{n^{\underline{b_A-1}}\,(n+2b_B-2)^{\underline{\underline{b_B}}}}\,\lambda^{d_B} n^{b_B}\kappa^c_{G_A\to}(T). \tag{87}$$

Proof. We combine Proposition 4.27 with Proposition 5.8:

$$\begin{aligned}\mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))} \kappa^c_{G\to}(\lambda v^{\otimes p} + T) &= \sum_{\substack{G=G_A\sqcup G_B\\ G_A \text{ open}\\ G_B \text{ closed}}} \frac{n^{\underline{b-1}}}{n^{\underline{b_A-1}}\,n^{\underline{b_B}}}\left(\mathop{\mathbb{E}}_{v}\kappa_{G_B}(\lambda v^{\otimes p})\right)\kappa^c_{G_A\to}(T)\\ &= \sum_{\substack{G=G_A\sqcup G_B\\ G_A \text{ open}\\ G_B \text{ closed}}} \frac{n^{\underline{b-1}}}{n^{\underline{b_A-1}}\,n^{\underline{b_B}}}\cdot\lambda^{d_B}\cdot\frac{n^{\underline{b_B}}}{(n+2b_B-2)^{\underline{\underline{b_B}}}} n^{b_B}\kappa^c_{G_A\to}(T),\end{aligned}$$

and simplifying gives the stated result. □

Proposition 5.14. *For any 1-open p-regular multigraph G,*

$$\mathop{\mathbb{E}}_{\substack{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))\\ W\sim\mathrm{Wig}}}\langle v,\kappa^c_{G\to}(\lambda v^{\otimes p}+W)\rangle=\frac{n^{\underline{b}}}{(n+2b-2)^{\underline{\underline{b}}}}\lambda^d n^b. \tag{88}$$

Proof. Proceeding similarly to the above using Proposition 4.27 with respect to the randomness in W,

$$\begin{aligned}&\mathop{\mathbb{E}}_{\substack{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))\\ W\sim\mathrm{Wig}}}\langle v,\kappa^c_{G\to}(\lambda v^{\otimes p}+W)\rangle\\ &=\mathop{\mathbb{E}}_{\substack{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))\\ W\sim\mathrm{Wig}}}\sum_{\substack{G=G_A\sqcup G_B\\ A\text{ open}\\ B\text{ closed}}}\frac{n^{\underline{b-1}}}{n^{\underline{b_A-1}}n^{\underline{\underline{b_B}}}}\kappa^c_{G_B}(W)\langle v,\kappa_{G_A\to}(\lambda v^{\otimes p})\rangle\end{aligned}$$

where the only non-zero term is when $B=\varnothing$, so that

$$\begin{aligned}&=\mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))}\langle v,\kappa_{G\to}(\lambda v^{\otimes p})\rangle\\ &=\lambda^d\mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))}\langle v,\kappa_{G\to}(v^{\otimes p})\rangle\\ &=\lambda^d\mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))}\langle v,m^!_{G\to}(v^{\otimes p})\rangle\end{aligned}$$

and now the same calculations as in Proposition 5.8 apply, giving

$$=\frac{n^{\underline{b}}}{(n+2b-2)^{\underline{\underline{b}}}}\lambda^d n^b,$$

which is the final result. □

Proof of Theorem 5.10. Recall that we are interested in producing an upper bound on the correlation, which we may restrict to equivariant polynomials without loss of generality by Proposition 3.24.

$$\mathrm{Corr}_{\le D}(\mathbb{P})=\sup_{\substack{f\in\mathbb{R}[Y]^n_{\le D}\\ f\text{ equivariant}\\ \mathbb{E}_{Y\sim\mathbb{P}}\|f(Y)\|^2\neq 0}}\frac{\mathbb{E}_{(v,Y)\sim\mathbb{P}}\langle v,f(Y)\rangle}{\sqrt{\mathbb{E}_{Y\sim\mathbb{P}}\|f(Y)\|^2}}. \tag{89}$$

Consider a particular equivariant f. By Theorem 3.15 and Lemma 4.35, it admits an expansion in the $\widehat{\kappa^c_{G\to}}$,

$$f(T)=\sum_G\alpha_G\widehat{\kappa^c_{G\to}}(T)=\sum_G\frac{n^{b_G/2}}{n^{\underline{b_G}}\cdot\sqrt{|\mathrm{eAut}(\mathrm{chop}(G)|}}\alpha_G\kappa^c_{G\to}(T). \tag{90}$$

where the sum is over p-regular 1-open multigraphs on at most D vertices. We now apply the technique of [SW22], using Jensen's inequality to give a lower bound on the norm of f:

$$\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}\|f(Y)\|^2=\mathop{\mathbb{E}}_{\substack{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))\\ W\sim\mathrm{Wig}}}\|f(\lambda v^{\otimes p}+W)\|^2 \tag{91}$$

$$\ge\mathop{\mathbb{E}}_{W\sim\mathrm{Wig}}\left\|\mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))}f(\lambda v^{\otimes p}+W)\right\|^2. \tag{92}$$

We can write the inner quantity as:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))} f(\lambda v^{\otimes p}+W)\\
&=\sum_G \frac{n^{b_G/2}}{n^{\underline{b_G}}\cdot\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}}\alpha_G \mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))} \kappa^c_{G\to}(\lambda v^{\otimes p}+T)\\
&=\sum_G \frac{n^{b_G/2}}{n^{\underline{b_G}}\cdot\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}}\alpha_G \sum_{\substack{G=G_A\sqcup G_B\\ G_A \text{ open}\\ G_B \text{ closed}}} \frac{n^{\underline{b_G-1}}}{n^{\underline{b_A-1}}(n+2b_B-2)^{\underline{\underline{b_B}}}}\ \lambda^{d_B}n^{b_B}\kappa^c_{G_A\to}(T)\\
&=\sum_G \frac{n^{b_G/2}}{(n-b_G+1)\cdot\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}}\alpha_G \sum_{\substack{G=G_A\sqcup G_B\\ G_A \text{ open}\\ G_B \text{ closed}}} \frac{1}{n^{\underline{b_A-1}}(n+2b_B-2)^{\underline{\underline{b_B}}}}\ \lambda^{d_B}n^{b_B}\kappa^c_{G_A\to}(T)
\end{aligned}
$$

$$
\begin{aligned}
&=\sum_G \frac{n^{b_G/2}}{(n-b_G+1)\cdot\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}}\alpha_G \sum_{\substack{G=G_A\sqcup G_B\\ G_A \text{ open}\\ G_B \text{ closed}}} \frac{n^{\underline{b_A}}\sqrt{|\mathrm{eAut}(\mathrm{chop}(G_A))|}}{n^{b_A/2}n^{\underline{b_A-1}}(n+2b_B-2)^{\underline{\underline{b_B}}}}\ \lambda^{d_B}n^{b_B}\widehat{\kappa^c_{G_A\to}}(T)\\
&=\sum_G \frac{1}{(n-b_G+1)\cdot\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}}\alpha_G\\
&\qquad\sum_{\substack{G=G_A\sqcup G_B\\ G_A \text{ open}\\ G_B \text{ closed}}} \frac{(n-b_A+1)\sqrt{|\mathrm{eAut}(\mathrm{chop}(G_A))|}}{(n+2b_B-2)^{\underline{\underline{b_B}}}}\ \lambda^{d_B}n^{3b_B/2}\widehat{\kappa^c_{G_A\to}}(T).
\end{aligned}
$$

Switching the order of summations, we find that we can write

$$
\mathop{\mathbb{E}}_{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))} f(\lambda v^{\otimes p}+W)=\sum_G (L\alpha)_G\cdot\widehat{\kappa^c_{G\to}}(T)=\sum_G\left(\sum_H R_{G,H}\alpha_H\right)\widehat{\kappa^c_{G\to}}(T). \tag{93}
$$

Here, R is an upper-triangular matrix when graphs are ordered by increasing number of vertices, with rows and columns indexed by $\mathcal{G}_{1,p\to}\sqcup\cdots\sqcup\mathcal{G}_{D,p\to}$. More specifically, we have $R_{G,H}\neq 0$ only when a subset of the connected components of H yield G. When this is so, then the entries of L are given by

$$
R_{G,H}=c_{G,H}\cdot\frac{n-b_G+1}{n-b_H+1}\cdot\sqrt{\frac{|\mathrm{eAut}(\mathrm{chop}(G))|}{|\mathrm{eAut}(\mathrm{chop}(H))|}}\cdot\frac{n^{3b_{H\setminus G}/2}}{(n+2b_{H\setminus G}-2)^{\underline{\underline{b_{H\setminus G}}}}}\cdot\lambda^{d_{H\setminus G}}, \tag{94}
$$

where $c_{G,H}$ is the number of ways that it is possible to choose the connected components of G out of those of H (i.e., the product of $\binom{m_H(C)}{m_G(C)}$ over all connected components C of G, where $m_G(C)$ is the number of times C occurs in G and likewise for $m_H(C)$).

Now, using Lemma 4.33 on the near-orthonormality of the $\widehat{\kappa^c_{G\to}}$, we find

$$
\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}\|f(Y)\|^2\geq\frac{1}{2}\|R\alpha\|^2=\frac{1}{2}\alpha^\top(R^\top R)\alpha. \tag{95}
$$

We define and rewrite using Proposition 5.14

$$\begin{aligned}\beta_G &:= \mathop{\mathbb{E}}_{\substack{v\sim\mathrm{Unif}(\mathbb{S}^{n-1}(\sqrt{n}))\\ W\sim\mathrm{Wig}}} \langle v, \widehat{\kappa^c_{G\to}}(\lambda v^{\otimes p}+W)\rangle \\ &= \frac{n^{b_G/2}}{n^{\underline{b_G}}\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}}\frac{n^{\underline{b_G}}}{(n+2b_G-2)^{\underline{\underline{b_G}}}}\lambda^{d_G}n^{b_G} \\ &= \frac{n^{3b_G/2}}{(n+2b_G-2)^{\underline{\underline{b_G}}}\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}}\lambda^{d_G}.\end{aligned}$$

We have $R_{G,G}=1$ for all G, so R is invertible. For our bound on the correlation, we then may compute

$$\mathrm{Corr}_{\le D}(\mathbb{P})^2 \le 2\sup_{\alpha\neq 0}\frac{\langle\alpha,\beta\rangle^2}{\|R\alpha\|^2} = 2\|R^{\top^{-1}}\beta\|^2. \tag{96}$$

Let us compute the inverse $\gamma := R^{\top^{-1}}\beta$. $R^\top$ is lower triangular and its diagonal is identically equal to 1, so we may compute the inverse recursively by back substitution:

$$\gamma_H = \beta_H - \sum_{G<H} R_{G,H}\gamma_G. \tag{97}$$

We may bound both the entries of β and of R as:

$$\begin{aligned}|\beta_G| &\lesssim \frac{n^{b_G/2}\lambda^{d_G}}{\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}} \\ &= n^{1/4}\frac{\left(n^{p/4}\lambda\right)^{d_G}}{\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}} \\ &\le n^{1/4}\frac{\left(c_pD^{-\frac{p-2}{4}}\right)^{d_G}}{\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))|}},\end{aligned}$$

$$\begin{aligned}|R_{G,H}| &\le \sqrt{\frac{|\mathrm{eAut}(\mathrm{chop}(G))|}{|\mathrm{eAut}(\mathrm{chop}(H))|}}c_{G,H}n^{b_{H\setminus G}/2}\lambda^{d_{H\setminus G}} \\ &= \sqrt{\frac{|\mathrm{eAut}(\mathrm{chop}(G))|}{|\mathrm{eAut}(\mathrm{chop}(H))|}}c_{G,H}\left(n^{p/4}\lambda\right)^{d_{H\setminus G}} \\ &\le \sqrt{\frac{|\mathrm{eAut}(\mathrm{chop}(G))|}{|\mathrm{eAut}(\mathrm{chop}(H))|}}c_{G,H}\left(c_pD^{-\frac{p-2}{4}}\right)^{d_{H\setminus G}}.\end{aligned}$$

Thus inductively we have, with $C(|\mathrm{conn}(H)|)$ denoting the number of distinct chains of subsets of the connected components of H,

$$\begin{aligned}|\gamma_H| &\le |\beta_H| + \sum_{G<H}|R_{G,H}|\cdot|\gamma_G| \\ &\lesssim n^{1/4}\frac{C(|\mathrm{conn}(H)|)}{\sqrt{|\mathrm{eAut}(\mathrm{chop}(H))|}}\left(c_pD^{-\frac{p-2}{4}}\right)^{d_H}\end{aligned}$$

and bounding the number of chains using Proposition D.2,

$$\leq n^{1/4} \frac{3^{|\mathrm{conn}(H)|}(|\mathrm{conn}(H)|)!}{\sqrt{|\mathrm{eAut}(\mathrm{chop}(H))|}} \left(c_p D^{-\frac{p-2}{4}}\right)^{d_H}.$$

Note that every closed connected component of any $H \in \mathcal{G}_{d,p\to}$ must have size at least 2 since p is odd, and only the open connected component can have size 1. Therefore, the number of connected components, if there are d vertices, is at most $1 + \frac{d-1}{2} = \frac{d+1}{2}$. Also, if $c_1, \ldots, c_m$ are the frequencies with which various closed connected components happen in G, then, since permutations of connected components are automorphisms and the "chop" operation does not affect the closed components of a graph,

$$|\mathrm{eAut}(\mathrm{chop}(H))| \geq \prod_{i=1}^{m} c_i!. \tag{98}$$

More specifically, suppose the total number of connected components of size 2 in H is ℓ. There are $\frac{p+1}{2}$ non-isomorphic connected p-regular graphs on two vertices (two vertices with an odd number of edges between them and a suitable number of loops on the ends). Suppose the numbers of these components are $c_1, \ldots, c_{(p+1)/2}$. We then have by a convexity argument

$$|\mathrm{eAut}(\mathrm{chop}(H))| \geq \prod_{i=1}^{(p+1)/2} c_i! \geq \left(\frac{2\ell}{p+1}\right)!^{\frac{p+1}{2}} \geq \left(\frac{2}{e(p+1)}\ell\right)^{\ell}. \tag{99}$$

We may organize the calculation of the norm of γ according to the number of connected components k, and the sizes a_1 of the open component and $a_2, \ldots, a_k$ of the closed components:

$$\|\gamma\|^2 \lesssim n^{1/2} \sum_{d=1}^{D} \left(c_p^2 D^{-\frac{p-2}{2}}\right)^d \sum_{k=1}^{\frac{d+1}{2}} 9^k k!^2 \sum_{\substack{a_1 \geq 1 \\ a_2,\ldots,a_k \geq 2 \\ a_1+a_2+\cdots+a_k = d}} \left(\frac{2}{e(p+1)} |\{i : a_i = 2\}|\right)^{-|\{i:a_i=2\}|} |\mathcal{G}_{a_1,p\to}| \prod_{i=2}^{k} |\mathcal{G}_{a_i,p}|$$

and in bounding the sizes of the sets of graphs by Propositions 5.9 and 5.12, let us replace c_p^2 with a larger c_p' that absorbs constants depending only on p, so that we obtain

$$\leq n^{1/2} \sum_{d=1}^{D} \left(c_p' D^{-\frac{p-2}{2}}\right)^d \left(9|\mathcal{G}_{d,p\to}| + \sum_{k=2}^{\frac{d+1}{2}} 9^k k!^2 \sum_{\substack{a_1 \geq 1 \\ a_2,\ldots,a_k \geq 2 \\ a_1+a_2+\cdots+a_k = d}} |\{i : a_i = 2\}|^{-|\{i:a_i=2\}|} \prod_{i=1}^{k} a_i^{\frac{p-2}{2} a_i} \right)$$

Here, again by a convexity argument, the choice of a_i that maximizes the expression in the inner sum will have the form $a_1 = 1$, $a_2 = \cdots = a_{s+1} = 2$, $a_{s+2} = \cdots = a_{k-1} = 3$, $a_k = d - 3k + s + 5$. The value of a term with these choices is at most $3^{\frac{p-2}{2}} d s^{-s} d^{\frac{p-2}{2}(d-3k+s+5)}$. Since $d^{\frac{p-2}{2}} \geq s$, to maximize this bound we take s as large as possible, i.e., $s = k - 2$. Thus we may bound, again replacing c_p' with c_p'' and absorbing a constant depending only on p,

$$\leq n^{1/2} \sum_{d=1}^{D} \left(c_p'' D^{-\frac{p-2}{2}} \right)^d \left(d^{\frac{p-2}{2}d} + \sum_{k=2}^{\frac{d+1}{2}} k!^2 (k-2)^{-(k-2)} d^{\frac{p-2}{2}(d-2k+3)} \sum_{\substack{a_1 \geq 1 \\ a_2, \ldots, a_k \geq 2 \\ a_1 + a_2 + \cdots + a_k = d}} 1 \right)$$

Here, the remaining sum is at most the number of integer partitions of d, which per Proposition D.3 is of order $\exp(O(\sqrt{d}))$. We have $k!/(k-2)^{k-2} \lesssim k^2 \leq d^2$, and $k! d^{-(p-2)k} \leq k^k d^{-k} \leq 1$. Putting everything together, we find

$$\leq n^{1/2} \sum_{d=1}^{D} \left(c_p'' e^{O(1/\sqrt{d})} \right)^d \left(1 + d^{\frac{3}{2}(p-2)+3} \right).$$

Finally, for c_p'' sufficiently small, the remaining summation is bounded by a series of the form $\sum_{d \geq 1} (1-\epsilon)^d d^K$ for some constant K, which is a convergent series. Thus we find $\mathrm{Corr}_{\leq D}(\mathbb{P})^2 \leq 2\|\gamma\|^2 = O(n^{1/2})$, completing the proof. □

6 Application 2: Distinguishing Wigner from Wishart Tensors

We first state the result we will prove formally (combining the upper and lower bound claims from the Introduction in Theorems 1.9 and 1.10).

Theorem 6.1 (Low-degree analysis for Wigner vs. Wishart detection)**.** *Suppose that μ_n are probability measures on $\mathrm{Sym}^p(\mathbb{R}^n)$ satisfying the following properties for $A \sim \mu_n$:*

1. *For all $i \in [n]^p$ having a repeated entry, $A_{i_1,\ldots,i_p} = 0$ almost surely.*
2. *There is a constant $C > 0$ such that, for all $i \in [n]^p$, $|A_{i_1,\ldots,i_p}| \leq C$ almost surely.*
3. *$\|A\|_F^2 = n^p$ almost surely.*

Let $Z_1, \ldots, Z_r \sim \mathrm{Gin}(n, 1/n)$ be i.i.d. and $A_1, \ldots, A_r \sim \mu_n$ be i.i.d., and write $\mathbb{P} = \mathbb{P}_{n,r}$ for the law of $r^{-1/2} \sum_{j=1}^r Z_j \cdot A_j$. There is a constant $a_{p,C} > 0$ such that the following holds. Suppose that $D = D(n) \leq \sqrt{n/2p^2}$ is given and $r = r(n)$ satisfies

$$r \geq a_{p,C} \cdot \left\{ \begin{array}{ll} n^p & \text{if } p \text{ is odd,} \\ n^{\frac{3}{2}p} & \text{if } p \text{ is even} \end{array} \right\}. \tag{100}$$

Then, $\mathrm{Adv}_{\leq D}(\mathbb{Q} = \mathrm{Wig}, \mathbb{P}_{n,r(n)}) = O(1)$.

Further, suppose that μ_n is supported on a single tensor A having all entries with no repeated indices in their position equal to $c = c(p,n) > 0$. Let $D = 3$ if p is even and $D = 4$ if p is odd, and suppose that $r = r(n)$ satisfies

$$r \ll \left\{ \begin{array}{ll} n^p & \text{if } p \text{ is odd,} \\ n^{\frac{3}{2}p} & \text{if } p \text{ is even} \end{array} \right\}. \tag{101}$$

Then, $\mathrm{Adv}_{\leq D}(\mathbb{Q} = \mathrm{Wig}, \mathbb{P}_{n,r(n)}) = \omega(1)$.

We establish some preliminary facts about the Ginibre ensemble.

Proposition 6.2. $\mathrm{Gin}(n,\sigma^2)$ *is orthogonally invariant: if* $Z \sim \mathrm{Gin}(n,\sigma^2)$ *and* $Q \in \mathcal{O}(n)$*, then* QZ *and* ZQ *both again have the law* $\mathrm{Gin}(n,\sigma^2)$.

Corollary 6.3. *For any random tensor T, the law of $Z \cdot T$ when $Z \sim \mathrm{Gin}(n, 1/n)$ independently of T is orthogonally invariant.*

A version of the Weingarten formula for $\mathbb{E}_{Q\sim\mathrm{Haar}(n)} Q^{\otimes \ell}$ (see Appendix C) also holds for tensor powers of $Z \sim \mathrm{Gin}(n,\sigma^2)$. In fact, the formula is simpler: it only contains the leading order terms of the Weingarten formula.

Proposition 6.4. $\mathbb{E}_{Z\sim\mathrm{Gin}(n,\sigma^2)} Z^{\otimes\ell} = \sigma^\ell \sum_\mu w(\mu) \otimes w(\mu)$.

Proof. The result follows directly from an application of the Isserlis–Wick formula. □

This simple fact implies remarkable simplifications for the cumulants of a tensor formed as $Z \cdot T$.

Corollary 6.5. *Let G be a p-regular multigraph on d vertices. Then,*

$$\mathop{\mathbb{E}}_{Z\sim\mathrm{Gin}(n,1/n)} \kappa_G(Z \cdot T) = \frac{n^{\underline{pd/2}}}{n^{pd/2}} m_G(T). \tag{102}$$

Moreover, let $G^{(0)} := G \setminus \mathrm{Frob}(G)$ be G with all Frobenii removed. Then,

$$\mathop{\mathbb{E}}_{Z\sim\mathrm{Gin}(n,1/n)} \kappa^c_G(Z \cdot T) = \frac{n^{\underline{pd/2}}}{n^{pd/2}} m_{G^{(0)}}(T)(\|T\|_F^2 - n^p)^{|\mathrm{Frob}(G)|}. \tag{103}$$

Proof. Write $b := pd/2$ as usual. For the first claim, we observe by Corollary 6.3 that

$$\begin{aligned} \mathop{\mathbb{E}}_{Z\sim\mathrm{Gin}(n,1/n)} \kappa_G(Z \cdot T) &= \mathop{\mathbb{E}}_{Z\sim\mathrm{Gin}(n,1/n)} m^!_G(Z \cdot T) \\ &= \mathop{\mathbb{E}}_{Z\sim\mathrm{Gin}(n,1/n)} \sum_{i_1,\ldots,i_b \text{ distinct}} \langle e_{i_1}^{\otimes 2} \otimes \cdots \otimes e_{i_b}^{\otimes 2}, T^{\otimes d} Z^{\otimes pd} \rangle \\ &= \frac{1}{n^b} \sum_{i_1,\ldots,i_b \text{ distinct}} \sum_\mu \langle e_{i_1}^{\otimes 2} \otimes \cdots \otimes e_{i_b}^{\otimes 2}, T^{\otimes d} w(\mu) \otimes w(\mu) \rangle \quad \text{(Proposition 6.4)} \end{aligned}$$

and here only the matching that corresponds to the one of equal indices in the first term in the inner product contributes, whereby

$$\begin{aligned} &= \frac{1}{n^b} \sum_{i_1,\ldots,i_b \text{ distinct}} m_G(T) \\ &= \frac{n^{\underline{b}}}{n^b} m_G(T), \end{aligned}$$

as claimed. For the second result, we first expand by definition over subsets of Frobenii, and then use the first result on each term:

$$\begin{aligned}\mathop{\mathbb{E}}_{Z\sim\mathrm{Gin}(n,1/n)}\kappa_G^c(Z\cdot T) &= \sum_{S\subseteq\mathrm{Frob}(G)}(-1)^{|S|}(n-b+p|S|)^{\underline{p|S|}}\mathop{\mathbb{E}}_{Z\sim\mathrm{Gin}(n,1/n)}\kappa_{G\setminus S}(Z\cdot T)\\ &= \sum_{S\subseteq\mathrm{Frob}(G)}(-1)^{|S|}(n-b+p|S|)^{\underline{p|S|}}\frac{n^{\underline{b-p|S|}}}{n^{b-p|S|}}m_{G\setminus S}(T)\end{aligned}$$

and by multiplicativity of the ordinary graph moments,

$$= m_{G^{(0)}}(T)\sum_{S\subseteq\mathrm{Frob}(G)}(-1)^{|S|}(n-b+p|S|)^{\underline{p|S|}}\frac{n^{\underline{b-p|S|}}}{n^{b-p|S|}}(\|T\|_F^2)^{|\mathrm{Frob}(G)|-|S|}$$

and writing $f := |\mathrm{Frob}(G)|$ and introducing $s := |S|$, we have

$$\begin{aligned}&= m_{G^{(0)}}(T)\sum_{s=0}^{f}\binom{f}{s}(-1)^s(n-b+ps)^{\underline{ps}}\frac{n^{\underline{b-ps}}}{n^{b-ps}}(\|T\|_F^2)^{f-s}\\ &= m_{G^{(0)}}(T)\sum_{s=0}^{f}\binom{f}{s}(-1)^s\frac{(n-b+ps)!}{(n-b)!}\frac{n!}{(n-b+ps)!}\frac{1}{n^{b-ps}}(\|T\|_F^2)^{f-s}\\ &= \frac{n^{\underline{b}}}{n^b}m_{G^{(0)}}(T)\sum_{s=0}^{f}\binom{f}{s}(-n^p)^s(\|T\|_F^2)^{f-s}\\ &= \frac{n^{\underline{b}}}{n^b}m_{G^{(0)}}(T)(\|T\|_F^2-n^p)^f\end{aligned}$$

by the binomial theorem, completing the proof. □

Proof of Theorem 6.1. We first express the expectations of the centered cumulants of $\mathbb{P}$ in terms of those of $\mathrm{Gin}(n,1/n)\cdot\mu$, and evaluate the latter using Corollary 6.5:

$$\begin{aligned}\mathop{\mathbb{E}}_{T\sim\mathbb{P}}\kappa_G^c(T) &= \mathop{\mathbb{E}}_{\substack{Z_i\sim\mathrm{Gin}(n,1/n)\\A_i\sim\mu}}\kappa_G^c\left(\frac{1}{\sqrt{r}}\sum_{i=1}^{r}Z_i\cdot A_i\right)\\ &= \sum_{G_1\sqcup\cdots\sqcup G_r=G}\frac{n^{\underline{b}}}{n^{\underline{b_1}}\cdots n^{\underline{b_r}}}\mathop{\mathbb{E}}_{\substack{Z_i\sim\mathrm{Gin}(n,1/n)\\A_i\sim\mu}}\prod_{i=1}^{r}\kappa_{G_i}^c\left(\frac{1}{\sqrt{r}}Z_i\cdot A_i;-\frac{1}{r}\mathbf{1}_{\mathrm{Frob}}\right)\\ &= r^{-d/2}\sum_{G_1\sqcup\cdots\sqcup G_r=G}\frac{n^{\underline{b}}}{n^{\underline{b_1}}\cdots n^{\underline{b_r}}}\prod_{i=1}^{r}\mathop{\mathbb{E}}_{\substack{Z\sim\mathrm{Gin}(n,1/n)\\A\sim\mu}}\kappa_{G_i}^c(Z\cdot A)\\ &= \frac{n^{\underline{b}}}{n^b}r^{-d/2}\sum_{G_1\sqcup\cdots\sqcup G_r=G}\prod_{i=1}^{r}\mathop{\mathbb{E}}_{A\sim\mu}m_{G_i^{(0)}}(A)(\|A\|_F^2-n^p)^{|\mathrm{Frob}(G_i)|}.\end{aligned}$$

We make a few observations.

First, if G contains any self-loops, then some $G_i^{(0)}$ contains a self-loop in every term of the sum. For this i, by our assumption that A is zero on entries with repeated indices, we have $m_{G_i^{(0)}}(A) = 0$ almost surely. Thus the entire sum is zero, so $\mathbb{E}_{T\sim\mathbb{P}} \kappa_G^c(T) = 0$ whenever G has self-loops.

Second, and similarly, if G contains any Frobenii, then some exponent of $\|A\|_F^2 - n^p$ is positive in each term in the sum. By our assumption we have $\|A\|_F^2 - n^p = 0$ almost surely when $A \sim \mu$, so $\mathbb{E}_{T\sim\mathbb{P}} \kappa_G^c(T) = 0$ whenever G has any Frobenii as well.

Lastly, let us consider the case where G has neither self-loops nor Frobenii. We may bound the ordinary graph moments naively by using our assumption that the entries of $A \sim \mu$ are almost surely uniformly bounded by a constant $C > 0$. This implies

$$|m_G(A)| \leq C^{|V(G)|} n^{|E(G)|}, \tag{104}$$

since $m_G(A)$ is a sum of $n^{|E(G)|}$ terms, each of which is a product of $|V(G)|$ factors of size at most C. Using this, we find

$$\begin{aligned} |\mathop{\mathbb{E}}_{T\sim\mathbb{P}} \kappa_G^c(T)| &\leq C^d n^b r^{-d/2} \sum_{G_1 \sqcup \cdots \sqcup G_r = G} 1 \\ &\leq C^d n^b r^{-d/2 + |\mathrm{conn}(G)|} \end{aligned}$$

where we have used that the number of partitions $G_1 \sqcup \cdots \sqcup G_r = G$ is just the number of assignments of each connected component of G to one of r bins, or $r^{|\mathrm{conn}(G)|}$.

Since G has no self-loops or Frobenii, it has no connected components on two vertices. Note that, depending on the parity of p, the smallest possible size of a connected component in G will differ: when p is even then it is 3, but when p is odd then it is 4. This will ultimately lead to the different thresholds depending on the parity of p in our result. Let us give this number a name:

$$\xi = \xi(p) := \left\{ \begin{array}{ll} 3 & \text{if } p \text{ is even,} \\ 4 & \text{if } p \text{ is odd} \end{array} \right\}. \tag{105}$$

Substituting the above bound into Corollary 4.24, we find

$$\begin{aligned} \mathrm{Adv}_{\leq D}(\mathrm{Wig}, \mathbb{P}) &\lesssim \sum_{d=0}^{D} (C^2 r^{-1} n^{p/2})^d \sum_{\substack{G \in \mathcal{G}_{d,p} \\ |V(K)| \geq \xi \text{ for all } K \in \mathrm{conn}(G)}} r^{2|\mathrm{conn}(G)|} \\ &\leq \sum_{d=0}^{D} (C^2 r^{-1} n^{p/2})^d \sum_{\ell=1}^{d/\xi} \sum_{\substack{\xi \leq a_1 \leq \cdots \leq a_\ell \\ a_1 + \cdots + a_\ell = d}} r^{2\ell} \prod_{i=1}^{\ell} |\mathcal{G}_{a_i,p}| \end{aligned}$$

and using Proposition 5.9, we have

$$\leq \sum_{d=0}^{D} (C^2 e^{p+1} p^{-p/2} r^{-1} n^{p/2})^d \sum_{\ell=1}^{d/\xi} \sum_{\substack{\xi \leq a_1 \leq \cdots \leq a_\ell \\ a_1 + \cdots + a_\ell = d}} r^{2\ell} \prod_{i=1}^{\ell} a_i^{\frac{p-2}{2} a_i}$$

and a convexity argument shows that the inner product over a_i is maximized when $a_1 = \cdots = a_{\ell-1} = \xi$ and $a_\ell = d - (\ell-1)\xi$. We may therefore continue bounding

$$\leq \sum_{d=0}^{D} (C^2 e^{p+1} p^{-p/2} \xi^{\frac{p-2}{2}} r^{-1} n^{p/2})^d \sum_{\ell=1}^{d/\xi} \sum_{\substack{\xi \leq a_1 \leq \cdots \leq a_\ell \\ a_1 + \cdots + a_\ell = d}} r^{2\ell} d^{d-(\ell-1)\xi}$$

Now, since $d \leq D \leq n^{1/2}$ by assumption, while $r \geq n^3$ (provided we choose $a_{p,C} \geq 1$), we have $r^2 \geq n^6$ while $d^\xi \leq d^4 \leq n^2$, so $r^2 \geq d^\xi$ and the largest term of the remaining sum is the one where $\ell = d/\xi$. Using Proposition D.3 to bound the number of terms in the sum, we find

$$\begin{aligned} &\leq \sum_{d=0}^{D} (C^2 e^{p+1} p^{-p/2} \xi^{\frac{p-2}{2}} r^{-1} n^{p/2})^d \cdot \exp(O(\sqrt{d})) r^{2d/\xi} d^\xi \\ &\leq \sum_{d=0}^{\infty} d^\xi \left(C^2 e^{p+1} p^{-p/2} \xi^{\frac{p-2}{2}} \exp(O(1/\sqrt{d})) \cdot r^{-(1-\frac{2}{\xi})} n^{p/2} \right)^d. \end{aligned}$$

Thus we find that, provided that $r \geq a_p n^{\frac{p}{2} \cdot \frac{\xi}{\xi-2}}$, the base of the exponent in the above series will be strictly smaller than 1, so the sum will converge and and the proof will be complete. When p is even, then $\xi = 3$ and the exponent of n in the condition above is $3p/2$, while when p is odd, then $\xi = 4$ and the exponent is p, so this gives precisely the stated result in both cases.

Finally, for the upper bound, note that for this choice of deterministic A we have $m_G(A) \asymp n^{|E(G)|}$ for each G, since each term in the sum in $m_G(A)$ is equal and close to 1. We thus have for any G on d vertices that

$$\mathop{\mathbb{E}}_{T \sim \mathbb{P}} \kappa_G^c(T) \asymp n^{\underline{pd/2}} r^{-d/2+|\mathrm{conn}(G)|} \tag{106}$$

by the same calculations as above. Let us take $d = \xi \in \{3, 4\}$ and G any connected p-regular graph on ξ vertices (say, a triangle with every edge repeated $p/2$ times when p is even, or a complete graph on 4 vertices with each edge of one perfect matching repeated $p-2$ times and the other edges occurring once when p is odd). We have $|\mathrm{conn}(G)| = 1$. By Corollary 4.24, we may lower bound the advantage by the contribution of just the term corresponding to this graph G, which, since it is of constant size, has $|\mathrm{eAut}(G)| = O(1)$, so

$$\begin{aligned} \mathrm{Adv}_{\leq D}(\mathrm{Wig}, \mathbb{P}) &\gtrsim n^{-p\xi/2} \left(\mathop{\mathbb{E}}_{T \sim \mathbb{P}} \kappa_G^c(T) \right)^2 \\ &\gtrsim n^{-p\xi/2} (n^{\underline{p\xi/2}})^2 r^{-\xi+2} \end{aligned}$$

and using Proposition D.1 on the falling factorials gives

$$\gtrsim r^{-\xi+2} n^{p\xi/2},$$

which diverges by our assumption both when p is even and when p is odd. □

Remark 6.6. *It appears difficult to substantially relax the assumption that $\|A\|_F^2 = n^p$ exactly. Indeed, considering terms corresponding to G consisting* only *of Frobenii, if we had a bound $|\|A\|_F^2 - n^p| \leq K$ (or, speaking more roughly, if the typical scale of this difference is K) then we see that*

our bounds would yield $|\mathbb{E}\kappa_G^c(T)| \le C^d n^{pd/2} K^{d/2}$*, and this would be the best bound we can achieve regardless of the value of* r*. Thus even if* K *is a constant this would not give an* $O(1)$ *bound on the advantage. It would be more natural to assume that* $\|A\|_F^2 - n^p$ *is centered and* $O(n^p)$*-subgaussian, say, in our result, but (at least with a naive bounding strategy as we have taken) such an assumption would therefore not suffice to control the above terms when* $D = \omega(1)$*. The same issue arises with relaxing the assumption that the diagonal entries of* $A \sim \mu$ *are exactly zero.*

Acknowledgments

Some of this work was carried out at the Simons Institute for the Theory of Computing and the Banff International Research Station for Mathematical Innovation and Discovery. We are grateful to Denis Bernard, Guy Bresler, Josh Grochow, Sam Hopkins, Nick Read, Tselil Schramm, Guilhem Semerjian, Piotr Śniady, and Dan Spielman for helpful discussions. C.M. is especially grateful to Alex Russell for his hospitality during a sabbatical long ago when they learned about free probability and explored diagrammatic methods.

References

[ABAČ13] Antonio Auffinger, Gérard Ben Arous, and Jiří Černỳ. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201, 2013.

[ABP73] M. Atiyah, R. Bott, and V.K. Patodi. On the heat equation and the index theorem. *Inventiones Math.*, 19:279–330, 1973.

[ADGM17] Anima Anandkumar, Yuan Deng, Rong Ge, and Hossein Mobahi. Homotopy analysis for tensor PCA. In *Conference on Learning Theory*, pages 79–104. PMLR, 2017.

[AGVP23] Octavio Arizmendi, Jorge Garza-Vargas, and Daniel Perales. Finite free cumulants: multiplicative convolutions, genus expansion and infinitesimal distributions. *Transactions of the American Mathematical Society*, 376(06):4383–4420, 2023.

[AHH12] Andris Ambainis, Aram W Harrow, and Matthew B Hastings. Random tensor theory: extending random matrix theory to mixtures of random product states. *Communications in Mathematical Physics*, 310(1):25–74, 2012.

[AMP16] Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: norm bounds and applications. *arXiv preprint arXiv:1604.03423*, 2016.

[AP18] Octavio Arizmendi and Daniel Perales. Cumulants for finite free convolution. *Journal of Combinatorial Theory, Series A*, 155:244–266, 2018.

[Ban10] Teodor Banica. The orthogonal Weingarten formula in compact form. *Letters in Mathematical Physics*, 91(2):105–118, 2010.

[BB20] Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. In *33rd Annual Conference on Learning Theory (COLT 2020)*, pages 648–847. PMLR, 2020.

[BB24] Kiril Bangachev and Guy Bresler. On the Fourier coefficients of high-dimensional random geometric graphs. *arXiv preprint arXiv:2402.12589*, 2024.

[BBH21] Matthew Brennan, Guy Bresler, and Brice Huang. De Finetti-style results for Wishart matrices: Combinatorial structure and phase transitions. *arXiv preprint arXiv:2103.14011*, 2021.

[BBN20] Matthew Brennan, Guy Bresler, and Dheeraj Nagaraj. Phase transitions for detecting latent geometry in random graphs. *Probability Theory and Related Fields*, 178(3):1215–1289, 2020.

[BCRT20] Giulio Biroli, Chiara Cammarota, and Federico Ricci-Tersenghi. How to iron out rough landscapes and get optimal performances: averaged gradient descent and its application to tensor PCA. *Journal of Physics A: Mathematical and Theoretical*, 53(17):174003, 2020.

[BDER16] Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

[BEAH+22] Afonso S Bandeira, Ahmed El Alaoui, Samuel Hopkins, Tselil Schramm, Alexander S Wein, and Ilias Zadik. The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. *Advances in Neural Information Processing Systems*, 35:33831–33844, 2022.

[BGG+16] Vijay V. S. P. Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Multiplicative approximations for polynomial optimization over the unit sphere. *Electron. Colloquium Comput. Complex.*, TR16-185, 2016.

[BGL17] Vijay Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. Sum-of-squares certificates for maxima of random tensors on the sphere. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2017*, volume 81 of *LIPIcs*, pages 31:1–31:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.

[BGS13] Valentin Bonzom, Razvan Gurau, and Matteo Smerlak. Universality in p-spin glasses with correlated disorder. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(02):L02003, 2013.

[BHK+19] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

[BM18] Nicolas Bousquet and Arnaud Mary. Reconfiguration of graphs with connectivity constraints. In *International Workshop on Approximation and Online Algorithms*, pages 295–309. Springer, 2018.

[Bol82] Béla Bollobás. The asymptotic number of unlabelled regular graphs. *Journal of the London Mathematical Society*, 2(2):201–206, 1982.

[Bon24] Remi Bonnin. Universality of the Wigner-Gurau limit for random tensors. *arXiv preprint arxiv:2404.14144*, 2024.

[Bra37] Richard Brauer. On algebras which are connected with the semisimple continuous groups. *Annals of Mathematics*, 38(4):857–872, 1937.

[CDN14] Chun-Feng Cui, Yu-Hong Dai, and Jiawang Nie. All real eigenvalues of symmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 35(4):1582–1601, 2014.

[CGL23a] Benoît Collins, Razvan Gurau, and Luca Lionni. The tensor Harish-Chandra–Itzykson–Zuber integral II: Detecting entanglement in large quantum systems. *Communications in Mathematical Physics*, pages 1–48, 2023.

[CGL23b] Benoît Collins, Razvan G Gurau, and Luca Lionni. The tensor Harish-Chandra–Itzykson–Zuber integral I: Weingarten calculus and a generalization of monotone Hurwitz numbers. *Journal of the European Mathematical Society*, 2023.

[CMSS06] Benoit Collins, James A. Mingo, Piotr Sniady, and Roland Speicher. Second order freeness and fluctuations of random matrices, III. Higher order freeness and free cumulants. *Documenta Mathematica*, 12, 07 2006.

[COGHK$^+$22] Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, Alexander S Wein, and Ilias Zadik. Statistical and computational phase transitions in group testing. In *Conference on Learning Theory*, pages 4764–4781. PMLR, 2022.

[Col03] Benoît Collins. Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability. *International Mathematics Research Notices*, 2003(17):953–982, 2003.

[CŚ06] Benoît Collins and Piotr Śniady. Integration with respect to the haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006.

[DdL$^+$22] Jingqiu Ding, Tommaso d'Orsi, Chih-Hung Liu, Stefan Tiegel, and David Steurer. Fast algorithm for overcomplete order-3 tensor decomposition. *arXiv preprint arXiv:2202.06442*, 2022.

[DDL23] Jian Ding, Hang Du, and Zhangsong Li. Low-degree hardness of detection for correlated Erdős-Rényi graphs. *arXiv preprint arXiv:2311.15931*, 2023.

[DHS20] Jingqiu Ding, Samuel B Hopkins, and David Steurer. Estimating rank-one spikes from heavy-tailed noise via self-avoiding walks. *arXiv preprint arXiv:2008.13735*, 2020.

[DMW23] Abhishek Dhawan, Cheng Mao, and Alexander S Wein. Detection of dense subhypergraphs by low-degree polynomials. *arXiv preprint arXiv:2304.08135*, 2023.

[Evn21] Oleg Evnin. Melonic dominance and the largest eigenvalue of a large random tensor. *Letters in Mathematical Physics*, 111(3):66, 2021.

[Ger31] S. Geršgorin. Über die abgrenzung der eigenwerte einer matrix. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, pages 749–754, 1931.

[GJJ+20] Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for Sherrington-Kirkpatrick via planted affine planes. In *61st Annual Symposium on Foundations of Computer Science (FOCS 2020)*, pages 954–965. IEEE, 2020.

[GJW24] David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Hardness of random optimization problems for Boolean circuits, low-degree polynomials, and Langevin dynamics. *SIAM Journal on Computing*, 53(1):1–46, 2024.

[Gur14] Razvan Gurau. Universality for random tensors. In *Annales de l'IHP Probabilités et statistiques*, volume 50, pages 1474–1525, 2014.

[Gur17] Răzvan Gheorghe Gurău. *Random tensors*. Oxford University Press, 2017.

[Gur20] Razvan Gurau. On the generalization of the Wigner semicircle law to real symmetric tensors. *arXiv preprint arXiv:2004.02660*, 2020.

[GW98] Roe Goodman and Nolan R. Wallach. *Representations and Invariants of the Classical Groups*. Cambridge University Press, 1998.

[Hak63] Seifollah Louis Hakimi. On realizability of a set of integers as degrees of the vertices of a linear graph II. Uniqueness. *Journal of the Society for Industrial and Applied Mathematics*, 11(1):135–147, 1963.

[Has20] Matthew B Hastings. Classical and quantum algorithms for tensor principal component analysis. *Quantum*, 4:237, 2020.

[HKP+17] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2017.

[Hop18] Samuel Hopkins. *Statistical inference and the sum of squares method*. PhD thesis, Cornell University, 2018.

[HR18] Godfrey H Hardy and Srinivasa Ramanujan. Asymptotic formulaæ in combinatory analysis. *Proceedings of the London Mathematical Society*, 2(1):75–115, 1918.

[HS17] Samuel B Hopkins and David Steurer. Efficient Bayesian estimation from few samples: community detection and related problems. In *58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 379–390. IEEE, 2017.

[HSS15] Samuel B. Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In Peter Grünwald, Elad Hazan, and Satyen Kale, editors, *Proc. 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 956–1006, 2015.

[HSSS16] Samuel B. Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, pages 178—191, 2016.

[Iss18] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12:134–139, 1918.

[JLM20] Aukosh Jagannath, Patrick Lopatto, and Leo Miolane. Statistical thresholds for tensor PCA. *Annals of Applied Probability*, 30(4):1910–1933, 2020.

[JP21] Chris Jones and Aaron Potechin. Almost-orthogonal bases for inner product polynomials. *arXiv preprint arXiv:2107.00216*, 2021.

[JPRX23] Chris Jones, Aaron Potechin, Goutham Rajendran, and Jeff Xu. Sum-of-squares lower bounds for densest k-subgraph. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 84–95, 2023.

[KVWX23] Pravesh Kothari, Santosh S Vempala, Alexander S Wein, and Jeff Xu. Is planted coloring easier than planted clique? In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5343–5372. PMLR, 2023.

[KWB22] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In Paula Cerejeiras and Michael Reissig, editors, *Mathematical Analysis, its Applications and Computation*, pages 1–50, Cham, 2022. Springer International Publishing.

[LMSY22] Siqi Liu, Sidhanth Mohanty, Tselil Schramm, and Elizabeth Yang. Testing thresholds for high-dimensional sparse random geometric graphs. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 672–677, 2022.

[Mas14] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703, 2014.

[MFC+19] Antoine Maillard, Laura Foini, Alejandro Lage Castellanos, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. High-temperature expansions and message passing algorithms. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(11):113301, 2019.

[Mik20] Dan Mikulincer. A CLT in Stein's distance for generalized Wishart matrices and higher order tensors. *arXiv preprint arXiv:2002.10846*, 2020.

[MR95] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995.

[MR11] Cristopher Moore and Alexander Russell. A graph integral formulation of the circuit partition polynomial. *Combinatorics, Probability & Computing*, 20(6):911, 2011.

[MR14] Andrea Montanari and Emile Richard. A statistical model for tensor PCA. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, page 2897–2905. MIT Press, 2014.

[MS17] James A. Mingo and Roland Speicher. *Free Probability and Random Matrices*, volume 35 of *Fields Institute Monographs*. Springer, 2017.

[MW19] Ankur Moitra and Alexander S Wein. Spectral methods from tensor networks. In *51st Annual ACM SIGACT Symposium on Theory of Computing (STOC 2019)*, pages 926–937, 2019.

[NS06] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 13. Cambridge University Press, 2006.

[NŚ11] Jonathan Novak and Piotr Śniady. What is...a free cumulant? *Notices of the American Mathematical Society*, 58:300–301, 2011.

[OR20] Mohamed Ouerfelli and Vincent Rivasseau. A new framework for tensor PCA based on trace invariants. 2020.

[OTR22] Mohamed Ouerfelli, Mohamed Tamaazousti, and Vincent Rivasseau. Random tensor theory for tensor decomposition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(7):7913–7921, 2022.

[Oue22] Mohamed Ouerfelli. *New perspectives and tools for Tensor Principal Component Analysis and beyond*. PhD thesis, Université Paris-Saclay, 2022.

[PR22] Aaron Potechin and Goutham Rajendran. Sub-exponential time sum-of-squares lower bounds for principal components analysis. *Advances in Neural Information Processing Systems*, 35:35724–35740, 2022.

[Pro76] Claudio Procesi. The invariant theory of $n \times n$ matrices. *Advances in mathematics*, 19(3):306–381, 1976.

[QCC18] Liqun Qi, Haibin Chen, and Yannan Chen. *Tensor eigenvalues and their applications*, volume 39. Springer, 2018.

[Qi05] Liqun Qi. Eigenvalues of a real supersymmetric tensor. *Journal of Symbolic Computation*, 40(6):1302–1324, 2005.

[Qi07] Liqun Qi. Eigenvalues and invariants of tensors. *Journal of Mathematical Analysis and Applications*, 325(2):1363–1377, 2007.

[RRS17a] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random CSPs below the spectral threshold. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 121–131, 2017.

[RRS17b] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random CSPs below the spectral threshold. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, page 121–131, 2017.

[RSWY23] Cynthia Rush, Fiona Skerman, Alexander S Wein, and Dana Yang. Is it easier to count communities than find them? In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2023.

[Sem24] Guilhem Semerjian. Matrix denoising: Bayes-optimal estimators via low-degree polynomials. *arXiv preprint arxiv:2402.16719*, 2024.

[Sub17] Eliran Subag. The complexity of spherical p-spin models — a second moment approach. *The Annals of Probability*, 45(5):3385–3450, 2017.

[SW22] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.

[WAM19] Alexander S. Wein, Ahmed El Alaoui, and Cristopher Moore. The Kikuchi hierarchy and tensor PCA. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019*, pages 1446–1468. IEEE Computer Society, 2019.

[Wei23] Alexander S Wein. Average-case complexity of tensor decomposition for low-degree polynomials. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, pages 1685–1698, 2023.

[Wen88] Hans Wenzl. On the structure of Brauer's centralizer algebras. *Annals of Mathematics*, 128(1):173–193, 1988.

[Wey46] Hermann Weyl. *The Classical Groups: Their Invariants and Representations*. 1946.

[Wic50] G. C. Wick. The evaluation of the collision matrix. *Physical Review*, 80:268–272, 1950.

[Wil99] Todd G Will. Switching distance between graphs with the same degrees. *SIAM Journal on Discrete Mathematics*, 12(3):298–306, 1999.

[WZ24] Yuchen Wu and Kangjie Zhou. Sharp analysis of power iteration for tensor PCA. *arXiv preprint arXiv:2401.01047*, 2024.

[ZJ09] Paul Zinn-Justin. Jucys-Murphy elements and Weingarten matrices. *arXiv preprint arXiv:0907.2719*, 2009.

A Characterizations of Invariants

Our goal in this Appendix is to give diagrammatic proofs of Theorems 3.2, 3.6, and 3.15. We first discuss the first two Theorems on invariant polynomials. Recall that we say that a function $f : \mathrm{Sym}^p(\mathbb{R}^n) \to \mathbb{R}$ is *invariant* if, for all $T \in \mathrm{Sym}^p(\mathbb{R}^n)$ and all $Q \in \mathcal{O}(n)$, $f(Q \cdot T) = f(T)$. Similarly, a function $f : \mathrm{Sym}^{p_1}(\mathbb{R}^n) \times \mathrm{Sym}^{p_2}(\mathbb{R}^n) \times \ldots \times \mathrm{Sym}^{p_m}(\mathbb{R}^n) \to \mathbb{R}$ is invariant if, for any symmetric tensors $T_1, \ldots, T_m$ of arity $p_1, \ldots, p_m$ respectively, $f(T_1, \ldots, T_m) = f(Q \cdot T_1, \ldots, Q \cdot T_m)$ for all $Q \in \mathcal{O}(n)$.

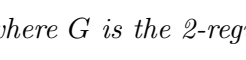
A priori, restricting T to $\mathrm{Sym}^p(\mathbb{R}^n)$ could weaken the definition of invariance. There are indeed polynomials that are invariant on symmetric tensors but not on general tensors: for instance, for 2×2 matrices, the function $m_{11}^2 + 2m_{12}^2 + m_{22}^2$ is the Frobenius norm $\|M\|_F^2$ and hence invariant, but only if M is symmetric. However, the following proposition shows that we can always replace such a polynomial with one that is invariant on general tensors by symmetrizing over $\mathcal{O}(n)$. In what follows we will identify invariant polynomials with this symmetrized version.

Proposition A.1. *Suppose f is an invariant polynomial. Then there is a polynomial g of the same degree which coincides with f on $\mathrm{Sym}^p(\mathbb{R}^n)$ and which is invariant on all tensors, i.e., $g(Q \cdot T) = g(T)$ for all $T \in (\mathbb{R}^n)^{\otimes p}$.*

Proof. Let $g(T) = \mathbb{E}_Q f(Q \cdot T)$, where the expectation is over the Haar measure on $\mathcal{O}(n)$. □

Next, we show Theorem 3.2, which says that any invariant polynomial is a linear combination of graph moments. This generalizes a similar fact for matrices: the invariant polynomials of a matrix are precisely the symmetric polynomials of its eigenvalues. These in turn can be written in terms of spectral moments, or equivalently traces of T's matrix powers.

Theorem A.2. *Let R be the ring of polynomials $f : \mathrm{Sym}^2(\mathbb{R}^n) \to \mathbb{R}$ in the entries of a matrix so that, for any orthogonal matrix Q, $f(T) = f(Q^\top T Q)$. Then R is generated by the polynomials $m_\ell(T) = \mathrm{tr}(T^\ell)$ for $\ell \geq 0$. In particular, any invariant homogeneous polynomial $f(T)$ of degree d is a linear combination of graph moments $\sum_i \alpha_i m_{G_i}(T)$ where each G_i is a disjoint union of cycles with a total of d vertices.*

Example A.3. *For instance, $f(T) = (\mathrm{tr}\, T)^2 \,\mathrm{tr}(T^2)\,\mathrm{tr}(T^3)$, which is a polynomial of degree 7, is $m_G(T)$ where G is the 2-regular multigraph*  .

Proof of Theorem 3.2. We illustrate the proof in Figure 10. First note that any polynomial f of degree d in a p-ary tensor T can be written as an inner product between $T^{\otimes d}$ and a vector of coefficients C, which we can also view as a dp-ary tensor.

Now, if f is invariant, it remains the same if we place a copy of any orthogonal matrix $Q^\top$ on each of the dp edges of $T^{\otimes d}$. But this is equivalent to applying $Q^{\otimes dp}$ to C: that is,

$$f(Q^\top \cdot T) = \langle Q^\top \cdot T^{\otimes d}, C\rangle = \langle T^{\otimes d}, Q \cdot C\rangle = \langle T^{\otimes d}, Q^{\otimes dp} C\rangle = \langle T^{\otimes d}, C\rangle = f(T)\,.$$

Since $f(T) = \langle T^{\otimes d}, Q^{\otimes dp} C\rangle$ for any $Q \in \mathcal{O}(n)$, we can symmetrize C by taking the expectation over the Haar measure, obtaining

$$f(T) = \langle T^{\otimes d}, \Pi_{dp} C\rangle\,, \tag{107}$$

where

$$\Pi_\ell = \mathop{\mathbb{E}}_{Q \in \mathcal{O}(n)} Q^{\otimes \ell}\,, \tag{108}$$

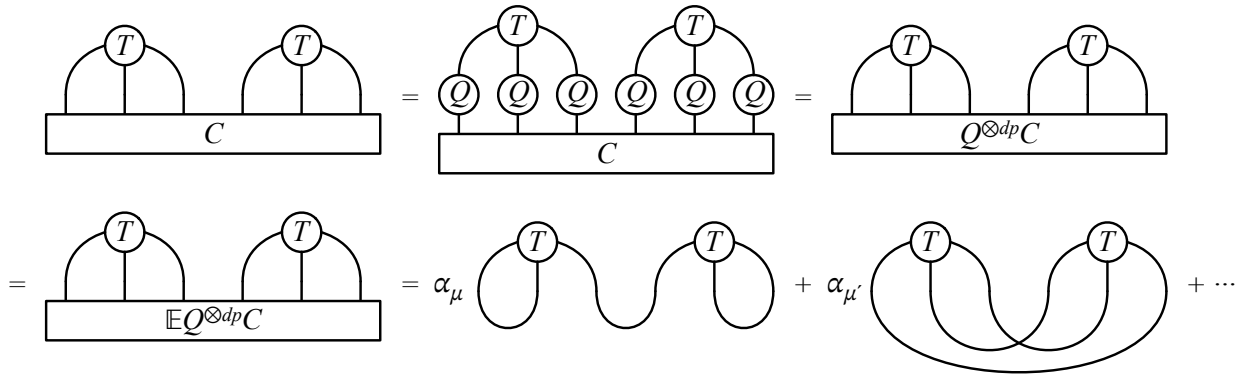


Figure 10: The proof of Theorem 3.2. Any homogeneous polynomial of degree d can be written as the inner product of $T^{\otimes d}$ with a dp-ary tensor of coefficients C. By hypothesis the function is unchanged if we symmetrize these coefficients by conjugating them with a Haar-random orthogonal matrix Q. But their image under the projection operator $\mathbb{E}\, Q^{\otimes dp}$ is a linear combination of matching vectors, each of which induces a multigraph connecting the copies of T.

the projection operator that we discussed in Section 3.3. Specifically, Π_ℓ projects onto the *trivial subspace* under the action of $\mathcal{O}(n)$, i.e., the set of vectors $w \in (\mathbb{R}^n)^{\otimes \ell}$ such that $Q^\top \cdot w = Q^{\otimes \ell} w = w$ for all $Q \in \mathcal{O}(n)$. As we present there, it follows from representation theory that Π_ℓ may be written $\Pi_\ell = \sum_{\mu,\nu} \mathrm{Wg}_{\mu,\nu} w(\mu) \otimes w(\nu)$ for some $\mathrm{Wg}_{\mu,\nu} \in \mathbb{R}$, μ, ν perfect matchings of $[\ell]$, and $w(\mu)$ the indicator vector of pairs of indices being equal under the matching μ.

Since the image of Π_ℓ is spanned by the $w(\mu)$, the symmetrized coefficients $\Pi_{dp} C$ in (107) are a linear combination $\sum_\mu \alpha_\mu w(\mu)$. But for each μ, $\langle T^{\otimes d}, w(\mu) \rangle$ is exactly a graph moment $m_{G(\mu)}(T)$, where $G(\mu)$ is the multigraph $G(\mu)$ formed by matching the dp half-edges of its vertices according to μ. (The reader may be familiar with the configuration model of random graphs [MR95] where μ is uniformly random.) Thus

$$f(T) = \sum_\mu \alpha_\mu m_{G(\mu)}(T)\,.$$

as illustrated in Figure 10. This completes the proof for invariant polynomials $f(T)$ of a single p-ary tensor. □

Proof of Theorem 3.6. For invariant polynomials of multiple tensors $f(T_1, \ldots, T_m)$, if p is homogeneous of degree d_i in each T_i and each T_i has arity k_i, we can write it as an inner product

$$f(T_1, \ldots, T_m) = \left\langle \bigotimes_{i=1}^m T_i^{\otimes d_i}, C \right\rangle ,$$

where the tensor of coefficients C has total arity $\ell = \sum_i d_i k_i$. The argument then goes through as before; see Figure 11 for an example. □

Again by similar arguments, we may also prove Theorem 3.15 on equivariant polynomials and open graph moments.

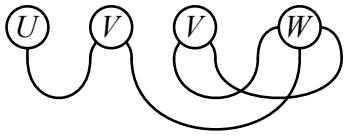


Figure 11: Generalizing the argument of Theorem 3.2 to mixed moments. The inner product of $U \otimes V \otimes V \otimes W$ with the matching vector shown gives the multigraph on the right in Figure 3 and the mixed moment (14).

Proof of Theorem 3.15. Let W be an ℓ-ary tensor of indeterminates. Then if $f(T)$ is ℓ-ary and equivariant, the inner product $P = \langle f(T), W\rangle$ is an invariant function of T and W. Similarly, if $f(T_1, \dots, T_m)$ is equivariant, then $F = \langle f(T_1, \dots, T_m), W\rangle$ is an invariant function of $\{T_1, \dots, T_m, W\}$.

If $f(T)$ is a homogeneous polynomial, then $F(T, W)$ is a homogeneous polynomial of degree 1 in W, i.e., which is multilinear in W's entries. By Theorem 3.2, F is a linear combination of mixed moments $m_G(T, W)$ where W corresponds to a single vertex w of degree ℓ and G's other vertices have degree p. Then each entry of $f(T)$ is a partial derivative of $F(T, W)$ by the corresponding entry of W,

$$f(T)_{i(E')} = \frac{\partial F(T, W)}{\partial W_{i(E')}}\,.$$

Taking this partial derivative removes W's vertex from G, leaving a p-regular open multigraph $G \setminus \{w\}$ with $|F| = \ell$ open edges. This completes the proof for $f(T)$, and mutatis mutandis for $f(T_1, \dots, T_m)$. □

B Properties of Wigner Tensors

B.1 Basic Properties

We give some further properties of $T \sim \mathrm{Wig}(p, n, \sigma^2)$. As a result of symmetrizing over all permutations π, the entries of T have different variances depending on the pattern of repetitions in their indices. Specifically, their variance is equal to the size of their stabilizer subgroup, i.e., the number of permutations in S_k which leave them unchanged.

Proposition B.1. *Let $T \sim \mathrm{Wig}(p, n, \sigma^2)$. For a given sequence of indices $i = (i_1, \dots, i_p) \in [n]^p$ and $j \in [n]$, let $c_j(i)$ be the number of times j appears in i. Then*

$$T_i \sim \mathcal{N}\left(0, \sigma^2 \prod_{j=1}^{n} c_j(i)!\right)\,.$$

Proof. Write $m = \prod_{j=1}^{n} c_j(i)!$. Then m is the number of permutations π such that $\pi(i) = i$, and the number of distinct $\pi(i)$ contributing to T_i in (5) is $p!/m$. For each of these the same $G_{\pi(i)}$ appears m times in (5), multiplying its variance by m^2. Since the $G_{\pi(i)}$ are independent, we then have $\operatorname{Var} T_i = (\sigma^2/p!) \times (p!/m) \times m^2 = \sigma^2 m$. □

This ensemble is orthogonally invariant, and for $p = 2$ we recover the Gaussian orthogonal ensemble of random matrix theory.

Proposition B.2. *For $T \sim \mathrm{Wig}(p, n, \sigma^2)$, for any $Q \in \mathcal{O}(n)$, $Q \cdot T$ has the same law as T. In particular, $T \sim \mathrm{Wig}(2, n, 1)$ has the law of the standard $n \times n$ Gaussian orthogonal ensemble, with $T_{ij} = T_{ji} \sim \mathcal{N}(0, 1 + \delta_{ij})$.*

B.2 Hardness of Computing Moments

In this section, we give the proof of Theorem 5.3 on the hardness of computing the expected graph moments of a Wigner tensor. In the course of developing the tools for the proof, we will find various useful reinterpretations of the quantities involved in the moments.

Definition B.3. *Let V be a finite set, $p : V \to \mathbb{N}$ a* degree sequence*, and $\mathcal{G}(V, p)$ the set of graphs (with loops and parallel edges allowed) on V such that each $v \in V$ has degree $p(v)$. $G, H \in \mathcal{G}(V, p)$ are related by a* switching *if there is a pair of distinct edges $\{v, w\}, \{v', w'\} \in E(G)$ such that H is formed by replacing both of these edges with the edges $\{v, v'\}, \{w, w'\}$. We view $\mathcal{G}(V, p)$ as a graph, where $G \sim H$ if G and H are related by a switching.*

The switching distance *between G and H, denoted $d_{\mathrm{switch}}(G, H)$, is the distance between G and H in $\mathcal{G}(V, p)$, or equivalently the minimum number of switchings required to reach H from G.*

We note that the degree sequence is preserved by the switching operation, so we must restrict our attention to a fixed degree sequence for the switching distance to be defined. Even so, it is not obvious that the switching distance is finite, but this is indeed the case. Also, we emphasize that, unlike with the sets $\mathcal{G}_{d,p}$ in the main text, we are working with *labelled* graphs here.

Proposition B.4 (Lemma 1 of [Hak63])**.** *Whenever $\mathcal{G}(V, p)$ is non-empty, then it is connected. Thus, $d_{\mathrm{switch}}(G, H) < \infty$ for any $G, H \in \mathcal{G}(V, p)$.*

Remark B.5 (Loops and parallel edges)**.** *We will rely on the results of the works [Hak63, Wil99, BM18], which variously work with not $\mathcal{G}(V, p)$ but the induced subgraph on the vertex set consisting either of simple graphs or loopless multigraphs (with parallel edges allowed). However, implicit in their results are the following facts:*

1. *The induced subgraphs of $\mathcal{G}(V, p)$ on both simple graphs and on loopless multigraphs are connected.*
2. *There is a path between simple graphs (respectively, loopless multigraphs) $G, H \in \mathcal{G}(V, p)$ of minimum length passing through only simple graphs (respectively, loopless multigraphs); that is, the distance between G and H in the induced subgraph on simple graphs (respectively, loopless multigraphs) equals the distance in $\mathcal{G}(V, p)$.*

Thus we will rephrase their results over $\mathcal{G}(V, p)$ without further comment.

Moreover, we may always reduce switching distance computations from a sequence of possibly large degrees to the case $p \equiv 1$, which corresponds to sets of *perfect matchings*.

Definition B.6. *For $G \in \mathcal{G}(V, p)$, let V' be a set of size $\sum_{v \in V} p(v)$, which we view as consisting of $v'_{a,b}$ for $a \in V$ and $b \in [p(a)]$. We say that a perfect matching $\mu \in \mathcal{G}(V', 1)$* realizes *$G$ if G is obtained from identifying $v'_{a,1}, \ldots, v'_{a,p(a)}$ into a single vertex for each a.*

In other words, μ realizes G if G would be obtained from μ under the configuration model; this is in somewhat different language the same definition used in the main text in constructing the graph Weingarten function.

Proposition B.7. $d_{\text{switch}}(G,H) = \min\{d_{\text{switch}}(\mu_G,\mu_H) : \mu_G \text{ realizes } G, \mu_H \text{ realizes } H\}$.

Definition B.8. *For G and H defined on the same vertex set, we write $G\triangle H$ for the graph containing the symmetric difference of the edge sets G and H, where if G has a edges between i and j and H has b edges between i and j, then $G\triangle H$ has $|a-b|$ edges between i and j. We view this graph as being colored, where, in the above setting, if $a > b$ then the edges between i and j are blue, and otherwise they are red.*

A symmetric circuit *in such a colored graph is a closed walk of even length alternating between red and blue edges. A* symmetric circuit partition *is a partition of the edges into symmetric circuits. We denote the number of circuits in the largest symmetric circuit partition by* $\text{circ}(G,H)$.

The following useful fact shows that the (minimum) switching distance is essentially equivalent to the maximum circuit partition.

Proposition B.9 (Theorem 2.5 of [Wil99], Theorem 22 of [BM18])**.** $d_{\text{switch}}(G,H) = \frac{1}{2}|E(G\triangle H)| - \text{circ}(G,H)$.

The proofs cited above treat simple and loopless graphs, respectively, but the same argument extends straightforwardly to graphs with loops as well. One way to see this is to use Proposition B.7 to rephrase the computation of $d_{\text{switch}}(G,H)$ as a similar computation of matchings realizing G and H, which are simple graphs and to which the results of [Wil99] apply.

Corollary B.10. $d_{\text{switch}}(G,H) \le |E(G)| = |E(H)|$.

Proof. The result follows since $|E(G\triangle H)| = 2|E(G-H)| \le 2|E(G)|$. □

We now move towards relating the switching distance to the quantities involved in the Wigner moments; in particular, we will relate it to the exponent $c_{\max}(G)$, which we recall satisfied:

$$\mathop{\mathbb{E}}_{W\sim\text{Wig}} m_G(W) \sim n^{c_{\max}(G)}. \tag{109}$$

Definition B.11. *For p a constant and $|V|$ even, denote by $\mathcal{F} = \mathcal{F}(V,p) \subset \mathcal{G}(V,p)$ the subset of multigraphs that consist of a disjoint union of Frobenii of degree p.*

Our first main result is that the exponent $c_{\max}(G)$ may be computed through the minimum switching distance to a set of Frobenii.

Lemma B.12. *For any p-regular G, $c_{\max}(G) = |E(G)| - d_{\text{switch}}(G,\mathcal{F})$.*

We note that, while $\mathcal{F}$ and $\mathcal{G}$ refer to sets of labelled graphs, the quantities being computed here do not depend on the labelling, since $\mathcal{F}$ is invariant under permutations of the vertex labels.

Proof. Let $v = |V(G)|$, identify $V(G)$ with $[v]$, and let μ be a perfect matching realizing G. View the vertex set of μ as $[v]\times[p]$, so that $(v,1),\dots,(v,p)$ are the "expanded" vertex set corresponding to $v \in V$. To any even edge coloring of G we may associate a perfect matching κ of $[v]$ and perfect matchings $\eta_{\{i,j\}}$ of $[p]$ for each $\{i,j\} \in \kappa$, so that vertices matched in κ have the same edge colors in their neighborhoods, and $\eta_{\{i,j\}}$ is a matching between half-edges of the same color incident with i and with j. We may view $\mu' := \bigsqcup_{\{i,j\}\in\kappa} \eta_{\{i,j\}}$ as a matching on the same vertex set as μ. The number of colors in the given edge coloring of G is at most the number of cycles into which $\mu \sqcup \mu'$ decomposes; conversely, there is an even edge coloring of G with precisely this number of edge

colors, formed by assigning a different color to the edges of G corresponding to the edges of μ lying in each cycle.

But, the possible μ' described above are precisely the matchings that realize a graph of (p-regular) Frobenii on $[v]$. Thus $c_{\max}(G)$ is equivalently the maximum number of cycles into which $\mu \sqcup \mu'$ decomposes for any μ' realizing a graph of Frobenii on $[v]$; moreover, the same holds for any μ realizing G. Let us write $\mathrm{cyc}(\mu \sqcup \mu')$ for this quantity. We have shown so far that

$$c_{\max}(G) = \max\{\mathrm{cyc}(\mu \sqcup \mu') : \mu \text{ realizes } G, \mu' \text{ realizes some } F \in \mathcal{F}\}. \tag{110}$$

Now, separating the 2-cycles in $\mu \sqcup \mu'$, which correspond to edges shared between μ and μ', from longer cycles and using Proposition B.9, we have

$$\begin{aligned}
\mathrm{cyc}(\mu \sqcup \mu') &= \mathrm{cyc}(\mu \triangle \mu') + |\mu \cap \mu'| \\
&= \mathrm{circ}(M, M') + \left(|E(\mu)| - \frac{1}{2}|E(\mu \triangle \mu')|\right) \\
&= |E(G)| - \left(\frac{1}{2}|E(\mu \triangle \mu')| - \mathrm{circ}(\mu, \mu')\right) \\
&= |E(G)| - d_{\mathrm{switch}}(\mu, \mu').
\end{aligned}$$

Finally, substituting, we find

$$\begin{aligned}
c_{\max}(G) &= |E(G)| - \min\{d_{\mathrm{switch}}(\mu, \mu') : \mu \text{ realizes } G, \mu' \text{ realizes some } F \in \mathcal{F}\} \\
&= |E(G)| - d_{\mathrm{switch}}(G, \mathcal{F}),
\end{aligned}$$

completing the proof. □

We also learn some interesting structural facts about maximum even edge colorings from this proof, using the interpretation of $c_{\max}(G)$ in (110).

Proposition B.13. *In any maximum even edge coloring, for any vertex, all edges adjacent to that vertex have distinct colors.*

Proof. Suppose otherwise. By (110), we then have a matching μ of $[d] \times [p]$ realizing G and another matching μ' realizing a disjoint union of Frobenii. Suppose without loss of generality that there are two edges incident with vertex 1 and having the same color, i.e., two edges in μ, touching vertices among $(1,1), \ldots, (1,p)$, and belonging to the same cycle in $\mu \sqcup \mu'$. Suppose again without loss of generality that these latter vertices are $(1,1)$ and $(1,2)$. Under μ', all vertices with first coordinate 1 are matched to vertices with some other first coordinate i. So, suppose $(1,1)$ is matched with (i,j) and $(1,2)$ with (i,k). Define μ'' by instead matching $(1,1)$ with (i,k) and $(1,2)$ with (i,j). Then, the cycle containing these edges in $\mu \sqcup \mu'$ is broken into two cycles, while all other cycles are unchanged. Thus, the original coloring must not have had the maximum number of colors, and we reach a contradiction. □

Proposition B.14. *In any maximum even edge coloring, every colored neighborhood occurs exactly twice.*

Proof. A similar argument applies in this case as well. In the matching interpretation, if there is a colored neighborhood occurring four (or more) times in a coloring, then the same collection of p cycles in $\mu \sqcup \mu'$ must pass through the vertices $(i_a, 1), \ldots, (i_a, p)$ for $a = 1, 2, 3, 4$ and some choice of $i_1, i_2, i_3, i_4 \in [d]$. Suppose without loss of generality that the (i_1, j) are matched to the (i_2, j) and the (i_3, j) to the (i_4, j) in μ' (if the cycles only pass through four of the same collections of vertices such a pairing must occur; if there are more such collections then one may find such pairs by following the matching of μ'. Then, it is possible to instead match either the (i_1, j) to the (i_3, j) or the (i_4, j) such that one of the p cycles is broken into two (by matching its endpoints differently), and no two of the other cycles are merged (by matching their endpoints to one another arbitrarily without joining two distinct cycles). Again, the total number of cycles must increase by at least 1, contradicting the maximality of the original choice of μ'. □

As a corollary of this result we learn that $w_{\max}(G)$, the sum of the "weights" of even edge colorings with $c_{\max}(G)$ colors as defined in the main text, is actually just a counting problem of the number of non-isomorphic even edge colorings with this number of colors; the weights necessarily always equal 1.

Finally, to prove Theorem 5.3, we will use the connection between $c_{\max}$ and the switching distance that we have developed as well as the following result.

Theorem B.15 (Theorem 3.2 of [Wil99])**.** *It is NP-hard to decide whether $d_{\text{switch}}(G, H) \geq d$ given $d \geq 0$ and simple graphs G, H on the same vertex set and having the same degree sequence.*

Remark B.16. *It may be tempting to try to use the cumulants we have defined to compute m_H and its Wigner expectation, and thereby to compute or estimate $c_{\max}$ and so a switching distance. The issue with such a strategy appears to be that computing with the relationship between the cumulants and the graph moments requires forming the graph Weingarten function $\text{Wg}_{G,H}$, which involves a summation over the matchings realizing G and H. There are exponentially many of these matchings, and indeed enumerating them would allow one to compute the switching distance by brute force.*

Proof of Theorem 5.3. We will show that if it is possible to compute $c_{\max}(G)$, then it is also possible to compute switching distances between simple graphs. Namely, let G, H be simple graphs on a shared vertex set V and with a shared degree sequence $p : V \to \mathbb{N}$. We will show that computing $d_{\text{switch}}(G, H)$ may be encoded in the computation of $c_{\max}(J)$ for a certain graph $J = J(G, H)$. By Proposition B.12, this is equivalent to computing $d_{\text{switch}}(J, \mathcal{F})$. Let $v = |V|$, identify V with $[v]$, and set $e = \frac{1}{2}\sum_{v \in V} p(v) = |E(G)| = |E(H)|$.

We form J on the vertex set $\{1, \ldots, v, 1', \ldots, v'\}$. For $i, j \in \{1, \ldots, v\}$, we draw one edge between i and j in J if $i \sim j$ in G. Likewise, for $i', j' \in \{1', \ldots, v'\}$, we draw one edge between i' and j' in J if $i \sim j$ in H. Finally, for each $i \in [v]$, we draw $3e - p(i)$ edges between i and i' in J. Thus, J is a disjoint union of one copy of G and one copy of H, along with a "very heavy matching" with many repeated edges between corresponding vertices in G and H under their joint labelling. Note also that J is $3e$-regular. The basic idea is that this heavy matching will force the nearest disjoint union of Frobenii to correspond to the matching of vertices with the same labels in G and H.

More formally, let F_0 be the graph of Frobenii on the heavy matching in J. We claim that $d_{\text{switch}}(J, \mathcal{F}) = d_{\text{switch}}(J, F_0)$; that is, that F_0 is a minimizer of the switching distance of J to any disjoint union of Frobenii. Indeed, we have $d_{\text{switch}}(J, F_0) \leq d_{\text{switch}}(G, H) + e \leq 2e$, because to reach F_0 we may first transform G into (a copy of) H by switchings, and then align pairs of corresponding

edges in the two copies of H with F_0 one at a time. (We also use the inequality $d_{\text{switch}}(G,H) \le e$ from Corollary B.10.) On the other hand, for any $F \in \mathcal{F} \setminus \{F_0\}$, to reach F from J we must change at least the $3e - p(i)$ edges between i and i' for some $i \in v$. Since $3e - p(i) \ge 3e - e \ge 2e$, we have $d_{\text{switch}}(J,F) \ge 2e$, proving the claim.

Finally, we claim that the inequality mentioned above is tight, that is, that

$$d_{\text{switch}}(J,\mathcal{F}) = d_{\text{switch}}(J,F_0) = d_{\text{switch}}(G,H) + e. \tag{111}$$

After showing this the proof will be complete.

By Proposition B.9, we have $d_{\text{switch}}(J,F_0) = \frac{1}{2}|E(J\triangle F_0)| - \text{circ}(J,F_0)$. Let us view the edges in $J\triangle F_0$ coming from J as "red" and those coming from F_0 as "blue." Then, by our construction of J, we have that $J\triangle F_0$ consists of two disjoint red copies of G and H, together with blue matchings between corresponding vertices in G and H, where the edge between i and i' is repeated $p(i)$ times. In particular, we have $|E(J\triangle F_0)| = 2e + \sum_{i=1}^{v} p(i) = 4e$, so $d_{\text{switch}}(J,F_0) = 2e - \text{circ}(J,F_0) = e + (e - \text{circ}(J,F_0))$. Thus it suffices to show that $d_{\text{switch}}(G,H) = e - \text{circ}(J,F_0)$. Further, since $d_{\text{switch}}(G,H) = |E(G-H)| - \text{circ}(G,H) = e - |E(G\cap H)| - \text{circ}(G,H)$, it also suffices to show $\text{circ}(J,F_0) = |E(G\cap H)| + \text{circ}(G,H)$.

Given a symmetric circuit partition of $G\triangle H$, we may produce one of $J\triangle F_0$ by traversing one blue edge (between G and H) between every edge of the given partition, and also adding circuits of length 4 including each pair of edges shared between G and H. This shows $\text{circ}(J,F_0) \ge |E(G\cap H)| + \text{circ}(G,H)$.

Conversely, we claim that, given a symmetric circuit partition of $J\triangle F_0$, there is another partition of at least the same size that contains circuits of length 4 including each pair of edges shared between G and H. Suppose $\{i,j\}$ is such an edge. Then, edges $\{i,j\}$ and $\{i',j'\}$ must belong to different circuits C, C' of the partition of $J\triangle F_0$. In C, $\{i,j\}$ must be surrounded by two blue edges m_1, m_2 of the heavy matching, and likewise in C', $\{i',j'\}$ must be surrounded by two blue edges m_1', m_2'. Then, we may form another symmetric circuit partition by replacing C, C' with $\{\{i,j\}, m_1, \{i',j'\}m_2\}, (C \setminus \{\{i,j\}, m_1\}) \sqcup (C' \setminus \{\{i',j'\}, m_1'\})$ which is of the same size.

So, there is a maximum symmetric circuit partition of $J\triangle F_0$ that contains all circuits of length 4 including pairs of edges shared between G and H. The remaining circuits in such a partition correspond to a symmetric circuit partition of $G\triangle H$. Thus we have the opposite inequality $\text{circ}(J,F_0) \le |E(G\cap H)| + \text{circ}(G,H)$ as well, and the proof is complete. □

Despite this hardness result, it is still possible to give some tractable and general bounds on the exponent $c_{\max}(G)$.

Proposition B.17. *For any simple p-regular G on d vertices,*

$$c_{\max}(G) \le \frac{(p+1)d}{4}. \tag{112}$$

Equality holds if and only if G has a transitive perfect matching *μ: a matching such that, whenever $\{v,v'\}, \{w,w'\} \in \mu$ and $v \sim w$, then $v' \sim w'$ as well.*

It is an interesting simplification of the problem of computing $c_{\max}(G)$ to check whether the condition for this bound being saturated holds or not. When $p = 3$, one may check that a transitive perfect matching exists if and only if each connected component of G is isomorphic to one of two graphs built from a union of two cycles: if we imagine forming a cubical complex from this graph

by adding a 2-cell between the "lateral" edges connecting edges of the transitive perfect matching, then the resulting surface must either be a cylinder without its top and bottom or the same with a "twist," i.e., a Möbius band. Once $p \geq 4$, such a simple classification seems elusive, and it is unclear whether to expect the problem to be easy or hard.

C Conditioning of Weingarten Matrices

We have discussed in Section 3.3 the structure of the projection operator $\Pi = \Pi_\ell = \mathbb{E}_Q Q^{\otimes \ell}$. As stated there, Π_ℓ projects onto the space spanned by the matching vectors $w(\mu)$ defined in (16) where μ is a perfect matching of $[\ell]$. Thus the projection may be written

$$\Pi_{\mu,\nu} = \sum_{\mu,\nu} (M^{-1})_{\mu,\nu}\, w(\mu) \otimes w(\nu), \tag{113}$$

where M^{-1} is the inverse of the Gram matrix

$$M_{\mu,\nu} = \langle w(\mu), w(\nu) \rangle ,$$

or the Moore–Penrose pseudoinverse if the $w(\mu)$ are overcomplete and M is not of full rank. For matching vectors, the Gram matrix is given by

$$M_{\mu,\nu} = \langle w_\mu, w_\nu \rangle = n^{\#\text{ of cycles in } \mu \sqcup \nu} = n^{\ell/2 - \Delta(\mu,\nu)} . \tag{114}$$

Here by $\mu \sqcup \nu$ we mean the 2-regular multigraph with ℓ vertices whose edges are the pairs in μ and ν, and each of these cycles gives a factor of n. We define $\Delta(\mu, \nu)$ as the minimum number of swaps it takes to transform μ to ν, where a swap changes two pairs in a matching from $\{(a,b),(c,d)\}$ to $\{(a,c),(b,d)\}$ or $\{(a,d),(b,c)\}$. The relation

$$\#\text{ of cycles in } \mu \sqcup \nu = \ell/2 - \Delta(\mu, \nu)$$

follows because $\mu \sqcup \nu$ has $\ell/2$ cycles if and only if $\mu = \nu$, and each swap on a shortest path from μ to ν merges two cycles into one. See Figure 12 for an example.

The inverse of M is called the *Weingarten function* and is denoted Wg. Thus

$$\Pi = \sum_{\mu,\nu} \mathrm{Wg}_{\mu,\nu}\, w(\mu) \otimes w(\nu) ,$$

where the sum is over all perfect matchings μ, ν of $[\ell]$. By permutation symmetry, $\mathrm{Wg}_{\mu,\nu} = (M^{-1})_{\mu,\nu}$ is a function only of the cycle structure of $\mu \sqcup \nu$, i.e., the number of cycles of each length that this graph has. For this reason Wg is often called a *function*, but for our purposes we will view it as a matrix and seek to understand its spectrum, or equivalently the spectrum of M.

For sufficiently large n, M is dominated by its diagonal, and a representation-theoretic argument [Bra37, Wen88] shows that M has full rank whenever $n \geq \ell/2$. However, we need the stronger property that its smallest eigenvalue is bounded above zero. Using a simple counting argument we will show that this holds, and moreover that the matrix is well-conditioned, whenever $n > (1+\epsilon)\ell^2$.

Proposition C.1. *Suppose that $n > \ell^2$. Then the Gram matrix M defined in* (114) *has full rank, and all of its eigenvalues lie in the interval $[n^{\ell/2}(1 - \ell^2/n), n^{\ell/2}(1 + \ell^2/n)]$.*

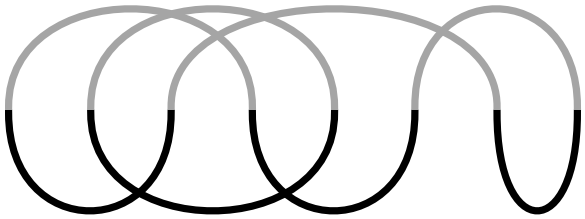


Figure 12: The inner product of two matching vectors $w(\mu), w(\nu)$ as defined in (16) is n raised to the number of cycles in their disjoint union $\mu \sqcup \nu$. In this example, $\ell = 8$ and there are two cycles in $\mu \sqcup \nu$, one of length 2 and one of length 6. Thus $\langle w(\mu), w(\nu) \rangle = n^2$ and μ and ν are $\Delta(\mu, \nu) = 2$ swaps apart.

Proof. Since there are only $\binom{\ell}{2}$ possible swaps, for any matching μ there are at most $\binom{\ell}{2}^t \leq (\ell^2/2)^t$ matchings ν such that $\Delta(\mu, \nu) = t$. Therefore, if we write

$$M = n^{\ell/2}(\mathbb{1} + H) \quad \text{where} \quad H_{\mu\nu} = \begin{cases} 0 & \mu = \nu \\ n^{-\Delta(\mu,\nu)} & \mu \neq \nu \,, \end{cases}$$

then if $n > \ell^2$ the sum of any row H_μ of H gives a bound on the operator norm of H,

$$|H| \leq \sum_{\nu \neq \mu} H_{\mu\nu} \leq \sum_{t=1}^{\infty} \left(\frac{\ell^2}{2n} \right)^t \leq \ell^2/n \,.$$

The Geršgorin circle theorem [Ger31] then implies that M has full rank whenever $n > \ell^2$ and that its eigenvalues lie in the stated interval. □

Since the eigenvalues of Wg are the reciprocals of those of M, Proposition C.1 implies that Wg is well-conditioned.

Corollary C.2. *Suppose that $n \geq 2\ell^2$. Then, all of the eigenvalues of $(\mathrm{Wg}_{\mu,\nu})$ lie in the interval $[\frac{1}{2} n^{-\ell/2}, 2n^{-\ell/2}]$.*

Remark C.3. *Proposition C.1 and Corollary C.2 are essentially tight, since when $n = o(\ell^2)$ then the largest eigenvalue of* Wg *is roughly $n^{-\ell/2}\, \mathrm{e}^{\ell^2/2n}$. We outline the argument, which we learned of from a personal communication with Piotr Śniady. Per Proposition 5 of [ZJ09] (originally due to [Col03]), the distinct eigenvalues of* Wg *correspond to Young diagrams λ of ℓ boxes with an even number of boxes in each row.*[8] *We write $\lambda/2$ for the same diagram where the number of boxes in each row is halved. We view these diagrams as sets of (i, j), the "coordinates" of the boxes of the diagram, with $i, j \geq 0$ so that the box in the top left corner has coordinates $(0, 0)$, the one below coordinates $(1, 0)$, the one to the right coordinates $(0, 1)$, and so forth. Then, the eigenvalue of* Wg *associated to λ is $1/(\prod_{(i,j)\in\lambda/2}(n + j - i))$. The largest eigenvalue of* Wg *is then the one corresponding to λ two columns of $\ell/2$ boxes each, which gives*

$$\frac{1}{\prod_{i=0}^{\ell/2-1}(n-i)} = n^{-\ell/2} \prod_{i=0}^{\ell/2-1} \frac{1}{1 - \frac{i}{n}} \approx n^{-\ell/2} \exp\left(\frac{\ell^2}{2n} \right), \tag{115}$$

[8] Actually, this correspondence could give another more explicit treatment of the conditioning results derived above, but we prefer to give a self-contained treatment using the simpler Geršgorin circle theorem.

as claimed.

In the main text we mostly encounter not this Weingarten function over matchings, but the "graph Weingarten function" that we define indexed by $\mathcal{G}_{d,p}$ (Definition 4.16). This has entries

$$\mathrm{Wg}_{G,H} = \sum_{\substack{\mu \text{ realizes } G \\ \nu \text{ realizes } H}} \mathrm{Wg}_{\mu,\nu} \tag{116}$$

for $G, H \in \mathcal{G}_{d,p}$. We also recall from Proposition 4.17 that the number of summands over each axis is

$$\#\{\mu \text{ realizes } G\} = \frac{p!^d d!}{|\mathrm{eAut}(G)|}. \tag{117}$$

Using this, we state a corollary on the spectrum of this compressed version of Wg in the form that will be useful in the main text.

Corollary C.4. *Suppose that $pd \leq \sqrt{n/2}$ is even. Then the symmetric matrix indexed by $G, H \in \mathcal{G}_{d,p}$ with entries*

$$n^{pd/2} \cdot \frac{\sqrt{|\mathrm{eAut}(G)| \cdot |\mathrm{eAut}(H)|}}{p!^d d!} \cdot \mathrm{Wg}_{G,H} \tag{118}$$

has all its eigenvalues lying in the interval $[\frac{1}{2}, 2]$.

Proof. Write $\ell := pd$. Define a matrix $\widetilde{J}$ indexed by $G \in \mathcal{G}_{d,p}$ and μ perfect matchings of $[pd]$ with

$$\widetilde{J}_{G,\mu} := \sqrt{\frac{|\mathrm{eAut}(G)|}{p!^d d!}} \mathbf{1}\{\mu \text{ realizes } G\} = \frac{1}{\sqrt{\#\{\mu \text{ realizes } G\}}} \mathbf{1}\{\mu \text{ realizes } G\}. \tag{119}$$

By construction, we have $\widetilde{J}\,\widetilde{J}^\top = \mathbb{1}$. On the other hand, calling X the matrix in the claim, we have

$$X = \widetilde{J}(n^{\ell/2}\mathrm{Wg})\widetilde{J}^\top. \tag{120}$$

By Corollary C.2, we then have

$$X \preceq 2\widetilde{J}\widetilde{J}^\top = 2\mathbb{1}, \tag{121}$$

and the lower bound follows similarly. □

By an identical proof and using Proposition 4.29, we also have the following variation for the Weingarten matrix appearing for open multigraphs.

Corollary C.5. *Suppose that $pd - 1 \leq \sqrt{n/2}$ is even. Then the symmetric matrix indexed by $G, H \in \mathcal{G}_{d,p\to}$ with entries*

$$n^{(pd-1)/2} \cdot \frac{\sqrt{|\mathrm{eAut}(\mathrm{chop}(G))| \cdot |\mathrm{eAut}(\mathrm{chop}(H))|}}{p!^{d-1}(p-1)!(d-1)!} \cdot \mathrm{Wg}_{\mathrm{chop}(G),\mathrm{chop}(H)} \tag{122}$$

has all its eigenvalues lying in the interval $[\frac{1}{2}, 2]$.

As an aside, let us mention that computing Wg explicitly is a notoriously difficult problem. We can get a sense of how its entries scale by writing it as a geometric series,

$$\mathrm{Wg} = M^{-1} = n^{-\ell/2}(\mathbb{1} + H)^{-1} = n^{-\ell/2} \sum_{t=0}^{\infty} (-1)^t H^t \,, \tag{123}$$

and expand this into a sum over paths through the space of perfect matchings. For a given pair of matchings μ, ν, let $\{\mu \sim \nu\}$ denote the set of paths σ where $\mu = \sigma_0 \neq \sigma_1 \neq \cdots \neq \sigma_t = \nu$ for some $t \geq 0$, and write $|\sigma| = t$. Define $\Delta(\sigma)$ as σ's total length in swap distance,

$$\Delta(\sigma) = \sum_{i=1}^{t} \Delta(\sigma_i, \sigma_{i-1}) \,,$$

and define $g(\sigma)$ as σ's *geodesic defect*, i.e., the difference between its length and the shortest-path distance between its endpoints,

$$g(\sigma) = \Delta(\sigma) - \Delta(\mu, \nu) \geq 0 \,.$$

Then (following [CŚ06, Ban10]) we have

$$\mathrm{Wg}_{\mu\nu} = n^{-\ell/2} \sum_{\sigma \in \{\mu \sim \nu\}} (-1)^{|\sigma|} \, n^{-\Delta(\sigma)} \tag{124}$$

$$= n^{-\ell/2 - \Delta(\mu,\nu)} \sum_{\sigma \in \{\mu \sim \nu\}} (-1)^{|\sigma|} \, n^{-g(\sigma)} \,. \tag{125}$$

Thus for graphs of constant size we have $\mathrm{Wg}_{\mu\nu} = O(n^{-\ell/2 - \Delta(\mu,\nu)})$. The prefactor for a given $\mathrm{Wg}_{\mu\nu}$, i.e., the sum of signed geodesics from μ to ν, might grow rapidly with ℓ.

For instance, for $D = 4$ where there are 3 perfect matchings, the rows of M and Wg where $\mu = \{(1,2),(3,4)\}$ are

ν	$\{(1,2),(3,4)\}$	$\{(1,3),(2,4)\}$	$\{(1,4),(2,3)\}$
$M_{\mu,\nu}$	n^2	n	n
$\mathrm{Wg}_{\mu,\nu}$	$\frac{n+1}{n(n-1)(n+2)}$	$\frac{-1}{n(n-1)(n+2)}$	$\frac{-1}{n(n-1)(n+2)}$

D Combinatorial Bounds

We gather some auxiliary combinatorial results that are used in the main text.

Proposition D.1. *Suppose $k \leq n/2$. Then, $n^k \exp(-k^2/n) \leq n^{\underline{k}} \leq n^k$.*

Proof. The upper bound is immediate. For the lower bound, we have:

$$\begin{aligned} n^{\underline{k}} &= n^k \cdot 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{k-1}{n}\right) \\ &= n^k \exp\left(\sum_{i=0}^{k-1} \log\left(1 - \frac{i}{n}\right)\right) \end{aligned}$$

and here, noting that $\log(1-x) \geq -2x$ for all $0 \leq x \leq \frac{1}{2}$, we have

$$\begin{aligned} &\geq n^k \exp\left(\frac{2}{n}\sum_{i=0}^{k-1} i\right) \\ &\geq n^k \exp\left(-\frac{k^2}{n}\right), \end{aligned}$$

as claimed. □

Proposition D.2. *A* chain *of subsets of* $[n]$ *is a sequence of strict inclusions* $\varnothing \subsetneq A_1 \subsetneq A_2 \subsetneq \cdots \subsetneq A_k \subsetneq [n]$. *The number of chains of subsets of* $[n]$ *is at most* $3^n n!$.

Proof. Call $C(n)$ the number of chains on $[n]$. The chains are in bijection with the ordered set partitions of $[n]$, where a chain as defined in the statement corresponds to the partition $(A_1, A_2 \setminus A_1, \cdots, [n] \setminus A_k)$. Every ordered set partition consists recursively of a choice of a non-empty set and an ordered set partition on the remaining elements. Thus, $C(0) = 1$ and for $n \geq 1$ we have

$$C(n) = \sum_{i=1}^{n} \binom{n}{i} C(n-i) = nC(n-1) + \sum_{i=2}^{n} \binom{n}{i} C(n-i). \tag{126}$$

We proceed by strong induction. Clearly the claim holds for $n = 0$. Suppose that the claim holds for all $C(m)$ with $0 \leq m \leq n-1$. Then,

$$\begin{aligned} C(n) &\leq 3^{n-1} n! + \sum_{i=2}^{n} \binom{n}{i} 3^{n-i}(n-i)! 3^{n-i} \\ &= 3^n n! \left(\frac{1}{3} + \sum_{i=2}^{n} \frac{3^{-i}}{i!}\right) \\ &\leq 3^n n! \left(\frac{1}{3} + \exp\left(\frac{1}{3}\right) - 1\right) \\ &< 3^n n!, \end{aligned}$$

completing the proof. □

Proposition D.3 (Section 2 of [HR18])**.** *An* integer partition *of* $d \geq 1$ *is a sequence of* $1 \leq a_1 \leq \cdots \leq a_\ell$ *such that* $a_1 + \cdots + a_\ell = d$. *There is an absolute constant* $C > 0$ *so that the number of integer partitions of* d *is at most* $\frac{1}{d}\exp(C\sqrt{d})$ *for all* $d \geq 1$.