Second-order finite free probability

Curran McConnell

A thesis submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Arts

Graduate program in Mathematics
York University
Toronto, Ontario

December 2023

Abstract

Finite free probability is a new field lying at the intersection of random matrix theory and non-commutative probability. It is called “finite” because unlike traditional free probability, which takes the perspective of operators on infinite-dimensional vector spaces, finite free probability focuses on the study of $d \times d$ matrices. Both fields study the behaviour of the eigenvalues of random linear transformations under addition. Finite free probability seeks in particular to characterize random matrices in terms of their (random) characteristic polynomials. I studied the covariance between the coefficients of these polynomials, in order to deepen our knowledge of how random characteristic polynomials fluctuate about their expected values. Focusing on a special case related to random unitary matrices, I applied the representation theory of the unitary group to derive a combinatorial summation expression for the covariance.

Acknowledgements

I would like to thank Nantel Bergeron, Kevin McGregor and Paul Skoufranis for their time and effort in offering edits and corrections to this thesis. Their feedback considerably improved the final result. Any remaining errors are my own.

I would like to thank Nantel Bergeron for his mentorship, particularly in the early phases of the project. His advice about how to choose good research questions was invaluable. Without it, I would have spent a long time going nowhere fast.

I also thank Benjamin Mirabelli. One of the open problems which he pointed in his PhD thesis [32] became the core research question for project.

I would like to thank the participants of the Algebraic Combinatorics seminar at the Fields Institute for Research in Mathematical Sciences for creating such a lively environment for research. This group of people includes Nantel Bergeron, Kelvin Chan, Lucas Gagnon, Félix Gelinas, Yohana Solomon, Farhad Soltani, Nancy Wallace, Mike Zabrocki and others. Representation theory is a very important topic for the work that this community does, and it was through conversations with members of this group that I first learned about many of the representation-theoretic techniques I applied in this thesis.

Table of Contents

Chapter 1

Introduction

1.1 The plan of this document

In this document, I will develop a formula for some "second-order" quantities in a setting where the "first-order" quantities have already been well-described. The quantities in question are the "moments" of a certain probability distribution, a concept which will be covered in detail in chapter 4. The distribution in question is the distribution of the characteristic polynomial of a random matrix that arises in the theory of *finite free probability*.

Finite free probability was originally developed by Marcus, Spielman and Srivastava in 2015 [27], as a development of some tools they used for their solution of the Kadison-Singer problem in 2013 [26]. It is a theory describing the behaviour of random matrices in terms of their random characteristic polynomials. The main thrust of their original paper [27] is to account for the expected value of the characteristic polynomial of certain random matrices. In particular, given a unitary, $d \times d$, Haar-distributed random matrix Q and self-adjoint matrices A and B, they developed a formula for the expected characteristic polynomial

$$\mathbb{E}[\chi(A + QBQ^*)],$$

where $\mathbb{E}$ is the expectation operator taking averages of random variables, and $\chi(A)$ is defined as the characteristic polynomial $\det(Ix - A)$ of a square matrix A. This formula for the expected characteristic polynomial resembled a convolution of the polynomials $\chi(A)$ and $\chi(B)$. This convolution was later related by Marcus [28] to the operation *free convolution of random variables* used in the field of *infinite free probability*, which at that point had existed for about three decades. This other free probability theory, usually just called *free probability* without the qualifier *infinite*, focuses on the eigenvalues of operators in infinite-dimensional W^*-algebras.

In the conclusion of his 2021 Princeton PhD thesis "Hermitian, Non-Hermitian and Multavariate Finite Free Probability", Benjamin Mirabelli posed the question of determining the covariance matrix of this random polynomial, which is the second-order analogue of the expected value of the polynomial. This covariance matrix would contain information about whether the polynomial clusters tightly about its average, or whether it spreads out a lot. The theory of infinite free probability already has a well-developed second-order theory. The long-term goal towards which this research project is a small contribution is to describe the relationship between second-order finite free probability and second-order infinite free probability. This parallels the move Marcus made by relating the convolution of polynomials occurring in finite free probability with the free convolution of random variables in infinite free probability. An essential prerequisite of this programme is a preliminary description of this covariance matrix.

1.2 A litany of algebraic structures

What follow are a large number of definitions we will use throughout the text. These are standard notations for algebraic structures, as may be found in [12] or [18]. Someone with recent experience of a graduate class in algebra will likely be familiar with all of these.

Definition 1.2.1. A monoid G is a set containing an identity element e, together with an operation $\cdot : G \times G \to G$, satisfying the following axioms.

$$\forall g, h, k \in G, (gh)k = g(hk)$$
$$\forall g \in G, eg = ge = g$$

Definition 1.2.2. Given a monoid G, we define the opposite monoid G^{op} as the set G equipped with the product

$$(g, h) \mapsto hg.$$

Definition 1.2.3. A left monoidal G-action of a monoid G on a set X is a function $. : G \times X \to X$ sending $(g, x) \mapsto g.x$, satisfying

$$h.g.x = hg.x$$
$$e.x = x$$

$\forall g, h \in G, x \in X$. We call X in this case a left monoidal G-set.

A right monoidal G-action is a left monoidal G^{op}-action. A right monoidal G-set is a left monoidal G^{op}-set.

There is in principle complete symmetry between left and right actions, but we will take the left to be the default direction of actions.

Definition 1.2.4. A homomorphism of left monoidal G-sets is a function $f : X \to Y$ such that

$$f(g.x) = g.f(x)$$

for all $g \in G, x \in X$.

A homomorphism of right monoidal G-sets is defined the same way.

Definition 1.2.5. A monoid homomorphism is a function $f : G \to H$ satisfying

$$f(gh) = f(g)f(h)$$

$\forall g, h \in G$, where G and H are monoids.

Definition 1.2.6. A group G is a monoid with the following property. For each $g \in G$ there exists some $g^{-1} \in G$ such that

$$gg^{-1} = g^{-1}g = e.$$

In other words, every group element has an inverse.

Definition 1.2.7. A left (respectively, right) G-action $\cdot : G \times X \to X$ is a left (right) monoidal action where G is a group. In this case, we call X a G-set.

Definition 1.2.8. If X is a G-set then we denote by X/G the set of orbits in X, i.e., the collection of subsets of X that are closed under the group action.

Definition 1.2.9. A group homomorphism is a monoid homomorphism $f : G \to H$, where each of G and H are groups.

Definition 1.2.10. An abelian group G is a group with the property that

$$ab = ba$$

for all $a, b \in G$. We may write the group operation of an abelian group as "+" and the identity element as "0".

Definition 1.2.11. A ring $(R, +, \cdot)$ is a set R such that $(R, +)$ forms an abelian group and $(R, \cdot)$ forms a monoid, satisfying the following axioms $\forall r, s, t \in R$.

$$(r + s)t = rt + st$$
$$r(s + t) = rs + rt$$

We may denote the multiplicative identity of a ring by 1.

Definition 1.2.12. A homomorphism of rings is a function that is both a group homomorphism with respect to addition, and a monoid homomorphism with respect to multiplication.

Definition 1.2.13. Given a ring R, we denote by $R^\times$ the group of units of R, i.e., the multiplicative group of those ring elements which possess a multiplicative inverse.

Definition 1.2.14. We call a ring R commutative if

$$ab = ba$$

for all $a, b \in R$.

Definition 1.2.15. A field F is a commutative ring such that $F^\times = F \setminus \{0\}$.

The most important fields for us are the real number line $\mathbb{R}$ and the complex number plane $\mathbb{C}$. The symbol $\mathbb{K}$ will denote a field that is either $\mathbb{R}$ or $\mathbb{C}$.

Definition 1.2.16. Let R be a ring. A left (respectively, right) R-module M is an abelian group equipped with a left (right) monoidal R-action with respect to R's multiplicative structure, satisfying

$$(r+s).m = r.m + s.m$$
$$r.(m+n) = r.m + r.n$$

$\forall r, s \in R, m, n \in M$. We will call the action *scalar multiplication* and suppress the "." notation for monoidal actions.

Definition 1.2.17. A morphism of left (respectively, right) R-modules is a group homomorphism which is also a morphism of left (right) monoidal R-sets. We call a morphism of left R-modules an R-linear transformation.

Definition 1.2.18. We say that an R-module M is finitely-generated if there is a finite subset $\{m_1, m_2, ..., m_n\}$ such that every element $m \in M$ may be written in the form

$$\sum_{i=1}^{n} r_i m_i$$

for some collection of $r_i \in R$.

Definition 1.2.19. We say that a module M is semisimple if it is the direct sum of its submodules. We say that a ring R is semisimple if every R-module is semisimple.

Definition 1.2.20. A vector space V is an F-module, where F is a field. We may also call V an F-vector space.

We say that an F-vector space is finite-dimensional if it is finitely-generated as an F-module.

Definition 1.2.21. An algebra A over a field F is an F-vector space equipped with a multiplicative operation $\cdot : A \times A \to A$, which may or may not be monoidal, or even associative, such that

$$(ra) \cdot (sb) = (rs)(ab)$$
$$a \cdot (b+c) = ab + ac$$
$$(a+b) \cdot c = ac + bc$$

for all $r, s \in F, a, b, c \in A$. We may say in this case that A is an F-algebra. We will typically write both the scalar action of F on A and the multiplication internal to A using juxtaposition, i.e., by writing ra and ab to denote the scalar product and the multiplicative operation, respectively.

1.3 Elementary topological notions

This thesis is concerned not just with algebra, but with certain topological/geometric structures built on top of algebraic ones. In this section, I will list the fundamental topological definitions required throughout the text.

Definition 1.3.1. A topological space is a set X equipped with a set $\mathcal{O}(X)$ of subsets of X, declared to be its open sets that is closed under finite intersections and arbitrary unions, and must contain both X and $\emptyset$. The elements of $\mathcal{O}(X)$ are declared to be the open sets of X. We call the set $\mathcal{O}(X)$ the *topology* of X.

Definition 1.3.2. We say a subset of a topological space is closed if it is the complement of an open set.

Definition 1.3.3. We say a function $f : X \to Y$ between topological spaces is continuous is the preimage of any open set in Y is open in X.

Definition 1.3.4. We say a topological space X is Hausdorff if, for all distinct $x, y \in X$, there exist open sets $U, V \subset X$ such that $x \in U, y \in V$ and $U \cap V = \emptyset$.

Definition 1.3.5. We say a subset K of a topological space X is compact if for every collection of open sets $\{U_i\}_{i\in I}$ such that $K \subseteq \bigcup_{i\in I} U_i$, there is a finite subset $J \subseteq I$ such that $K \subseteq \bigcup_{j\in J} U_j$.

Definition 1.3.6. Let X be a topological space. We say it has the discrete topology if for all $A \subseteq X$, A is open.

Definition 1.3.7. A basis $\mathcal{B}$ of a topological space X is a set of subsets of X such that

$$\mathcal{O}(X) = \left\{ \bigcup_{B\in\Gamma} B \mid \Gamma \subseteq \mathcal{B} \right\}.$$

Thus a topological space may be specified by its basis, rather than by its topology.

1.4 Topological algebraic structures

Here I will combine the algebraic and topological concepts defined previously in the chapter.

A *topological X*, where "X" is an algebraic structure, is just an instance of that algebraic structure that is also a topological space. When an algebraic structure has a finite underlying set, we will implicitly imbue it with a discrete topology.

Observe that the operation in a monoid G is a G-action on G, i.e., that G is a G-set with respect to its own operation. Groups are monoids, and modules are groups, and in general, nearly all of the algebraic structures we have discussed share the property that their data may be given in the form of one or more monoidal G-actions. (The multiplicative operation in an algebra is the sole exception so far, but the examples we will see will in fact often be monoidal.)

When we discuss a *continuous* X A, where "X" is an algebraic structure, we just mean that A is Hausdorff, and that every monoidal action out of which A is composed is a continuous function. For example, a continuous R-module M has the property that the scalar action $R \times M \to M$ is continuous, as is the abelian operation $M \times M \to M$. A continuous group has the property that the group operation $G \times G \to G$ is continuous.

1.5 Metrics, norms and inner products

The definitions in this section produce geometric structures built on top of the topological ones just defined. These concepts undergird the quantitative methods used in this thesis.

The definitions here are standard, and may be found for instance in [34].

Definition 1.5.1. A metric space (X, d) is a set X together with a function

$$d : X \times X \to [0, \infty)$$

with the following properties:

- $d(x, y) = d(y, x)$ for all $x, y \in X$
- $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$, and
- $d(x, y) = 0$ if and only if $x = y$.

A metric space possesses a natural topology given by the basis of ϵ-balls

$$B_\epsilon(x) = \{y \in X \mid d(x, y) < \epsilon\}$$

for all $x \in X, \epsilon > 0$.

Definition 1.5.2. A sequence (s_n) in a metric space X is Cauchy if for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $m, n \geq N$ implies

$$d(s_m, s_n) < \epsilon.$$

Definition 1.5.3. A sequence (s_n) in a metric space X converges to a point s if for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $n \geq N$ implies

$$d(s_n, s) < \epsilon.$$

Definition 1.5.4. A metric space is complete if every Cauchy sequence converges.

Definition 1.5.5. A normed vector space $(V, \|\cdot\|)$ is a vector space V together with a function

$$\|\cdot\| : V \to [0, \infty)$$

with the following properties:

- $\|u+v\| \leq \|u\| + \|v\|$ for all $u, v \in V$,
- $\|\lambda u\| = |\lambda| \|u\|$ for all $\lambda \in \mathbb{K}, u \in V$, and
- $\|u\| = 0$ if and only if $u = 0$.

A normed vector space possesses a natural metric given by

$$d(x, y) = \|y - x\|.$$

Definition 1.5.6. An inner product space $(V, \langle \cdot | \cdot \rangle)$ is a vector space V together with a function

$$\langle \cdot | \cdot \rangle : V \times V \to \mathbb{K}$$

with the following properties:

- $\langle u|v\rangle = \overline{\langle v|u\rangle}$ for all $u, v \in V$.
- $\langle u|v + w\rangle = \langle u|v\rangle + \langle u|w\rangle$ for all $u, v, w \in V$.
- $\langle u|\alpha v + w\rangle = \alpha\langle u|v\rangle + \langle u|w\rangle$ for all $u, v, w \in V, \alpha \in \mathbb{K}$.
- $\langle u|u\rangle \geq 0$ for all $u \in V$. There is strict equality if and only if $u = 0$.

The first condition is equivalent to the requirement that $\langle u|v\rangle = \langle v|u\rangle$ when $\mathbb{K} = \mathbb{R}$. An inner product space possesses a natural norm given by

$$\|v\| = \sqrt{\langle v|v\rangle}.$$

Definition 1.5.7. The standard inner product for a vector space $\mathbb{C}^d$ is given by

$$\langle u|v\rangle = \sum_{i=1}^{d} \overline{u_i} v_i.$$

Definition 1.5.8. A Banach space is a complete normed vector space.

Definition 1.5.9. A Hilbert space is a complete inner product space.

Chapter 2

Linear algebra

In this chapter, I develop some basic facts about linear algebra, including a few lesser-known formulae relating to determinants. The ambient field is always assumed to be the complex numbers if not otherwise stated.

2.1 Core definitions and notations

The proofs in this thesis will involve a decent amount of clerical work to keep track of matrix and vector indices. We will adopt a few conventions to make this easier. Many of these notations could be found in a textbook like [2] or [12]. However, some are less standard, coming from [27].

- We will use the notation $[d] = \{1, 2, ..., d\}$.
- If S is a finite set of integers, then there is a unique order-preserving bijection $\alpha : [|S|] \to S$. We will equivocate in our notation between S and α, using the symbol S to represent both. E.g., we may write $\{2, 4, 5\}(2) = 4$.
- We will regard $S(T)$ to be the set $\{S(t) \mid t \in T\}$, for all $T \subseteq [|S|]$.
- We will regard $\overline{S}(T)$ to be the set $T \setminus S(T)$.
- We will regard a sequence $(s_i)_{i=1}^m$ in the set X to be a function $[m] \to X$.
- If A is a $m \times n$ matrix and $1 \leq i \leq m, 1 \leq j \leq n$, then $A_{i,j}$ is the (i, j)th entry of A.
- We denote the set of $n \times n$ matrices over the field F by $\mathcal{M}_n(F)$.
- We will denote by $A_{\cdot,i}$ the ith column of a matrix, and regard it as a vector.
- We denote the standard basis of a vector space F^n by $\{e_i\}$, where e_i has a 1 for its ith entry and zeroes elsewhere.

- Over any field F, and for all $n \geq 1$, there is a function $\det : (F^n)^n \to F$ known as the determinant. This is the unique alternating n-multilinear function with the property that

$$\det(e_1, e_2, ..., e_n) = 1.$$

We will typically consider det to have the domain $M_n(F)$, and regard it as mapping

$$A \mapsto \det(A_{\cdot,1}, A_{\cdot,2}, ..., A_{\cdot,n}).$$

- Over any field, one can construct the zero vector space $V = \{0\}$ consisting of only the origin. There is precisely one 0×0 matrix representing the unique linear endomorphism $A : V \to V, 0 \mapsto 0$. We declare the determinant of this matrix to equal 1.

- If A is a $m \times n$ matrix and $S \subseteq [m], T \subseteq [n]$, then we denote by $A_{S,T}$ the (S,T)-submatrix of A. We denote by A_S the (S,S)-submatrix of A, and call any such submatrix *principal*.

- We call the determinant of a submatrix of a matrix A a *minor* of A. We call the determinant of a principal submatrix a principal minor of A. We denote by $[A]_{S,T}$ the (S,T)-minor of A, and we denote by $[A]_S$ the principal (S,S)-minor of A.

- The previous conventions imply in particular that $[A]_\emptyset = 1$ for all matrices A.

- We will denote the set of subsets $T \subseteq S$ for which $|T| = k$ by $\binom{S}{k}$.

- If S is a finite set of numbers, we will denote by $\|S\|_1$ the sum of the elements of S. I'm borrowing this nonstandard notation from [27].

- A vector $v \in \mathbb{C}^d$ is implicitly a $d \times 1$ matrix, or column vector. A scalar $z \in \mathbb{C}$ is implicitly a 1×1 matrix.

- For complex matrices A, we will reserve the notation A^* to mean the adjoint of A (conjugate transpose), and we will reserve $\overline{A}$ for the entrywise conjugation of A. Since a scalar $z \in \mathbb{C}$ is a symmetric 1×1 matrix, we have $z^* = \overline{z}$, but we will try to stick to overline notation to better signal when a value is a scalar.

- We will adopt Iverson bracket notation [44]. If P is a statement with a well-defined truth value, then $[P]$ will equal 1 if P is true, and will equal 0 if P is false. For example, $[1 = 0] = 0$.

- We define Dirac delta notation in terms of the Iverson bracket, such that $\delta_{i,j} = [i = j]$.

- We will denote the kernel of a linear map A by $\ker A$. We will denote the image by $\operatorname{im} A$.

- We will say that a square matrix is diagonal if all of its non-zero entries $a_{i,j}$ have the property that $i = j$.
- We will say that a square matrix is upper-triangular if all of its non-zero entries $a_{i,j}$ have the property that $i \leq j$.

2.2 Unitary matrices

Unitary matrices are a generalization of the complex unit circle to higher dimensions. This thesis is going to be concerned with *random* unitary matrices later on, for understanding which the fundamentals of concrete, non-random unitary matrices will be important.

In this section, we will assume V is an inner product space.

Proposition 2.2.1 ([2])**.** If $A \in \operatorname{End} V$, then there is some $A^* \in \operatorname{End} V$ for which

$$\langle u, Av \rangle = \langle A^* u, v \rangle$$

for all $u, v \in V$.

Definition 2.2.2. A matrix U is unitary if $UU^* = I$.

Proposition 2.2.3 ([2])**.** Let $U : V \to V$ be a matrix. The following are equivalent.

1. U is unitary.
2. $U^*U = I$.
3. The columns of U, seen as vectors, are an orthonormal basis for V.
4. The rows of U, seen as vectors, are an orthonormal basis for V.
5. $\langle v|w \rangle = \langle Uv|Uw \rangle$ for all $v, w \in V$.

2.3 Self-adjoints, idempotents and orthogonal projections

The algebraic concepts outlined in this section are fundamentally linked with the geometry of (complex) Euclidean space, which is why they will be critical to many of the arguments later featured in this thesis.

Definition 2.3.1. A matrix A is self-adjoint if $A^* = A$.

Proposition 2.3.2 ([2])**.** A matrix T is self-adjoint if and only if for all $u, v \in V$, we have the identity

$$\langle u|Tv \rangle = \langle Tu|v \rangle.$$

Definition 2.3.3. We call a square matrix P idempotent if $P^2 = P$.

Definition 2.3.4. We call a square matrix P an orthogonal projection if P is idempotent and $\ker P = (\operatorname{im} P)^{\perp}$.

Proposition 2.3.5 ([14])**.** A selfadjoint idempotent P is an orthogonal projection.

Proposition 2.3.6 ([2])**.** If P and Q are orthogonal projections onto the same column space then $P = Q$.

Proof. The columns of an orthogonal projection matrix onto a subspace W consist of the projections of the standard basis vectors onto W. $\square$

Definition 2.3.7. We call a matrix T normal if $TT^* = T^*T$.

Proposition 2.3.8 ([2])**.** Every self-adjoint operator is normal.

Theorem 2.3.9 ([2])**.** If a matrix A is normal, then it is unitarily equivalent to a diagonal matrix. That is, there exists some unitary matrix U such that

$$UAU^*$$

is a diagonal matrix.

Lemma 2.3.10 (Exercise in [2])**.** The eigenvalues of a self-adjoint matrix must be real.

The following corollary is a consequence of the fact that every self-adjoint matrix is normal, and that the eigenvalues of a self-adjoint matrix must be real.

Corollary 2.3.11 (Deduced from propositions in [2])**.** If a A is self-adjoint, then it is unitarily equivalent to a diagonal matrix with real entries. That is, there exists some unitary matrix U such that

$$UAU^*$$

is a diagonal matrix with real entries.

2.4 Eigenvalues, characteristic polynomials and determinants

In this section, I include some proofs from [29] and [16] for certain classical facts about determinants, the proofs of which are a little hard to come by. I also prove a fact assumed in [27] about the minors of diagonal matrices, which is specialized enough that the reader may not have seen it explicitly before, though it should come as no surprise. We will assume that the inner product space V is finite-dimensional.

Definition 2.4.1. A generalized eigenvector of a matrix $T : V \to V$ is some nonzero $v \in V$ for which there exists $\lambda \in \mathbb{C}, k \geq 1$ such that

$$(T - \lambda I)^k v = 0.$$

In this case, we say λ is the eigenvalue corresponding to v. If $k = 1$, we simply call v an *eigenvector*.

Definition 2.4.2. The multiplicity of an eigenvalue λ of $T : V \to V$ is the size of a maximal linearly independent subset $\{b_1, b_2, ..., b_r\}$ of V, composed of generalized eigenvectors b_i of T, all with eigenvalue λ.

Definition 2.4.3. The characteristic polynomial $\chi(T)$ of a matrix $T : V \to V$ is the product of binomials $\prod_\lambda (x - \lambda)^{m_\lambda}$ over all eigenvalues λ of T, where m_λ is the multiplicity of λ in T.

Proposition 2.4.4 ([21] states this fact in the language of change-of-basis matrices)**.** The eigenvalues of a linear transformation, counted with multiplicity, are invariant under conjugation by $\mathrm{GL}(d)$. Consequently, for any $d \times d$ matrix A and $B \in \mathrm{GL}(d)$,

$$\chi(A) = \chi(BAB^{-1}).$$

Proposition 2.4.5 ([1])**.** The determinant of a square matrix is the product of its eigenvalues, counted with multiplicity.

We will develop some important properties of matrix minors in this section, including properties of the determinant itself. These facts will be key lemmas for the proofs we will develop related to finite free probability.

We will make use of the following well-known expansions of the determinant in order to prove some lesser-known facts.

Theorem 2.4.6 (Leibniz formula, [2])**.** For a $d \times d$ matrix A,

$$\det A = \sum_{\sigma \in \mathfrak{S}_d} \mathrm{sgn}(\sigma) \prod_{i=1}^{d} A_{i,\sigma(i)}.$$

Theorem 2.4.7 (Laplace expansion, [16])**.** For a $d \times d$ matrix A and for all $i \in [d]$,

$$\det A = \sum_{j=1}^{d} (-1)^{i+j} A_{i,j} [A]_{[d]\setminus\{i\},[d]\setminus\{j\}}.$$

Proposition 2.4.8 ([29])**.** Given $d \times d$ matrices A and B, $k \leq d$ and $S, T \in \binom{[d]}{k}$, we have

$$[A + B]_{S,T} = \sum_{i=0}^{k} \sum_{U,V \in \binom{[k]}{i}} (-1)^{\|U\|_1 + \|V\|_1} [A]_{U(S),V(T)} [B]_{\overline{U}(S),\overline{V}(T)}.$$

Proof. For $S \subseteq [d]$, let X_U be given such that $(X_U)_{\cdot,i} = A_{\cdot,i}$ when $i \in S$ and $(X_U)_{\cdot,i} = B_{\cdot,i}$ when $i \notin S$. Then

$$\begin{aligned}\det(A+B) &= \det(A_{\cdot,1}+B_{\cdot,1},...,A_{\cdot,d}+B_{\cdot,d})\\ &= \sum_{i=0}^{d}\sum_{U\in\binom{[d]}{i}} \det X_U\end{aligned}$$

by the multilinearity of the determinant. Now, fix U, and consider $\det X_U$. We have by the Laplace expansion, that for all $i \in [d]$,

$$\begin{aligned}\det X_U &= \sum_{j=1}^{d}(-1)^{i+j}(X_U)_{i,j}[X_U]_{[|U|]\setminus\{i\},[|U|]\setminus\{j\}}\\ &= \sum_{j=1}^{d}(-1)^{i+j}X_{U(i),U(j)}[X]_{\overline{\{i\}}(U),\overline{\{j\}}(U)}.\end{aligned}$$

We will iteratively Laplace-expand $\det X_U$ column-wise along each $i \in U$, noting that if $U = \emptyset$ then $\det X_U = \det B$. We then acquire

$$\begin{aligned}\det X_U &= \sum_{V\in\binom{[d]}{|U|}}(-1)^{\|U\|_1+\|V\|_1}[X]_{U,V}[X]_{\overline{U}([d]),\overline{V}([d])}\\ &= \sum_{V\in\binom{[d]}{|U|}}(-1)^{\|U\|_1+\|V\|_1}[A]_{U,V}[B]_{\overline{U}([d]),\overline{V}([d])}.\end{aligned}$$

Then

$$\begin{aligned}\det(A+B) &= \sum_{i=0}^{d}\sum_{U\in\binom{[d]}{i}}\sum_{V\in\binom{[d]}{|U|}}(-1)^{\|U\|_1+\|V\|_1}[A]_{U,V}[B]_{\overline{U}([d]),\overline{V}([d])}\\ &= \sum_{i=0}^{d}\sum_{U,V\in\binom{[d]}{i}}(-1)^{\|U\|_1+\|V\|_1}[A]_{U,V}[B]_{\overline{U}([d]),\overline{V}([d])}.\end{aligned}$$

The equality in the proposition statement is a consequence of simply restricting rows and columns of each $A+B$ from $[d]$ to S and T, respectively. ☐

Proposition 2.4.9 ([27])**.** If $D \in \operatorname{End}\mathbb{C}^d$ is a diagonal $d \times d$ matrix, and $S, T \in \binom{[d]}{k}$, then $S \neq T$ implies $[D]_{S,T} = 0$.

Proof. Suppose without loss of generality that $i \in T$ but $i \notin S$. Then the column in $D_{S,T}$ corresponding to the ith column of D contains only zeroes. By the linearity of the determinant in each column of its argument, we have that $[D]_{S,T} = 0$. ☐

Definition 2.4.10. Let A be a matrix. By $\sigma_k(A)$, we denote the coefficient of $(-1)^k x^{d-k}$ in the characteristic polynomial $\chi(A)$.

Lemma 2.4.11 ([2])**.** For each $d \times d$ matrix A, there exists a $d \times d$ nonsingular matrix X such that XAX^{-1} is upper-triangular.

Proposition 2.4.12 ([8])**.**

$$\sigma_k(A) = \sum_{S \in \binom{[d]}{k}} [A]_S.$$

Proof. Consider the eigenvalues of A to be given with multiplicity in a tuple $(\lambda_1, \lambda_2, ..., \lambda_d)$. Since we are dealing with $\chi(A)$, we can assume without loss of generality by 2.4.4 and 2.4.11 that A is upper-triangular. We will furthermore assume that λ_i is the i, ith entry of A, i.e., the eigenvalues are distributed along the diagonal of A. We have

$$\begin{aligned}
\chi(A) &= \prod_{\lambda} (x - \lambda)^{m_\lambda} \\
&= \sum_{i=0}^{d} \sum_{S \in \binom{[d]}{d-i}} x^{d-i} \prod_{j \in [d] \setminus S} (-\lambda_j) \\
&= \sum_{i=0}^{d} \sum_{S \in \binom{[d]}{d-i}} x^{d-i} (-1)^i \prod_{j \in [d] \setminus S} \lambda_j,
\end{aligned}$$

so the coefficient of $x^{d-i}(-1)^i$ in $\chi(A)$ is

$$\begin{aligned}
\sum_{S \in \binom{[d]}{d-i}} \prod_{j \in [d] \setminus S} \lambda_j &= \sum_{S \in \binom{[d]}{i}} \prod_{j \in S} \lambda_j \\
&= \sum_{S \in \binom{[d]}{i}} [A]_S.
\end{aligned}$$

The final equality relies on the fact that the eigenvalues are, without loss of generality, distributed along the diagonal of A as described at the beginning of this proof. $\square$

Definition 2.4.13. Let V be a vector space. The underlying set of the tensor algebra $T(V)$ is given by

$$T(V) = \bigoplus_{k=0}^{\infty} V^{\otimes k},$$

where we define $V^{\otimes 0} = \mathbb{C}$.

This is a vector space under componentwise scalar multiplication and addition, and it is a ring via the product

$$(u, v) \mapsto u \otimes v.$$

Definition 2.4.14. If V is a vector space over a field of characteristic zero, then denote by $\bigwedge V$ the exterior algebra

$$T(V)/I,$$

where I is the ideal generated by $\{v \otimes v \mid v \in V\}$.

We denote the product in this ring by $u \wedge v$, calling it the *wedge product*. Denote by $\bigwedge^k V$ the sub-vector space of $\bigwedge V$ comprising wedges of exactly k vectors. We refer to this as the kth exterior power of V.

If we have a basis $\{e_i\}_{i=1}^d$ for V, then we have a natural basis $\{f_S\}_{S\in\binom{[d]}{k}}$ for $\bigwedge^k V$, where f_S is defined as the wedge product $\bigwedge_{i\in S} e_i$.

Definition 2.4.15. Given a square matrix $T : V \to V$, denote by $\bigwedge^k T$ the matrix mapping $\bigwedge^k V \to \bigwedge^k V$ with entries

$$\left(\bigwedge^k T\right)_{A,B} = [T]_{A,B}.$$

Theorem 2.4.16 (Cauchy-Binet, [24], [16])**.** $\bigwedge^k$ is a monoid homomorphism with respect to matrix multiplication. That is, if we let A and B be $d \times d$ matrices, and let $0 \leq k \leq d$, then

$$\bigwedge^k AB = \bigwedge^k A \bigwedge^k B.$$

Entrywise, we may express this via the well-known identity,

$$[AB]_{S,T} = \sum_{U\in\binom{[d]}{k}} [A]_{S,U}[B]_{U,T}.$$

Proof. Let $k = |S| = |T|$. Note that

$$[AB]_{S,T} = \det\left((AB)_{S,T}\right) = \det[(A)_{S,[d]}(B)_{[d],T}].$$

$$
\begin{aligned}
[AB]_{S,T} &= \sum_{\sigma\in\mathfrak{S}_k} \operatorname{sgn}(\sigma) \prod_{i=1}^{k} (A_{S,[d]}B_{[d],T})_{i,\sigma(i)} \\
&= \sum_{\sigma\in\mathfrak{S}_k} \operatorname{sgn}(\sigma) \prod_{i=1}^{k} \sum_{j=1}^{d} A_{S(i),j}B_{j,T(\sigma(i))} \\
&= \sum_{\sigma\in\mathfrak{S}_k} \operatorname{sgn}(\sigma) \sum_{(j_1,\dots,j_k)\in[d]^k} \prod_{i=1}^{k} A_{S(i),j_i}B_{j_i,T(\sigma(i))} \\
&= \sum_{\sigma\in\mathfrak{S}_k} \operatorname{sgn}(\sigma) \sum_{(j_1,\dots,j_k)\in[d]^k} \left(\prod_{i=1}^{k} A_{S(i),j_i}\right)\left(\prod_{i=1}^{k} B_{j_i,T(\sigma(i))}\right) \\
&= \sum_{(j_1,\dots,j_k)\in[d]^k} \left(\prod_{i=1}^{k} A_{S(i),j_i}\right)\left(\sum_{\sigma\in\mathfrak{S}_k} \operatorname{sgn}(\sigma) \prod_{i=1}^{k} B_{j_i,T(\sigma(i))}\right) \\
&= \sum_{(j_1,\dots,j_k)\in[d]^k} \left(\prod_{i=1}^{k} A_{S(i),j_i}\right) \det(B_{j_{[k]},T}),
\end{aligned}
$$

where we define $B_{j_{[k]},T}$ to be the matrix for which $(B_{j_{[k]},T})_{r,s} = B_{j_r,T(s)}$. If $j_r = j_s$ for distinct $j, s \in [k]$, then the summand corresponding to that tuple $(j_1, ..., j_k)$ is equal to zero. So we have

$$
\begin{aligned}
[AB]_{S,T} &= \sum_{\substack{(j_1,\dots,j_k)\in[d]^k \\ r\neq s\Rightarrow j_r\neq j_s}} \left(\prod_{i=1}^{k} A_{S(i),j_i}\right) \det(B_{j_{[k]},T}) \\
&= \sum_{1\leq j_1<\dots<j_k\leq d} \sum_{\sigma\in\mathfrak{S}_k} \left(\prod_{i=1}^{k} A_{S(i),j_{\sigma(i)}}\right) \det(B_{(j\circ\sigma)_{[k]},T}) \\
&= \sum_{1\leq j_1<\dots<j_k\leq d} \sum_{\sigma\in\mathfrak{S}_k} \operatorname{sgn}(\sigma) \left(\prod_{i=1}^{k} A_{S(i),j_{\sigma(i)}}\right) \det(B_{j_{[k]},T}) \\
&= \sum_{1\leq j_1<\dots<j_k\leq d} \det(A_{S,j_{[k]}}) \det(B_{j_{[k]},T}) \\
&= \sum_{U\in\binom{[d]}{k}} \det(A_{S,U}) \det(B_{U,T}) \\
&= \sum_{U\in\binom{[d]}{k}} [A]_{S,U}[B]_{U,T}.
\end{aligned}
$$

$\square$

Chapter 3

Measure theory

In this chapter, I will develop just enough measure theory to serve as a foundation for probability theory. I will also define the Haar measure.

3.1 Measure spaces

The concept of a measure space allows us to talk rigorously about the "volume" of a set in that space. Since Kolmogorov's axiomatization of probability theory in terms of measure spaces in 1933, measure spaces have been the standard way to model probabilistic phenomena. That is the use we will put this concept to in this thesis.

Effective references for measure theory are [36] and [3].

Definition 3.1.1. A σ-algebra Σ on a set X is a collection of subsets of X for which

- $\emptyset \in \Sigma$,
- if $A \in \Sigma$ then $A^c \in \Sigma$, and
- if $(A_n)_{n\in\mathbb{N}}$ is a sequence in Σ, then $\bigcup_{n\in\mathbb{N}} A_n \in \Sigma$.

Definition 3.1.2. A measurable space (X, Σ) is a set X endowed with a σ-algebra Σ.

Definition 3.1.3. A measure μ on a measurable space (X, Σ) is a function $\Sigma \to [0, \infty]$ for which

- $\mu(\emptyset) = 0$, and
- if (A_n) is a sequence of disjoint sets in Σ, then $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

The second axiom is called *countable additivity.* We will often suppress Σ and speak of X itself as being a measurable space. Furthermore, we refer to an element $A \in \Sigma$ as being a measurable set of X.

Definition 3.1.4. A measure space is a triple (X, Σ, μ), where (X, Σ) is a measurable space, and μ is a measure on that space. We often abuse notation by suppressing Σ and μ and referring to X itself as a measure space.

Definition 3.1.5. If X is a topological space, then we call the smallest σ-algebra containing all the open sets of X the *Borel subsets* of X. If a set A is an element of the Borel subsets of X, we call A a Borel set. If μ is a measure on the (X, Σ), where Σ are the Borel subsets of X, we call μ a Borel measure.

Definition 3.1.6. If X is a topological space, (X, Σ) is a measurable space, and Σ contains the Borel subsets of X, then we call a measure μ on X inner regular if for all $A \in \Sigma$,

$$\mu(A) = \sup_{K \in \mathcal{K}(X), K \subseteq A} \mu(K),$$

where $\mathcal{K}(X)$ is the set of compact subsets of X. We call μ outer regular if for all $A \in \Sigma$,

$$\mu(A) = \inf_{O \in \mathcal{O}(X), A \subseteq O} \mu(O),$$

where $\mathcal{O}(X)$ is the set of open subsets of X. We call μ regular if it is both inner regular and outer regular.

Definition 3.1.7. If μ is a measure on X, then we call μ a probability measure if $\mu(X) = 1$. We will often write probability measures with the symbol $\mathbb{P}$.

Definition 3.1.8. We refer to a function $f : X \to Y$ between measurable spaces X and Y as a measurable function if the preimage of any measurable set in Y is a measurable set in X. If the Σ-algebra on X lies within the collection of Borel subsets of X, then we say that f is Borel-measurable.

Definition 3.1.9. We say a measure μ over a Hausdorff space X is compactly supported if there exists some compact set K for which $\mu(X \setminus K) = 0$. In such a case, we define $\mathrm{supp}(K)$ to be the smallest such K.

3.2 Functional analysis

We will need to use some tools from functional analysis to make a couple of subtle points about measures, but this vast and vibrant theory will remain at the margins of this thesis.

I used Gert Pederson's [34] as a reference for this section.

Definition 3.2.1. If V and W are normed vector spaces, we call a linear map $T : V \to W$ a bounded operator if

$$\sup_{\|x\|_V \leq 1} \|Tx\|_W$$

exists.

Definition 3.2.2. If V and W are normed vector spaces, then let $\mathcal{B}(V, W)$ denote the vector space of bounded operators $T : V \to W$. The norms of V and W induce the *operator norm* on $\mathcal{B}(V, W)$, given by

$$\|T\| = \sup_{\|x\|_V \leq 1} \|Tx\|_W.$$

Definition 3.2.3. Let X be a compact Hausdorff space. Then denote by $C_{\mathbb{K}}(X)$ the algebra of continuous functions $X \to \mathbb{K}$, where $\mathbb{K}$ is a field. If the field is evident from context, we will suppress the subscript. $C(X)$ is a Banach space with respect to the uniform norm, i.e., the norm

$$\|f\| = \sup_{x \in X} |f(x)|.$$

Definition 3.2.4. Let X be a compact Hausdorff space. Then denote by $\mathcal{M}(X)$ the space of signed measures on X.

Definition 3.2.5. A functional f on a normed vector space V is an element of the space of bounded linear maps $\mathcal{B}(V, \mathbb{K})$, which we denote as V^*.

Proposition 3.2.6. Given a compact Hausdorff space X, $C(X)^* = \mathcal{M}(X)$.

Theorem 3.2.7 (Stone-Weierstrass, real)**.** Suppose X is a compact Hausdorff topological space, and A is subalgebra of the real algebra $C_{\mathbb{R}}(X)$ containing the constant functions and separating points in X. Then A is dense in $C(X)$ with respect to the uniform norm.

3.3 Haar measure

The concept of Haar measure exists to generalize an important observation about the real line. That is, the length of the interval $(0, 1)$ is equal to the length of the interval $(0 + x, 1 + x)$ for any real number x. Our geometric intuition makes it hard to imagine things being otherwise with the real line, but there is *a priori* no reason to expect that a given measure space equipped with a group operation will have such a tidy relationship between its operation and its measure.

In this section we will reproduce several essential results which we will not derive, since the proofs are quite involved. A good reference for these facts is [11].

Definition 3.3.1. If G is a topological group, then we call a measure μ over the Borel subsets of G a *left Haar measure* (respectively, *right Haar measure*) if μ satisfies:

- $\mu(A) = \mu(gA)$ (respectively, $\mu(A) = \mu(Ag)$) for all Borel sets $A \subseteq G$ and all $g \in G$,
- μ is regular,

- if $K \subseteq G$ is compact, then $\mu(K) < \infty$, and
- $\mu(G) \neq 0$.

Theorem 3.3.2 ([11])**.** Every locally compact Hausdorff topological group G enjoys at least one Haar measure. In fact, this measure is unique up to a multiplicative constant, i.e., for all Haar measures μ and ν on G, there exists some $c > 0$ for which $\mu = c\nu$.

Under certain circumstances, the left and right Haar measure on a group will be the same measure.

Proposition 3.3.3 ([11])**.** If G is an abelian, locally compact Hausdorff topological group, then a left Haar measure on G is also a right Haar measure on G, and vice versa.

Proof. Since G is abelian, we can equate

$$\mu(gA) = \mu(Ag),$$

and so the properties of a left Haar measure satisfy the right Haar measure axioms, and vice versa. $\square$

Theorem 3.3.4 ([11])**.** If G is a compact Hausdorff topological group, then a left Haar measure on G is also a right Haar measure on G, and vice versa.

Thus for compact groups, we speak of *Haar measure*, dropping the left/right distinction. By convention, we normalize a Haar measure on a compact group G so $\mathbb{P}$ is a probability measure, and we write it with the symbol $\mathbb{P}$.

documentclass[../main.tex]subfiles

Chapter 4

Probability

The goal of this chapter is to establish the relationship between a measure over a real or complex vector space, and that measure's moments. In particular, these moments often identify the measure uniquely. The moments of a distribution, especially the first four, have broad use in statistics. These are used to compute to a probability distribution's mean, variance, skewness and kurtosis, each of which measures an important qualitative property of the distribution. Moments higher than the fourth become harder to interpret, but are still useful in numerical algorithms and proofs. This chapter will primarily concern itself with the classification problem, but the reader should keep in mind that these moments have broad practical applications.

We will be assuming some definitions and theorems from measure theory and functional analysis, which may be found in the the previous chapter. The first chapter of [31] has a good introductory section on the moments of a probability measure, in the context of free probability.

4.1 Foundations

Definition 4.1.1. If Ω is a measure space equipped with a σ-algebra Σ and a probability measure $\mathbb{P}$, then we call $(\Omega, \Sigma, \mathbb{P})$ a probability space. We refer in this setting to the set Ω as a sample space, we refer to Σ as an event space. Abusing notation, we will often suppress Σ and $\mathbb{P}$ and refer to Ω itself as a probability space.

Definition 4.1.2. We refer to a measurable function $X : \Omega \to \mathbb{R}$ of a probability space Ω as a (real) random variable on Ω. We denote by $\mathcal{M}_{\mathbb{R}}(\Omega)$ the set of random variables on Ω. We will often suppress reference to the probability space over which a random variable is defined.

Definition 4.1.3. We associate to every random variable X a probability measure μ_X on $\mathbb{R}$. We define this measure as

$$\mu_X(A) = \mathbb{P}(X^{-1}(A)).$$

Definition 4.1.4. Given a probability space Ω, we define the expectation operator $\mathbb{E} : \mathcal{M}_{\mathbb{R}}(\Omega) \to \mathbb{R}$ over Ω in the following way. For any random variable X on Ω, we define

$$\mathbb{E}[X] = \int_{\mathbb{R}} t \, d\mu_X(t)$$

if the right hand side is a real number. The expectation operator is in general a partial function, i.e., there are many X for which $\mathbb{E}[X]$ is not defined. We will often suppress reference to the probability space over which $\mathbb{E}$ is defined.

4.2 Moments of random variables

Definition 4.2.1. Given a measure μ on $\mathbb{R}$, we refer to the quantity $\int t^k \, d\mu(t)$ as the kth moment of μ, for all integers $k \geq 0$. This quantity does not always exist; we say that μ possesses a kth moment if the quantity exists. We may ignore the zeroth moment of μ when it is known we are working with probability measures, since all probability measures have a zeroth moment equal to one.

It is not entirely self-evident what the most convenient way of representing measures is for the purposes of writing proofs or performing computations, since measures can in the general case contain an enormous amount of information. One convenient technique is to summarize a measure on the real line in terms of its moments. These moments come in the form of a *moment sequence* $(m_k)_{k \geq 0}$ taking values in $\mathbb{R} \cup \{*\}$, where the symbol $*$ stands in for an undefined quantity. One question we may ask is whether a moment sequence uniquely identifies the measure it originates from, or whether two distinct measures may have identical moment sequences. It turns out that in the situations we will be considering in this thesis (i.e., compactly supported distributions) the distribution is identified uniquely by its moment sequence.

We will define the moments of a random variable to be the moments of its corresponding measure.

Definition 4.2.2. The kth central moment of a random variable X is the quantity

$$\mathbb{E}[(X - \mathbb{E}(X))^k].$$

The first central moment of any random variable is identically zero. We call the second central moment of X the variance of X, and denote it $\mathrm{Var}(X)$. The third and fourth central moments are named skewness and kurtosis, respectively, though we won't be making use of these quantities.

Lemma 4.2.3 ([4])**.** If $f : \mathbb{R} \to \mathbb{R}$ is a measurable function, then either

$$\mathbb{E}[f(X)] = \int f(t) \, d\mu_X(t),$$

or both sides of the equality are undefined.

Proof. If Ω is the underlying probability space, with measure $\mathbb{P}$, then

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} t \, d\mu_{f \circ X}(t) = \int_{\Omega} f \circ X(\omega) \, d\mathbb{P}(\omega).$$

Likewise,

$$\int_{\mathbb{R}} f(t) \, d\mu_X(t) = \int_{\Omega} f(X(\omega)) \, d\mathbb{P}(\omega) = \int_{\Omega} f \circ X(\omega) \, d\mathbb{P}(\omega).$$

The two expressions in the equation reduce to the same expression, so they are either both undefined, or they are equal. □

Corollary 4.2.4 ([4])**.** In particular, $\mathbb{E}[X^k] = \int t^k \, d\mu_X(t)$.

Proposition 4.2.5 (Exercise in [31])**.** If X is a random variable for which the associated measure μ_X has compact support, then X has all moments.

Proof. Let

$$M = \sup\{|r| : r \in \operatorname{supp}(\mu_X)\}.$$

Then by 4.2.3,

$$\left|\mathbb{E}[X^k]\right| \leq \int |t|^k \, d\mu_X(t) \leq M^k.$$

□

Moment sequences do not always uniquely identify measures in the non-compact case. However, a substantial proportion of the measures which arise in applications may be uniquely identified by their moment sequence. First, we will state without proof a fairly general sufficient condition known as Carleman's condition.

Theorem 4.2.6 (Carleman's condition, [7])**.** Let X be a random variable possessing all moments. For each integer $k \geq 1$, let m_k be the kth moment of X. If

$$\sum_{k=1}^{\infty} m_{2k}^{-\frac{1}{2k}} = \infty,$$

then the sequence (m_k) uniquely identifies the measure μ_X. I.e., for any other random variable Y possessing the same moment sequence as X, we can infer that $\mu_X = \mu_Y$.

I will now develop a simpler argument that applies only to the case of compactly supported measures.

Lemma 4.2.7 (Exercise in [39])**.** Let μ and ν be measures on the compact interval $[a, b]$. Then $\mu = \nu$ if and only if $(m_k)_{k=0}^{\infty}$ is the moment sequence of both μ and ν.

Proof. This lemma is an application of the Stone-Weierstrass theorem 3.2.7. The algebra of polynomials on $[a, b]$ is a subalgebra of $C([a, b])$ containing the constant functions, and it separates points. Therefore, the subalgebra is uniformly dense. For any function $f \in C([a, b])$ and for any $\epsilon > 0$, there exists a polynomial p such that $|f(x) - p(x)| < \epsilon$ for all $x \in [a, b]$. So

$$\int_a^b |f(t) - p(t)|\, d\mu(t) < \epsilon m_0 \quad \text{and} \quad \int_a^b |f(t) - p(t)|\, d\nu(t) < \epsilon m_0.$$

This means, in particular, that $|\int_a^b f(t)\, d\gamma(t) - \int_a^b p(t)\, d\gamma(t)| < \epsilon m_0$ for $\gamma = \mu$ and $\gamma = \nu$.

Note also that, for any polynomial $p(x) = \sum_i \alpha_i x^i$, we have

$$\begin{aligned}
\int_a^b p(t)\, d\mu(t) &= \int_a^b \sum_i \alpha_i t^i\, d\mu(t) \\
&= \sum_i \alpha_i \int_a^b t^i\, d\mu(t) \\
&= \sum_i \alpha_i m_i \\
&= \sum_i \alpha_i \int_a^b t^i\, d\nu(t) \\
&= \int_a^b p(t)\, d\nu(t).
\end{aligned}$$

Since we can approximate $\int_a^b f(t)\, d\mu(t)$ and $\int_a^b f(t)\, d\nu(t)$ simultaneously by one sequence of polynomials $(p_i(x))$, and since μ and ν are equal on polynomials, we have the equality

$$\int_a^b f(t)\, d\mu(t) = \int_a^b f(t)\, d\nu(t)$$

for all $f \in C([a, b])$. Since the space of (signed) measures $\mathcal{M}([a, b])$ is the dual of the space of continuous functions $C([a, b])$ (see 3.2.6), we may conclude $\mu = \nu$. $\square$

I had difficulty sourcing a citation for the following lemma, however there is nothing novel about it, and I am sure it has been stated before. Perhaps it can be found in the writings of Felix Hausdorff, since he worked on related topics in the 1920s.

Lemma 4.2.8. Let μ be a measure on the real line. Then μ has compact support if and only if its even moments are exponentially bounded, i.e., there exists an $M \geq 0$ such that

$$\int t^{2k}\, d\mu(t) \leq M^{2k}$$

for all $k \geq 0$. (If any even moments are undefined, we will consider this a failure to be exponentially bounded.)

Proof. Denote the moment sequence of μ by (m_k) for $k \geq 0$.

Let μ have compact support in $[-R, R]$ for some $R \in \mathbb{R}$.

$$\int t^{2k}\, d\mu(t) \leq R^{2k} m_0 \leq (Rm_0)^{2k}.$$

That is to say, the even moments of μ are exponentially bounded with respect to the constant $M = Rm_0$.

Now suppose that μ is not compactly supported, that is, for all $R \geq 0$, $\mu(\mathbb{R} \setminus [-R, R]) > 0$. If the even moments of μ are not defined, then they certainly aren't exponentially bounded. So assume they are defined. Let $M > 0$. Then by hypothesis, $\mu(\mathbb{R} \setminus [-M, M]) > 0$. Denote this quantity by W. Then

$$\begin{aligned}
\int t^{2k}\, d\mu(t) &= \int_{-M}^{M} t^{2k}\, d\mu(t) + \int_{\mathbb{R}\setminus[-M,M]} t^{2k}\, d\mu(t) \\
&\geq \int_{\mathbb{R}\setminus[-M,M]} t^{2k}\, d\mu(t) \\
&> WM^{2k}.
\end{aligned}$$

Let $\delta > 0$ such that

$$\int_{\mathbb{R}\setminus[-M,M]} t^{2k}\, d\mu(t) = W(M+\delta)^{2k}.$$

Then

$$\int t^{2k}\, d\mu(t) \geq W(M+\delta)^{2k}.$$

So as $k \to \infty$,

$$\begin{aligned}
\int t^{2k}\, d\mu(t)/M^{2k} &\geq W(M+\delta)^{2k}/M^{2k} \\
&\to \infty.
\end{aligned}$$

Therefore, there exists some $K \geq 0$ for which $k \geq K$ implies

$$\int t^{2k}\, d\mu(t) > M^{2k}.$$

Since M was arbitrary, the even moments of μ are not exponentially bounded. $\square$

The following theorem can be seen as a mere corollary of Carlemann's condition, and thus is rarely stated in this form.

Theorem 4.2.9. Let μ be a measure on the real line, with compact support. Suppose a measure ν has the same moments as μ. Then $\mu = \nu$.

Proof. We will argue by the contrapositive, assuming $\mu \neq \nu$. There are two cases. In the first case, ν is not compactly supported. Then by 4.2.8, we can distinguish μ from ν by comparing the growth rates of their even moments. In the second case, μ and ν are both compactly supported. Then they are also measures on the compact interval

$$[\min(\operatorname{supp}(\mu) \cup \operatorname{supp}(\nu)), \max(\operatorname{supp}(\mu) \cup \operatorname{supp}(\nu))].$$

By 4.2.7, the two measures must differ in their moments. $\square$

Definition 4.2.10. Suppose we have a collection of random variables $\{X_i\}_{i \in I}$. Given a product $\prod_{k=1}^{n} X_{i_k}$ where $i : [n] \to I$, we refer to the expectation of that product as a *mixed moment* of the $\{X_i\}$. We say that a mixed moment formed in this way has degree n. A *quadratic moment* is a mixed moment of degree 2, a *cubic moment* is a mixed moment of degree 3, and so on.

The title of this thesis, "Second-order finite free probability", relates to the fact that we will be investigating the quadratic moments of a certain family of random variables.

4.3 Random vectors

Like many concepts definable over real numbers, we can extend the concept of random variables into multiple dimensions to acquire a random vector. A particular random vector lies at the centre of this thesis, though the following definition is not yet the right one, because it does not incorporate complex numbers. We will deal that in the following two sections.

Definition 4.3.1. A random vector X is a measurable function $X : \Omega \to \mathbb{R}^n$, where Ω is a probability space. If π_i is a projection $\mathbb{R}^n \to \mathbb{R}$ onto the ith component of $\mathbb{R}^n$, then we refer to the random variable $\pi_i(X)$ as a component of X.

Definition 4.3.2. If Ω is a probability space, then we denote by $\mathcal{M}_{\mathbb{R}^n}(\Omega)$ the set of random vectors $\Omega \to \mathbb{R}^n$.

Definition 4.3.3. If X is a random vector in $\mathbb{R}^n$, we can speak of its associated joint distribution μ_X, which is a probability measure on $\mathbb{R}^n$. We again define

$$\mu_X(A) = \mathbb{P}(X^{-1}(A)).$$

Definition 4.3.4. Given a probability space Ω and an integer $n \geq 2$, we can again define an expectation operator $\mathbb{E} : \mathcal{M}_{\mathbb{R}^n}(\Omega) \to \mathbb{R}^n$ by

$$\mathbb{E}[X] = \int_{\mathbb{R}^n} t \, d\mu_X(t).$$

Just like in the one-dimensional case, $\mathbb{E}$ is in general a partial function on random vectors.

Definition 4.3.5. If X is a random vector, then we refer to the mixed moments of the components of X as the moments of X.

Like with the one-dimensional case, a multivariate distribution may be identified more or less effectively by its moments. The following is a sufficient condition developed in 2001 by de Jeu.

Theorem 4.3.6 (Extended Carleman's condition [20])**.** Let μ be a measure on $\mathbb{R}^n$ such that

$$\int_{\mathbb{R}^n} \|t\|^k d\mu(t) < \infty$$

for all $k \geq 0$. Let $\{v_1, ..., v_n\}$ be a basis for $\mathbb{R}^n$. For $j = 1, ..., n$ and $k = 0, 1, 2, ...$, define

$$s_j(k) = \int_{\mathbb{R}^n} \langle v_j | t \rangle^k \, d\mu(t).$$

If each of the sequences $\{s_j(k)\}_{m=1}^{\infty}$ satisfies Carleman's condition

$$\sum_{k=1}^{\infty} s_j(k)^{-\frac{1}{2k}} = \infty,$$

then μ is uniquely identified by its mixed moments.

However, we will only be trying to characterize measures of compact support in this thesis. The proof that they are uniquely identified by their moments is again an easy application of the Stone-Weierstrass theorem.

Proposition 4.3.7. Let μ be a measure with compact support over $\mathbb{R}^n$. If ν is another measure with the same mixed moments as μ, then $\mu = \nu$.

Proof sketch. The polynomial ring $\mathbb{R}[x_1, ..., x_n]$ is yet again a subalgebra of $C(\text{supp}(\mu))$ containing the constant functions and separating points. Therefore, we can adapt a version of 4.2.7 to this setting.

Likewise, we can replicate a version of 4.2.8. Let $\iota_j : \mathbb{R} \to \mathbb{R}^n$ be the injection into the jth component. Then define, for $1 \leq j \leq n$,

$$\mu_j(A) = \mu(\{\iota_j(a) \mid a \in A\}).$$

The measure μ is compactly supported if and only if each of these projections μ_j is compactly supported. We can detect from the mixed moments $\int_{\mathbb{R}^n} t_j^{2k} \, d\mu(t)$ whether each μ_j is compactly supported, therefore, we can determine whether μ is compactly supported from its moments.

Thus we can use moment sequences to detect any differences between μ and ν both in the case in which ν is compactly supported, and the case in which ν is not compactly supported. $\square$

4.4 Complex random variables

In this section, we will define complex random variables in terms of two-dimensional random vectors, analogous to the way in which $\mathbb{C}$ can be defined as $\mathbb{R}^2$ equipped with a multiplication operation.

See [25] for more on complex random variables.

Definition 4.4.1. A complex random variable X is a random vector $X : \Omega \to \mathbb{R}^2$, where the codomain is identified with $\mathbb{C}$.

Definition 4.4.2. Given a complex random variable X, we call the quantity

$$\mathbb{E}[X^{\epsilon_1} X^{\epsilon_2} ... X^{\epsilon_k}]$$

a kth *-moment of X if $\epsilon_i \in \{1, *\}$ for all $i = 1, ..., k$. (This is a slight abuse of notation–what I mean is that we may take the adjoint of some of the X_i.) If $\epsilon_i = 1$ for all i then we call the quantity the kth moment of X.

I had a difficult time sourcing this lemma in the literature, but it's a fairly straightforward application of the algebra of complex numbers and the linearity of the expectation operator. I suspect that the $*$-moments of complex random vectors is a niche enough topic that most of the writing about them is targeted at experts, who likely take this fact for granted.

Lemma 4.4.3. Suppose X is a complex random variable. Then moments of X as a random vector are a finite linear combination of the *-moments of X as a complex random variable, and vice versa.

Proof. By conventional manipulations of complex numbers, we have the following identities.

$$\Re X = \frac{X + X^*}{2}$$
$$\Im X = \frac{X - X^*}{2}$$
$$X = \Re X + i \Im X = \frac{X + X^*}{2} + i \frac{X - X^*}{2}$$

We can thus recover the linear *-moments of X as finite linear combinations of the linear mixed moments of its real and imaginary parts, and vice versa. For higher-order moments we have

$$\begin{aligned}
&\mathbb{E}[X^m(X^*)^n] \\
&= \mathbb{E}[(\Re X + i\Im X)^m \cdot (\Re X - i\Im X)^n] \\
&= \mathbb{E}\left[\left(\sum_{i=0}^{m}\binom{m}{i}(\Re X)^i(i)^{m-i}(\Im X)^{m-i}\right)\left(\sum_{j=0}^{n}\binom{n}{j}(\Re X)^j i^{n-j}(\Im X)^{n-j}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{m}\sum_{j=0}^{n} i^{m+n-i-j}\binom{m}{i}\binom{n}{j}(\Re X)^{i+j}(\Im X)^{m+n-i-j}\right] \\
&= \sum_{i=0}^{m}\sum_{j=0}^{n} i^{m+n-i-j}\binom{m}{i}\binom{n}{j}\mathbb{E}[(\Re X)^{i+j}(\Im X)^{m+n-i-j}],
\end{aligned}$$

demonstrating that the $*$-moments of X may be expressed as a finite linear combination of the mixed moments of its real and imaginary parts. Furthermore, we can derive

$$\begin{aligned}
\mathbb{E}[(\Re X)^m(\Im X)^n] &= \mathbb{E}\left[\left(\frac{X+X^*}{2}\right)^m\left(\frac{X-X^*}{2}\right)^n\right] \\
&= 2^{-m-n}\mathbb{E}\left[\sum_{i=0}^{m}\binom{m}{i}X^i(X^*)^{m-i}\sum_{j=0}^{n}\binom{n}{j}(-1)^{n-j}X^j(X^*)^{n-j}\right] \\
&= 2^{-m-n}\sum_{i=0}^{m}\sum_{j=0}^{n}(-1)^{n-j}\binom{m}{i}\binom{n}{j}\mathbb{E}[X^{i+j}(X^*)^{m+n-i-j}],
\end{aligned}$$

proving that the mixed moments of the real and imaginary parts of X may be expressed as a finite linear combination of the $*$-moments of X. ☐

Corollary 4.4.4. If μ is a compactly supported measure over $\mathbb{C}$, then it is uniquely determined by its *-moments.

Proof. This is an application of 4.3.7 and 4.4.3. ☐

Definition 4.4.5. Given a complex random variable X, we call the quantity

$$\mathbb{E}\left[\prod_{i=1}^{k}(X-\mathbb{E}(X))^{\epsilon_i}\right]$$

a kth central *-moment of X if $\epsilon_i \in \{1, *\}$ for all $i = 1, ..., k$. We call the second central *-moment

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X)\overline{(X - \mathbb{E}(X))}]$$

the variance of X. We call the second central *-moment

$$\mathrm{PVar}(X) = \mathbb{E}[(X - \mathbb{E}(X)(X - \mathbb{E}(X))]$$

the pseudovariance of X.

4.5 Complex random vectors

Finally, we can define the concept of complex random vectors. The "characteristic polynomial" mentioned in the title of this thesis is going to be framed as a complex random vector, so this is an important idea.

See [25] for more on complex random vectors.

Definition 4.5.1. A complex random vector X is a measurable function $X : \Omega \to \mathbb{C}^n$, where Ω is a probability space. By definition, each component of X is a complex random variable.

Definition 4.5.2. The moments of a complex random vector X are the mixed *-moments of the components of X.

Proposition 4.5.3. If μ is a compactly supported measure over $\mathbb{C}^n$, then it is uniquely determined by its mixed *-moments.

4.6 Random unitary matrices

In this thesis, we're going to be talking a lot about "Haar-distributed random unitary matrices". Sometimes we will be emphasizing its nature as a "random matrix", which is elaborated in the first definition below, and sometimes we will be emphasizing its nature as a "Haar-distributed random variable", which is elaborated in the second definition below.

Definition 4.6.1. A random $m \times n$ matrix X is a measurable function $X : \Omega \to \mathbb{K}^{mn}$, where Ω is a probability space and $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. We denote the set of such random matrices by $\mathcal{M}_{M_{m\times n}(\mathbb{K})}(\Omega)$. We identify $\mathbb{K}^{mn}$ with the additive group of $m \times n$ matrices over $\mathbb{K}$. If $m = n$, then we consider $\mathcal{M}_{M_{m\times n}(\mathbb{K})}(\Omega)$ to be a ring, inheriting matrix multiplication as its product.

Definition 4.6.2. Let $\mathcal{G}$ be a compact group. Let Ω be a probability space, and let $G : \Omega \to \mathcal{G}$ be a measurable function. Then we call G a random element of $\mathcal{G}$.

Furthermore, let μ_G be the measure on $\mathcal{G}$ given by $\mu_G(A) = \mathbb{P}(G^{-1}(A))$. We call this the distribution of G.

If μ_G is a Haar measure on $\mathcal{G}$, then we say that G is a Haar-distributed element of $\mathcal{G}$.

Chapter 5

Representation theory of compact groups

5.1 Introduction

Definition 5.1.1. A representation $\rho : G \times V \to V$ is a group action on a $\mathbb{C}$-vector space that is linear in its second argument.

We may write the domain and codomain of a representation by $\rho : G \to \operatorname{Aut} V$, denoting by $\operatorname{Aut} V$ the space of $\mathbb{C}$-linear automorphisms of V.

Definition 5.1.2. Let G be a group. Then the group ring (or algebra) $\mathbb{C}G$ may be given by the vector space of finitely-supported formal sums

$$\sum_{g \in G} \alpha_g g,$$

with the multiplicative rule given by

$$\left(\sum_{g \in G} \alpha_g g \right) \left(\sum_{g \in G} \beta_g g \right) = \sum_{g \in G} \sum_{h \in G} \alpha_h \beta_{h^{-1} g} g.$$

One may find in any suitable textbook on representation theory, for example, [12], the fact that the data of a representation of a finite group G is identical to the data of a $\mathbb{C}G$-module, and that all $\mathbb{C}G$-modules may be decomposed into a direct sum of finitely-generated modules.

This basic insight continues to hold for compact infinite groups. In this chapter, we will develop some of the elementary theory of finite-dimensional representations of compact groups. All groups in sight will be compact, and we will assume all representations and modules are continuous. We will also assume that the $\mathbb{C}$-vector spaces over which we will work are inner product spaces.

We will also assume by default that the integral over a compact group is taken with respect to the Haar measure, see 3.3.

5.2 G-invariant inner products

Definition 5.2.1. Let $\rho : G \to \operatorname{Aut} V$ be a representation, and let $\langle\cdot|\cdot\rangle$ be an inner product on V. We say that ρ is unitary, and that $\langle\cdot|\cdot\rangle$ is G-invariant, if

$$\langle \rho(g)v|\rho(g)w\rangle = \langle v|w\rangle$$

for all $g \in G, v, w \in V$.

There are many times when we will request that a representation of a group G be unitary. The point of the following proposition is that it is very easy to acquire such a representation when working in finite dimensions. Given a non-invariant inner product, we can always manufacture an invariant one by taking averages.

Proposition 5.2.2 ([13])**.** Given a representation ρ of a compact group G on a finite-dimensional vector space V with inner product $\langle\cdot|\cdot\rangle$, then ρ is unitary with respect to the inner product

$$(u|v) = \int \langle \rho(g)u|\rho(g)v\rangle \, dg.$$

5.3 Peter-Weyl theorem

There are several Peter-Weyl theorems. We are interested in the formulation that allows us to decompose $\mathbb{C}G$-modules into a direct sum of simple modules, similar to how one can decompose $\mathbb{C}^d$ into a direct sum of copies of $\mathbb{C}$.

Theorem 5.3.1 ([5], Peter-Weyl)**.** Any continuous, unitary representation $\rho : G \to \mathcal{H}$ of a compact group G on a Hilbert space $\mathcal{H}$ is the direct sum of finite-dimensional irreducible representations.

This fact has a few immediate corollaries. I don't have a citation for these, but they immediate enough that I imagine that these or very similar statements have been used as easy exercises in courses on the subject.

Corollary 5.3.2. Let G be a compact group. Any continuous $\mathbb{C}G$-module which is $\mathbb{C}$-linearly homeomorphic to a Hilbert space is semisimple. In particular, all finitely-generated $\mathbb{C}G$-modules are semisimple.

Corollary 5.3.3. Let G be a compact group. If a continuous $\mathbb{C}G$-module is $\mathbb{C}$-linearly homeomorphic to a Hilbert space, then it cannot be simultaenously infinite-dimensional and simple.

This tells us that the only obstructions to the semisimplicity of $\mathbb{C}G$-modules are the topological pathologies of infinite-dimensional vector spaces. Requiring that any topological $\mathbb{C}G$-modules be $\mathbb{C}$-linearly homeomorphic to a Hilbert space is sufficient to brush these pathologies under the rug. We will only be directly concerned in this thesis with finite-dimensional representations, so avoiding these pathologies is advantageous.

5.4 Characters

Let G be a compact group, and $\rho : G \to \operatorname{Aut} V$ be a representation of G on a finite-dimensional inner product space V. We define the character of ρ to be the function

$$\chi_\rho : G \to \mathbb{C}$$

given by

$$\chi_\rho(g) = \operatorname{Tr} \rho(g).$$

We call a representation corresponding to a simple $\mathbb{C}G$-module irreducible, and we call the character of an irreducible representation irreducible too.

Note that $\chi_\rho(e) = \dim \mathcal{H}$. This is true because χ is a homomorphism, and the identity element of $\operatorname{Aut} \mathcal{H}$ may be written (in any basis) as a diagonal matrix with 1s along the diagonal. If we sum all of those 1s, we get the dimension of $\mathcal{H}$.

Finally, we will consider characters to be elements of the vector space of functions $G \to \mathbb{C}$. We will consider the subspace spanned by a group's characters to be a Hilbert space, according to the inner product

$$\langle \chi_\mu | \chi_\nu \rangle = \int_G \overline{\chi_\mu(g)} \chi_\nu(g) \, dg.$$

This is sesquilinear due to the linearity of integration. It is non-degenerate because the integrand is nonnegative, and because it is strictly positive in some open neighbourhood of the identity element $e \in G$.

I will now present some important calculational identities and facts about characters, without proof. The reader may consult [13] or [12] for more details on how this works for finite groups.

Proposition 5.4.1 ([12])**.** If g is conjugate to h in G, then

$$\chi_\rho(g) = \chi_\rho(h)$$

for any character χ_ρ of G.

Proposition 5.4.2 ([12])**.** If $\rho : G \to \operatorname{Aut} V$ is a unitary representation, then

$$\chi_\rho(g^{-1}) = \overline{\chi_\rho(g)}$$

for all $g \in G$. This is the character of the dual representation ρ^*, which corresponds to the dual module $\operatorname{Hom}(V, \mathbb{C})$.

Proposition 5.4.3 ([12])**.**

$$\chi_\mu + \chi_\nu = \chi_{\mu \oplus \nu}$$

for all characters χ_μ, χ_ν of G. Furthermore, $\mu \oplus \nu$ is unitary if μ and ν are.

Proposition 5.4.4 ([13])**.**

$$\chi_\mu \chi_\nu = \chi_{\mu \otimes \nu}$$

for all characters χ_μ, χ_ν of G. Furthermore, $\mu \otimes \nu$ is unitary if μ and ν are.

Proposition 5.4.5 ([12])**.** If μ and ν are irreducible unitary representations of a compact group G, then

$$\langle \chi_\mu | \chi_\nu \rangle = [\mu = \nu].$$

In other words, the irreducible characters of G are an orthonormal set within the vector space generated by characters.

In fact, since the sum of characters corresponds to the direct sum of representations, the irreducible characters form a $\mathbb{C}$-basis for the $\mathbb{C}$-vector space generated by characters of finitely-generated representations of G. Stronger yet, they are an $\mathbb{N}$-basis for the characters of finitely-generated representations of G, in the sense that the latter are nonnegative integral combinations of the former.

Lemma 5.4.6 ([38], Schur's lemma)**.** Let V and W be simple $\mathbb{C}G$-modules. Then for all $\mathbb{C}G$-module homomorphisms $f : V \to W$, either $f = 0$, or $V = W$ and $f = cI$ for some $c \in \mathbb{C}$.

Corollary 5.4.7 ([37])**.** Let λ, μ be irreducible representations $G \to \operatorname{Aut} \mathbb{C}^d$. If T is any matrix such that $T\lambda(g) = \mu(g)T$ for all $g \in G$, then either T is invertible, or $T = 0$.

Proof. The compositions $T\lambda$ and μT are representations of G, and in fact they are the same one. Since λ and μ are simple, $T\lambda$ and μT are either both zero, or they are equivalent to λ and μ, which are equivalent to each other. In that case, T is invertible. $\square$

Corollary 5.4.8 ([37])**.** Let $\rho : G \to \operatorname{Aut} V$ be a representation, and let $\mathcal{C}$ be its commutant in $\operatorname{End} V$. Then $\mathcal{C} = \{cI \mid c \in \mathbb{C}\}$ if and only if ρ is irreducible.

5.4.1 Operations on representations

Proposition 5.4.9 ([35] describes this character using the vocabulary of symmetric functions)**.** Let $\rho : G \to \operatorname{Aut} V$ be a unitary representation. Then the exterior power representation

$$\bigwedge^k \rho : G \to \operatorname{Aut} \bigwedge^k V$$

given by

$$\left(\bigwedge^k \rho \right)(g) = \bigwedge^k (\rho(g))$$

is a unitary representation, and its character, which we denote by $\bigwedge^k \chi_\rho$, is

$$\bigwedge^k \chi_\rho(g) = \sum_{S \in \binom{[\dim V]}{k}} [\rho(g)]_{S,S},$$

i.e., the sum of principal k-minors of $\rho(g)$.

Proof. Let $g \in G$. Then

$$\begin{aligned}\left(\bigwedge^k \rho(g)^*\right)_{S,T} &= \overline{\bigwedge^k \rho(g)_{T,S}} \\ &= \overline{[\rho(g)]_{T,S}} \\ &= [\rho(g)^*]_{S,T}\end{aligned}$$

so

$$\begin{aligned}\left(\bigwedge^k \rho(g) \bigwedge^k \rho(g)^*\right)_{S,T} &= \sum_{R \in \binom{[\dim V]}{k}} [\rho(g)]_{S,R}[\rho(g)^*]_{R,T} \\ &= [\rho(g)\rho(g)^*]_{S,T} \\ &= [I]_{S,T} \\ &= \delta_{S,T}\end{aligned}$$

by the Cauchy-Binet theorem 2.4.16. Thus

$$\bigwedge^k \rho(g) \bigwedge^k \rho(g)^* = I.$$

This implies $\bigwedge^k \rho$ is unitary.

Finally,

$$\begin{aligned}\bigwedge^k \chi_\rho(g) &= \operatorname{Tr} \bigwedge^k \rho(g) \\ &= \sum_{S,S \in \binom{[\dim V]}{k}} [\rho(g)]_{S,S}.\end{aligned}$$

□

Proposition 5.4.10. Let $\rho : G \to \operatorname{Aut} V$ be a unitary representation. Then the adjoint representation

$$\operatorname{Adj} \rho : G \to \operatorname{Aut} \operatorname{End} V$$

given by

$$\operatorname{Adj} \rho(g) : M \mapsto gMg^*$$

is a unitary representation, and its character, which we denote by $\operatorname{Adj} \chi_\rho$, is

$$\operatorname{Adj} \chi_\rho(g) = \sum_{i=1}^{\dim V} |g_{i,i}|^2.$$

(Note: this representation may also be seen as $\rho \otimes \rho^* : G \to \operatorname{Aut} V \otimes V^*$, but we will be using the concrete matrix perspective much more than we would have used the dual-module perspective.)

Proof. Let $n = \dim V$. Recall that, given some orthonormal basis for V, we can select a basis for $\operatorname{End} V$ in the form of the matrices $\{e_{i,j}\}_{i,j=1}^n$ which are nonzero only in their i, jth entry, where they bear a 1. Then $\langle A|B\rangle = \operatorname{Tr} A^*B$ gives us an inner product for $\operatorname{End} V$ under which this basis is orthonormal. Furthermore, $\operatorname{End} V$ has dimension n^2 as a vector space. Thus for all $g \in G$, $\operatorname{Adj} \rho(g)$ is a $n^2 \times n^2$ matrix in the general linear group $\operatorname{Aut} \operatorname{End} V$. We will denote by $(\operatorname{Adj} \rho(g))_{e_{i,j},e_{k,l}}$ the entry of $\operatorname{Adj} \rho(g)$ in the row corresponding to $e_{i,j}$ and the column corresponding to $e_{k,l}$. In other words, the entry $(\operatorname{Adj} \rho(g))_{e_{i,j},e_{k,l}}$ is equal to the inner product $\langle e_{k,l}| \operatorname{Adj} \rho(g)e_{i,j}\rangle$. We will now compute the value of this entry.

It is

$$\begin{aligned}
(\operatorname{Adj} \rho(g))_{e_{i,j},e_{k,l}} &= \langle e_{k,l}| \operatorname{Adj} \rho(g)e_{i,j}\rangle \\
&= \operatorname{Tr} e_{k,l}^* \rho(g)e_{i,j}\rho(g)^* \\
&= \operatorname{Tr} e_{l,k}\rho(g)e_{i,j}\rho(g)^* \\
&= \sum_{m=1}^n (e_{l,k}\rho(g)e_{i,j}\rho(g)^*)_{m,m} \\
&= \sum_{m=1}^n \sum_{p=1}^n (e_{l,k})_{m,p}(\rho(g)e_{i,j}\rho(g)^*)_{p,m} \\
&= (\rho(g)e_{i,j}\rho(g)^*)_{k,l} \\
&= \sum_{q=1}^n \rho(g)_{k,q}(e_{i,j}\rho(g)^*)_{q,l} \\
&= \sum_{q=1}^n \sum_{r=1}^n \rho(g)_{k,q}(e_{i,j})_{q,r}(\rho(g)^*)_{r,l} \\
&= \rho(g)_{k,i}(\rho(g)^*)_{j,l} \\
&= \rho(g)_{k,i}\overline{\rho(g)_{l,j}}.
\end{aligned}$$

We can then compute

$$
\begin{aligned}
((\operatorname{Adj}\rho(g))((\operatorname{Adj}\rho(g))^*))_{e_{i,j},e_{k,l}} &= \sum_{p,q=1}^{n} (\operatorname{Adj}\rho(g))_{e_{i,j}e_{p,q}}((\operatorname{Adj}\rho(g))^*)_{e_{p,q},e_{k,l}} \\
&= \sum_{p,q=1}^{n} (\rho(g)_{p,i}\overline{\rho(g)_{q,j}})\overline{\rho(g)_{p,k}\overline{\rho(g)_{q,l}}} \\
&= \sum_{p,q=1}^{n} \rho(g)_{p,i}\overline{\rho(g)_{q,j}}\rho(g)_{q,l}\overline{\rho(g)_{p,k}} \\
&= \left(\sum_{p=1}^{n}\overline{\rho(g)_{p,k}}\rho(g)_{p,i}\right)\left(\sum_{q=1}^{n}\overline{\rho(g)_{q,j}}\rho(g)_{q,l}\right) \\
&= (\rho(g)^*\rho(g))_{k,i}(\rho(g)^*\rho(g))_{j,l} \\
&= I_{k,i}I_{j,l} \\
&= [k=i][l=j] \\
&= [e_{i,j}=e_{k,l}].
\end{aligned}
$$

Thus $\operatorname{Adj}\rho$ is unitary. □

5.5 Exterior powers of the unitary group

In this section, we will outline some important representation-theoretic properties of the unitary group $\mathcal{U}(d)$. Most of the literature on this topic treats instead the *special unitary group*, which may be written as $\mathcal{U}(d)/\mathcal{U}(1)$. See for example [19].

Definition 5.5.1. Fix $d \geq 1$. For $1 \leq i < j \leq d$, denote by $\phi_i(\theta)$ the diagonal matrix in $\operatorname{End}\mathbb{C}^d$ with the property that $(\phi_i(\theta))_{i,i} = e^{i\theta}$, and such that all other diagonal entries are equal to 1.

Lemma 5.5.2 ([40] states this with different notation)**.** Each matrix $\phi_i(\theta)$ is unitary.

Proof. Notice

$$(\phi_i(\theta)\phi_i(\theta)^*)_{i,i} = e^{i\theta}e^{-i\theta} = 1.$$

It is straightforward to see from here that $\phi_i(\theta)\phi_i(\theta)^* = I$. □

Definition 5.5.3. Fix $d \geq 2$. Denote by $r_{i,j}(\theta)$ the matrix in $\operatorname{End}\mathbb{C}^d$ with the properties that

$$(r_{i,j}(\theta))_{\{i,j\},\{i,j\}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

and

$$r_{i,j}(\theta)_{[d]\setminus\{i,j\},[d]\setminus\{i,j\}} = I.$$

Lemma 5.5.4 ([40] states this with different notation)**.** Each matrix $r_{i,j}(\theta)$ is unitary.

Proof. Notice that

$$(r_{i,j}(\theta)r_{i,j}(\theta)^*)_{i,i} = (r_{i,j}(\theta)r_{i,j}(\theta)^*)_{j,j} = \cos^2\theta + \sin^2\theta = 1.$$

Also,

$$(r_{i,j}(\theta))_{i,j} = (r_{j,i}(\theta))_{j,i} = \cos\theta\sin\theta - \cos\theta\sin\theta = 0.$$

It is straightforward to see from here that $r_{i,j}(\theta)r_{i,j}(\theta)^* = I$. □

The authors of [17] cite the following as a "classical fact". While my deduction here is fairly elementary, I have been unable to find a specific reference for this fact. Most reference sources provide a full classification of the "nice" representations of the unitary group. Due to a strong inclusion/restriction duality between $\mathcal{U}(d)$ and $\mathrm{GL}(d)$, it is also often the case that reference sources treat only the general linear case. In order to avoid developing more theory than necessary, I have used elementary techniques to prove the following claim. It may be seen as a corollary of the general classification of polynomial $\mathrm{GL}(d)$-irreps developed in [22].

The following fact is explained in [19], in relation to the closely related *special unitary group*, using the vocabulary of "highest weight theory" and Young tableaux. But I first learned it from [17], which cites an exercise in [13] as its source.

Proposition 5.5.5 ([13] as an exercise, [19] in different vocabulary)**.** Let $\rho : \mathcal{U}(d) \to \operatorname{Aut}\mathbb{C}^d$ be the tautological representation $g \mapsto g$. Then the exterior powers $\bigwedge^k \rho$ for $0 \le k \le d$ form a family of distinct irreducible representations.

Proof. We can prove irreducibility by showing that the commutant of $\bigwedge^k \rho$ in $\operatorname{End}\bigwedge^k \mathbb{C}^d$ is the set of scalar matrices cI for $c \in \mathbb{C}$. Then by Schur's lemma, $\bigwedge^k \rho$ must be irreducible. I will put this part off until the end of the proof. For now, I will work to show that the representations are inequivalent. One invariant of a representation is its dimension. The exterior powers come in equidimensional pairs $(\bigwedge^0 \rho, \bigwedge^d \rho), (\bigwedge^1 \rho, \bigwedge^{d-1} \rho), \ldots$, with the possible exception of the power $\bigwedge^{d/2} \rho$ when d is even. We need to find an invariant which distinguishes the equidimensional exterior powers. I will use the representations' characters to distinguish them.

Let $k \ge 0$ be an integer, and make the assumption that if d is even then $k \ne d/2$. Denote by χ the character of $\bigwedge^k \rho$, and denote by κ the character of $\bigwedge^{d-k} \rho$. These are given by

$$\chi(g) = \sum_{T \subseteq \binom{[d]}{k}} \left(\bigwedge^k \rho(g) \right)_{T,T} = \sum_{T \subseteq \binom{[d]}{k}} [g]_T$$

and

$$\kappa(g) = \sum_{T \subseteq \binom{[d]}{d-k}} \left(\bigwedge^{d-k} \rho(g) \right)_{T,T} = \sum_{T \subseteq \binom{[d]}{d-k}} [g]_T.$$

Now, $\bigwedge^k \phi_l(\theta)$ is a diagonal unitary matrix, where

$$\left(\bigwedge^k \phi_l(\theta) \right)_{U,U} = \begin{cases} e^{i\theta} & l \in U \\ 1 & \text{otherwise} \end{cases}.$$

There are $\binom{d-1}{k-1}$ entries of the diagonal which are equal to $e^{i\theta}$, and $\binom{d-1}{k}$ entries which are equal to 1. Conversely, there are $\binom{d-1}{d-k-1}$ entries of the diagonal of $\bigwedge^{d-k} \phi_k(\theta)$ which are equal to $e^{i\theta}$ and $\binom{d-1}{d-k}$ entries which are equal to 1.

Thus

$$\chi(\phi_l(\theta)) = \binom{d-1}{k-1} e^{i\theta} + \binom{d-1}{k}$$

and

$$\kappa(\phi_l(\theta)) = \binom{d-1}{d-k-1} e^{i\theta} + \binom{d-1}{d-k}$$

Suppose we set $\theta = \frac{\pi}{2}$. Then

$$\Re \chi(\phi_l(\theta)) = \binom{d-1}{k}$$

and

$$\Re \kappa(\phi_l(\theta)) = \binom{d-1}{d-k} = \binom{d-1}{k-1}.$$

These two expressions will never be equal, except in the case that $d - k = k$. But this possibility is ruled out by the hypothesis that $k \neq d/2$. Thus there exists $\theta \in \mathbb{R}$ for which

$$\chi(\phi_l(\theta)) \neq \kappa(\phi_l(\theta)),$$

and thus the representations $\bigwedge^k \rho$ and $\bigwedge^{d-k} \rho$ are inequivalent. We can conclude that the exterior powers of ρ are all distinct representations.

Now I will show that these representations are all irreducible. Let $M \in \operatorname{End} \bigwedge^k \mathbb{C}^d \cong \operatorname{Mat}_{\binom{d}{k}}(\mathbb{C})$, and suppose that

$$\left(\bigwedge^k g \right) M \bigwedge^k g^* = M$$

for all $g \in \mathcal{U}(d)$. First, let's try to establish that $M_{S,T} = 0$ for $S \neq T$, $S, T \in \binom{[d]}{k}$.

Assume that $l \in S, l \notin T$. Then

$$\begin{aligned}\left(\left(\bigwedge^k \phi_l(\theta)\right) M \bigwedge^k \phi_l(-\theta)\right)_{S,T} &= \sum_{U,V \in \binom{[d]}{k}} \left(\bigwedge^k \phi_l(\theta)\right)_{S,U} M_{U,V} \left(\bigwedge^k \phi_l(-\theta)\right)_{V,T} \\ &= \left(\bigwedge^k \phi_l(\theta)\right)_{S,S} M_{S,T} \left(\bigwedge^k \phi_l(-\theta)\right)_{T,T} \\ &= e^{i\theta} M_{S,T},\end{aligned}$$

Thus the only condition under which M is in the commutant of $\bigwedge^k \rho$ is if all off-diagonal entries $M_{S,T}$ are equal to zero.

We have established that the commutant of $\bigwedge^k \rho$ contains only diagonal matrices. Now, we will use the elementary rotation matrices to show that these diagonal entries must all be equal, i.e., that the commutant comprises the scalar matrices. Let S, T again be distinct elements of $\binom{[d]}{k}$. We will try to establish that $M_{S,S} = M_{T,T}$. Let's now consider the relations forced by the unitary matrix $r_{i,j}(\theta)$, where $i \in S \setminus T$, $j \in T \setminus S$.

$$\begin{aligned}\left(\left(\bigwedge^k r(\theta)\right) M \bigwedge^k r(-\theta)\right)_{S,S} &= \sum_{U,V \in \binom{[d]}{k}} \left(\bigwedge^k r(\theta)\right)_{S,U} M_{U,V} \left(\bigwedge^k r(-\theta)\right)_{V,S} \\ &= \left(\bigwedge^k r(\theta)\right)_{S,S} M_{S,S} \left(\bigwedge^k r(-\theta)\right)_{S,S} \\ &+ \left(\bigwedge^k r(\theta)\right)_{S,T} M_{T,S} \left(\bigwedge^k r(-\theta)\right)_{S,S} \\ &+ \left(\bigwedge^k r(\theta)\right)_{S,S} M_{S,T} \left(\bigwedge^k r(-\theta)\right)_{T,S} \\ &+ \left(\bigwedge^k r(\theta)\right)_{S,T} M_{T,T} \left(\bigwedge^k r(-\theta)\right)_{T,S} \\ &= M_{S,S} \cos^2\theta + M_{T,T} \sin^2\theta,\end{aligned}$$

since M is already known to be a diagonal matrix. Setting the derivative of this expression with respect to θ to zero, we find that this matrix entry is invariant for all θ only if

$$2M_{S,S} \sin\theta\cos\theta = 2M_{T,T} \sin\theta\cos\theta,$$

i.e., only if $M_{S,S} = M_{T,T}$. Since T and S were arbitrary, we know that all diagonal entries of M must be equal to each other for M to be in the commutant of $\bigwedge^k \rho$. But the matrices with this property are just the scalar matrices, which commute with the whole matrix algebra. Therefore, the exterior powers of $\bigwedge^k \rho$ are irreducible. $\square$

Chapter 6

Weingarten calculus

6.1 The fundamental theorem

The exposition in this section is based off of the excellent pedagogical resource [9].

Suppose that d is a positive integer, G is a compact Hausdorff group and $(\rho, \mathcal{H})$ is a continuous, unitary representation of that group over a d-dimensional complex vector space $\mathcal{H}$. After choosing an orthonormal basis $\{e_i\}_{i=1}^d$ for $\mathcal{H}$, we may interpret the images $\rho(g)$ of elements $g \in G$ as being $d \times d$ matrices, with entries $\rho_{i,j}(g) = \langle e_i | \rho(g) e_j \rangle$. Various practical applications involve the solutions of integrals of the form $\int \prod_{l=1}^n \rho_{i_l, j_l}(g)\, dg$, for g Haar-distributed on G. Solving integrals of this form is the *raison d'être* of the Weingarten calculus. In fact, the early precursors of this technique consisted of *ad hoc* identities proved by quantum physicists for $G = \mathcal{U}(d)$, the group of unitary matrices.

Now let's try to develop a systematic approach to this problem. In the product just given, we will think of i and j as functions $[n] \to [d]$, and we will see our integral as a quantity

$$I_{i,j} = \int_G \prod_{l=1}^n \rho_{i(l),j(l)}(g)\, dg,$$

called a Weingarten integral. For each $g \in G$, the product inside the integral is the same as the entry $\rho_{i,j}^{\otimes n}(g)$ of the matrix $\rho^{\otimes n}(g)$. So we will rewrite

$$I_{i,j} = \int_G \rho_{i,j}^{\otimes n}(g)\, dg.$$

Thus if we can tabulate all the entries of the matrix $P \in \operatorname{End} \mathcal{H}^{\otimes n}$ given by

$$P = \int_G \rho^{\otimes n}(g)\, dg,$$

we will have written down the solutions of all Weingarten integrals for products of length n.

This matrix P has some special properties, which we will use to determine its entries.

First, by the compactness of G and the translation-invariance of the Haar measure,

$$\begin{aligned} P^2 &= \int_G \rho^{\otimes n}(g)\, dg \int_G \rho^{\otimes n}(g)\, dg \\ &= \int_G \int_G \rho^{\otimes n}(g)\rho^{\otimes n}(h)\, dgdh \\ &= \int_G \int_G \rho^{\otimes n}(gh)\, dgdh \\ &= \int_G \int_G \rho^{\otimes n}(g)\, dgdh \\ &= \int_G \rho^{\otimes n}(g)\, dg \\ &= P, \end{aligned}$$

so P is idempotent.

Second, since ρ is a unitary representation, and since the Haar measure is invariant under the involution $g \mapsto g^{-1}$,

$$P^* = \int_G \rho^{\otimes n}(g)^*\, dg = \int_G \rho^{\otimes n}(g)^{-1}\, dg = \int_G \rho^{\otimes n}(g^{-1})\, dg = P,$$

so P is selfadjoint. A selfadjoint idempotent such as P must be an orthogonal projection by 2.3.5.

Now, for all $v \in \mathcal{H}, g \in G$, we have $\rho^{\otimes n}(g)Pv = Pv$ by the translation-invariance of the Haar measure. And if $\rho^{\otimes n}(g)v = v$ for all $g \in G$, then

$$Pv = \int_G \rho^{\otimes n}(g)\, dg\, v = \int_G \rho^{\otimes n}(g)v\, dg = v,$$

so $v \in \operatorname{im} P$. Therefore, the image of P is the set $(\mathcal{H}^{\otimes n})^G$ of all $v \in \mathcal{H}^{\otimes n}$ which are invariant under the representation $\rho^{\otimes n}$.

Since P is an orthogonal projection onto the subspace $(\mathcal{H}^{\otimes n})^G$, the problem of finding the entries of P is closely related to the problem of finding a basis for the trivial $\mathbb{C}G$-submodules of $\mathcal{H}^{\otimes n}$.

Let A be a $d^n \times m$ matrix whose columns are formed by the elements of some basis of the m-dimensional subspace $(\mathcal{H}^{\otimes n})^G \subseteq \mathcal{H}^{\otimes n}$. Then we have the following well-known matrix factorization, stated without proof in [9]. (The authors say it is a result "familiar from matrix analysis".)

Lemma 6.1.1 (Stated in [9], proof outlined to me by my supervisor Paul Skoufranis, also proven in [41])**.**

$$P = A(A^*A)^{-1}A^*.$$

Proof. We know $A = QR$ where R is an upper triangular matrix, and Q is a unitary matrix whose columns form an orthonormal basis of $\operatorname{im} P$. Since A has linearly independent columns, R is nonsingular, and thus

$$\begin{aligned} P &= QQ^* \\ &= (AR^{-1})((R^*)^{-1}A^*) \\ &= A(R^*R)^{-1}A^* \end{aligned}$$

Since $R^*R = R^*Q^*QR = A^*A$, we have

$$A(A^*A)^{-1}A^*.$$

□

We will now name the constitutent parts of $P = A(A^*A)^{-1}A^*$.

Definition 6.1.2. We call the $m \times m$ matrix A^*A the Gram matrix of $\rho^{\otimes n}$ with respect to the chosen bases of $\mathcal{H}^{\otimes n}$ and $(\mathcal{H}^{\otimes n})^G$.

Definition 6.1.3. We call the inverse of the Gram matrix the Weingarten matrix of $\rho^{\otimes n}$ with respect to the chosen bases of $\mathcal{H}^{\otimes n}$ and $(\mathcal{H}^{\otimes n})^G$. We denote this matrix by W.

We can thus write

$$P = AWA^*.$$

In summary, we have the following.

Theorem 6.1.4 (Fundamental theorem of Weingarten calculus, [9])**.** Let d be a positive integer, and let $(\rho, \mathcal{H})$ be a continuous, unitary, d-dimensional representation of a compact Hausdorff group G over a Hilbert space $\mathcal{H}$, equipped with an orthonormal basis $\{e_i\}_{i=1}^d$. Suppose that $n \geq 1$ and i, j are functions $[n] \to [d]$. Let A be the $d^n \times m$ matrix whose columns are formed by some basis of the subspace $(\mathcal{H}^{\otimes n})^G$, written in terms of the basis $\{e_{i_1} \otimes \ldots \otimes e_{i_n}\}_{i:[n]\to[d]}$ of $\mathcal{H}^{\otimes n}$. Let W be the Weingarten matrix with respect to those bases. Then the Weingarten integral

$$I_{i,j} = \int_G \prod_{l=1}^n U_{i(l),j(l)}(g)\, dg$$

is equal to

$$(AWA^*)_{i,j} = \sum_{r,s=1}^m A_{i,r} W_{r,s} A^*_{s,j}.$$

On a less formal note, we can say that the task of computing the integrals $I_{i,j}$ reduces to the task of finding any set of vectors S whose span is the G-invariant subspace of $\mathcal{H}^{\otimes n}$. Once such an S has been found, we can reduce down to a basis using the Gram-Schmidt process, and then the fundamental theorem gives us an arithmetic procedure for computing $I_{i,j}$.

Chapter 7

Finite free probability

Finite free probability was first developed by Marcus, Spielman and Srivastava in a 2015 paper entitled "Finite free convolutions of polynomials". This effort was a development of some ideas that had developed during their study of interlacing families of polynomials, through which they had settled a famous open problem in analysis. That problem, the Kadison-Singer problem, was initially expressed in terms of infinite-dimensional operator algebras, but the solution that Marcus, Spielman and Srivastava developed emerged out of the study of matrices (i.e., the study of finite-dimensional operator algebras).

Finite free probability is thus *finite* in the sense that it is concerned primarily with the behaviour of matrices. The notion of *probability* is involved because one of the key operations in finite free probability is calculating averages over random matrices. The reasoning behind the moniker *free* is a little more sophisticated.

7.1 Free probability

Free probability is a noncommutative probability theory, i.e., it is a probability theory that concerns itself with *random variables* for which the product XY may not be identical to YX. This is unthinkable when one formalizes random variables (as we have) as measurable functions $\Omega \to \mathbb{R}$, so free probability requires a completely different foundation than the conventional Kolmogorov axioms of probability theory. Most of the material in this overview is adapted from [42] or [31]. Here we will take an informal overview, just enough to describe the connection between free probability and finite free probability.

Definition 7.1.1. A random matrix ensemble is a family of random matrices $\{M_d\}$, indexed by dimension.

Definition 7.1.2. For a $d \times d$ matrix X, we define

$$\operatorname{tr} X = \frac{1}{d} \operatorname{Tr} X.$$

It turns out that certain ensembles of random matrices, in the limit as their dimension tends towards infinity, begin to behave like random variables with respect to the expectation

$$\mathbb{E}[X] = \operatorname{tr} X.$$

In particular, the matrix moments

$$\mathbb{E}[X^k] = \operatorname{tr} X^k$$

end up playing a role analogous to that of the moments of conventional probability theory,

$$\mathbb{E}[X^k] = \int_{\mathbb{R}} t^k d\mu_X(t).$$

We even associate to X the measure on $\mathbb{R}$ corresponding to X's moment sequence $(\operatorname{tr} X^k)_{k=0}^{\infty}$, supposing one exists.

The basic ingredients required to "do free probability" are a $\mathbb{C}$-algebra $\mathcal{A}$ of *noncommuting random variables* (*n.c.r.v.*'s) and a $\mathbb{C}$-linear operator $\phi : \mathcal{A} \to \mathbb{C}$ which we call the *expectation operator*. We call $\mathcal{A}$ and ϕ together a *noncommutative probability space*, and we call elements of that space *noncommutative random variables*. We will give an example in terms of group theory. First, we will need to define an important construction on groups.

Definition 7.1.3. The free product of groups G and H, denoted $G * H$, may be defined as the set of words over the alphabet $G \cup H$, under the equivalence relation

$$e_G \sim e_H,$$

and with the equivalence relations

$$g_1 \star g_2 \sim g_1 g_2 \forall g_1, g_2 \in G$$

and

$$h_1 \star h_2 \sim h_1 h_2 \forall h_1, h_2 \in H.$$

This is a group under the operation of concatenation of words.

Notice in particular that the non-identity elements of G and H never commute with each other, even if G and H are abelian, and reduced words can be written in alternating form

$$g_1 h_1 g_2 h_2 ... g_n h_n.$$

Then one can take the group algebra $\mathcal{A} = \mathbb{C}(G * H)$, and the expectation operator $\phi : \mathcal{A} \to \mathbb{C}$ which gives us, for any formal sum of group elements, the coefficient of the identity element in that sum. In symbols,

$$\phi\left(\sum_{g \in G} z_g g\right) = z_e,$$

where e is the group identity. Then something fascinating arises. Any alternating product $a_1 b_1 a_2 b_2 ... a_k b_k$, where $a_i \in \langle G \rangle \subset \mathcal{A}$ for all i, and $b_i \in \langle H \rangle \subset \mathcal{A}$ for all i, has the property that if $\phi(a_i) = 0$ and $\phi(b_i) = 0$ for all i, then

$$\phi(a_1 b_1 a_2 b_2 ... a_k b_k) = 0.$$

The reason for this identity is that generators belonging to each of the groups H and G cannot commute with each other, by the definition of the free product of groups, and so no term in formal sums $a_i = \sum_{g \in G} z_g^i g$ or $b_j = \sum_{h \in H} w_h^j h$ ever finds itself multiplied on the immediate left or right by its own inverse.

It may not be immediately obvious, but this property is deeply analogous to the traditional concept of *independence* in probability theory, and we call any two subalgebras of a noncommutative probability space satisfying this property *freely independent* with respect to each other. Furthermore, we call noncommutative random variables a and b freely independent if they generate freely independent subalgebras.

Given the previous description of freely independent noncommutative random variables in the group algebra $\mathbb{C}(G * H)$, one might wonder whether anything analogous is possible in the setting of random matrices. One conventional approach described in [31] uses random matrix ensembles. We may consider a random $d \times d$ matrix as representing a distribution whose sample space consists of probability mass evenly distributed among the eigenvalues of the matrix. For example, a matrix with eigenvalues $\{1, i, 2, 2, 3 - i\}$ (counted with multiplicity) would represent the distribution with the following probability table:

$$\begin{aligned} \mathbb{P}[X = 1] &= 20\% \\ \mathbb{P}[X = i] &= 20\% \\ \mathbb{P}[X = 2] &= 40\% \\ \mathbb{P}[X = 3 - i] &= 20\%. \end{aligned}$$

The sequence of probability distributions corresponding to the matrices in the ensemble is a sequence of discrete probability distributions, which may converge weakly towards some particular probability distribution on the complex plane. In this way, we can associate (well-behaved) random matrix ensembles with probability distributions, and so regard a matrix ensemble $(X_n)_{n=1}^{\infty}$ as a n.c.r.v., using the expectation operator

$$\mathbb{E}[(X_n)_{n=1}^{\infty}] = \lim_{n \to \infty} \operatorname{tr} X_n.$$

One might ask whether free independence arises in this setting, and in what form. Suppose that we had two matrices, A and B, but they have been given to us in two potentially distinct orthonormal coordinate systems. Without loss of generality, we could take the coordinate system used for A as our reference point, and write B as UBU^* for some free variable U taking values in the unitary group $\mathcal{U}(d)$. If we truly have no knowledge of the difference U between the two coordinate systems, we could represent our ignorance probabilistically by making U a *random* variable with Haar distribution.

Given this setup, it turns out that in many cases involving ensembles A and B, the limit of an ensemble A of random matrices is freely independent from the limit of the ensemble UBU^*, where U is an ensemble of Haar unitary matrices, and B is another ensemble that is not necessarily independent, even in the traditional sense, from A. When this occurs, the measure corresponding to $A + UBU^*$ may be computed from the measures corresponding to A and B, an operation known as *free convolution.* A power series transform known as the $\mathcal{R}$-transform often provides the easiest way to perform this operation. Furthermore, it is often easier to compute these free convolutions in the limit as $d \to \infty$ than to understand the phenomena at a fixed finite dimension d, so in some applications, free probability is used to approximate the behaviour of high, but finite-dimensional phenomena.

This is where finite free probability picks up. Rather than focus on the limit of the ensemble $A+UBU^*$ as its dimension d tends towards infinity, finite free probability provides a lens onto the distribution of its eigenvalues at fixed d by considering the distribution of its characteristic polynomial. These two perspectives are in fact describing the same phenomena, as Marcus showed that the $\mathcal{R}$-transform of a random matrix ensemble's eigenvalue distribution may be recovered via certain transforms of their characteristic polynomials.

Proposition 7.1.4 ([28])**.** Let A and B be fixed matrices, and let U be a random Haar-distributed unitary. Then the distribution of $\chi(A + UBU^*)$ is uniquely determined by its mixed *-moments.

Proof. The function $f : \mathrm{Mat}_d(\mathbb{C}) \to \mathbb{C}^{d+1}, X \mapsto \chi(A + XBX)^*$ is a continuous function of matrices. Thus for a Haar-distributed unitary U, the distribution of $A + UBU^*$ is compactly supported. The statement follows from 4.5.3. $\square$

We will be working with the expression $A+QBQ^*$ for random matrix Q and deterministic A and B. In particular, we will want Q to be Haar-distibuted over some compact subgroup $\mathcal{G} \subseteq \mathcal{U}(d)$. For any subgroup $\mathcal{G}$ of $\mathrm{GL}(d)$, we say that the matrices A and B are $\mathcal{G}$-similar if there exists $G \in \mathcal{G}$ for which $A = GBG^{-1}$. Write this as $A \sim_{\mathcal{G}} B$. We say that A is $\mathcal{G}$-diagonal if there exists a diagonal B such that $A \sim_{\mathcal{G}} B$.

Lemma 7.1.5 ([28])**.** If $\mathcal{G}$ is a compact subgroup of $\mathcal{U}(d)$, if Q is a Haar-distributed random element of $\mathcal{G}$, and if A and B are $\mathcal{G}$-similar to matrices A' and B', respectively, then we have the equality of characteristic polynomials

$$\chi(A + QBQ^*) = \chi(A' + QB'Q^*).$$

Proof. Let $G, H \in \mathcal{G}$ such that $GAG^* = A'$ and $HBH^* = B'$. Then, beginning via the unitary invariance of the characteristic polynomial,

$$\begin{aligned}\chi(A + QBQ^*) &= \chi(G(A + QBQ^*)G^*)\\ &= \chi(GAG^* + GQBQ^*G^*)\\ &= \chi(A' + QBQ^*).\end{aligned}$$

The final equality relies on the translation-invariance of Q. By the same logic,

$$\begin{aligned}\chi(A + QBQ^*) &= \chi(A' + QH^*HBH^*HQ^*)\\ &= \chi(A' + QH^*B'HQ^*)\\ &= \chi(A' + QB'Q^*)\end{aligned}$$

□

Definition 7.1.6. Call a set of $d \times d$ matrices *permuting* if it contains all of the permutation matrices.

I haven't seen the following claim written down anywhere, but that's just because it's an application of the very well-understood group actions of $\mathfrak{S}_n$ to a specific case.

Lemma 7.1.7. Suppose B is a diagonal $d \times d$ matrix with diagonal entries $(b_i)_{i=1}^d$, and $\mathcal{G}$ is a permuting subgroup of $\mathcal{U}(d)$. Then B is $\mathcal{G}$-similar to any other diagonal matrix sharing the same diagonal entries, counted with multiplicity.

Proof. Let B be a diagonal matrix as in the theorem statement. For any two indices $i, j \in [d]$, we can construct a permutation matrix Σ which acts from the left to swap the ith and jth standard basis vectors in $\mathbb{C}^d$. This action is its own inverse, so it expresses an action of $\mathbb{Z}/2\mathbb{Z}$ on $\mathbb{C}^d$. It is also unitary. The previous facts imply $\Sigma^* = \Sigma^{-1} = \Sigma$, so Σ is selfadjoint too.

Elementary manipulations show left multiplication by Σ swaps the ith and jth rows of B, and right multiplication by Σ swaps the ith and jth columns of B, so the action $B \mapsto \Sigma B \Sigma$ swaps the diagonal entries b_i and b_j. It is a standard fact (see [12]) that the symmetric group $\mathfrak{S}_d$ is generated by its transpositions, so any permutation matrix Σ can be decomposed such that $\Sigma B \Sigma^*$ is the result of swapping a sequence of pairs of diagonal entries of B. Since these permutation matrices belong in $\mathcal{G}$, B is $\mathcal{G}$-similar to each of these matrices $\Sigma B \Sigma^*$. □

The previous lemmas culminate in the following proposition. This result is likely familiar to practitioners of free probability, but I haven't seen it written down as a discrete fact.

Proposition 7.1.8. Suppose $\mathcal{G}$ is a compact, permuting subgroup of $\mathcal{U}(d)$, and that A and B are $\mathcal{G}$-diagonal. Then $A \mapsto \chi(A+QBQ^*)$ is a symmetric function of the eigenvalues of A, and $B \mapsto \chi(A + QBQ^*)$ is a symmetric function of the eigenvalues of B.

We can also develop the following, which again is likely familiar to practitioners of free probability.

Proposition 7.1.9. Suppose $\mathcal{G}$ is a compact subgroup of $\mathcal{U}(d)$. Then $(A, B) \mapsto \chi(A + QBQ^*)$, as a function from matrices to distributions, is a symmetric function of two variables.

Proof. Since the characteristic polynomial is unitarily invariant,

$$\begin{aligned}\chi(A+QBQ^*) &= \chi(Q^*(A+QBQ^*)Q)\\ &= \chi(Q^*AQ+Q^*QBQ^*Q)\\ &= \chi(Q^*AQ+B),\end{aligned}$$

which is equal in distribution to

$$\chi(B+QAQ^*)$$

via the Harr equality of measures property, $\mu_Q\mu_{Q^{-1}} = \mu_{Q^*}$. □

7.2 Minor-orthogonality

The unitary group shares a property known as minor-orthogonality with many of its compact subgroups. The concept was first defined in [27].

Definition 7.2.1. We call a random $d \times d$ matrix R minor-orthogonal if

$$\mathbb{E}[[R]_{S,T}[R^*]_{U,V}] = \frac{1}{\binom{d}{k}}[S=V][T=U]$$

for all $k \in [d+1], S, T, U, V \in \binom{[d]}{k}$.

Why does this property deserve this name? The nomenclature derives from the fact that $(X,Y) \mapsto \mathbb{E}[X\overline{Y}]$ is an inner product on complex random variables, inducing the norm

$$\|X\| = \int_{\mathbb{C}} |t|^2 d\mu_X(t).$$

Thus minor-orthogonality means that the matrix minors of R are pairwise orthogonal in this particular inner product space; furthermore, the norm of an individual minor must be $1/\binom{d}{k}$.

The following definition will be useful for proving representations are minor-orthogonal.

Definition 7.2.2 ([17])**.** We say that a random matrix Y has the *exterior power property* if it is equal in distribution to some $\rho(g)$, where g is a Haar-distributed random matrix over a compact group of matrices G, where ρ is a unitary and continuous representation, and where the exterior powers $\bigwedge^k \rho$ are distinct and irreducible G-representations.

The reasons why various random matrices are minor-orthogonal were proven over time by several different group. Marcus, Spielman and Srivastava [27] proved minor-orthogonality for the unitary group and the group of *signed permutation matrices* isomorphic to $S_n \ltimes \mathbb{Z}_2$. Their proofs were based on elementary (yet non-obvious) manipulations of sums. Hall, Puder and Sawin [17] later developed a representation-theoretic approach which applies to all representations possessing the exterior power property defined above, without explicitly using the language of Weingarten calculus. Later, Campbell and Yin [6] developed proofs for the unitary, orthogonal and hyperoctahedral cases (the last of which subsumes the case of signed permutation matrices). In their proofs, they used formulae developed using Weingarten calculus, including the one that will be developed in chapter 8 for the unitary group.

However, the formulae Campbell and Yin apply are very powerful, and commensurately complex. To apply their approach to new groups, one must first derive this very general formula, and then specialize it back again to the very special case of checking for minor-orthogonality. It seems more parsimonious to me to check whether a given group has the exterior power property than to first derive a general Weingarten formula for it, and then to perform ad-hoc algebraic manipulations over that formula. In what follows, I will attempt to synthesize various aspects of all the above authors' proofs. In particular, I will aim to emulate the generality of Hall, Puder and Sawin's approach, while using the conventional language of Weingarten calculus, following Campbell and Yin.

Theorem 7.2.3 ([17])**.** Let Y be a random matrix with the exterior power property. Then Y is minor-orthogonal.

Proof. By hypothesis, Y is equal in distribution to $\rho(g)$ for some Haar-distributed random matrix g over some compact group of matrices G, and for some continuous, unitary representation $\rho : G \to \mathbb{C}^d$. Furthermore, $\bigwedge^k \rho$ is irreducible for all $0 \leq k \leq d$. We will, without loss of generality, dispense with Y and work to prove that $\rho(g)$ is minor-orthogonal. (We have no need here of representing multiple random matrices with their dependence/independence properties, and focusing on the representation is helpful for the proof.) Finally, we will equivocate between g as a Haar-distributed random matrix over G and g as a particular element of G.

We will use Weingarten calculus to establish the conclusion of this theorem. The representation we want is $\operatorname{Adj} \bigwedge \rho : G \to \operatorname{Aut} \operatorname{End} \bigwedge \mathbb{C}^d$, where $\operatorname{Adj} \bigwedge \rho = \operatorname{Adj} \bigoplus_{k=0}^{d} \bigwedge^k \rho$ is the adjoint of the direct sum of all of the kth exterior powers of ρ. Its entries take the form

$$\begin{aligned}(\operatorname{Adj} \textstyle\bigwedge \rho)_{(S,T),(U,V)} &= [\rho(g)]_{U,S}\overline{[\rho(g)]_{V,T}} \\ &= [\rho(g)]_{U,S}[\rho(g)^*]_{T,V}\end{aligned}$$

for various $S, T, U, V \subseteq [d]$, where $|U| = |S|$ and $|T| = |V|$. (See 5.4.1 for more details on these constructions over ρ.) We want these entries because they are the expressions over which we must integrate to prove $\rho(g)$ is minor-orthogonal.

We want to know the invariant subspace $(\operatorname{End} \bigwedge \mathbb{C}^d)^G$ with respect to this representation, i.e., the set of matrices $M \in \operatorname{End} \bigwedge \mathbb{C}^d$ for which

$$\bigwedge \rho(g) M \bigwedge \rho(g)^* = M$$

for all $g \in G$. This is the same as the set of matrices M which commute with $\bigwedge \rho(g)$ for all $g \in G$. Since $\bigwedge \rho$ is a direct sum of irreducible representations, by Schur's lemma, any matrix M which commutes with its image is a block-scalar matrix. More specifically, for each component $\bigwedge^k \mathbb{C}^d$ of $\bigwedge \mathbb{C}^d$, the matrices which commute with $\bigwedge^k \rho$ are the scalar matrices in $\operatorname{End} \bigwedge^k \mathbb{C}^d$. Because $\dim \bigwedge^k \mathbb{C}^d = \binom{d}{k}$, each such scalar matrix is $cI_{\binom{d}{k}}$, for some $c \in \mathbb{C}$. Recombining these components, we get

$$M = \bigoplus_{k=0}^{d} c_k I_{\binom{d}{k}}$$

for some sequence of complex numbers $(c_k)_{k=0}^d$. In other words, $\{I_{\binom{d}{k}}\}_{k=0}^d$ forms a basis for the invariant subspace $(\operatorname{End} \bigwedge \mathbb{C}^d)^G$.

We will index the standard basis of $\bigwedge \mathbb{C}^d$ with subsets $S \subseteq [d]$, acquiring a basis $\{e_S\}_{S \subseteq [d]}$. Building off of this basis, we can consider the matrix algebra $\operatorname{End} \bigwedge \mathbb{C}^d$ to be generated by the basis $\{\mathfrak{e}_{S,T}\}_{S,T \subseteq [d]}$, where $\mathfrak{e}_{S,T}$ is the $\binom{d}{k} \times \binom{d}{k}$ matrix possessing a 1 in the S,Tth entry, and 0s everywhere else. In this basis, a matrix $I_{\binom{d}{k}} \in \operatorname{End} \bigwedge \mathbb{C}^d$ may be expressed as

$$I_{\binom{d}{k}} = \sum_{S \in \binom{[d]}{k}} \mathfrak{e}_{S,S}.$$

Thus the invariant subspace $(\operatorname{End} \bigwedge \mathbb{C}^d)^G$ has a basis of cardinality $d+1$, accounting for each $I_{\binom{d}{k}}$ for $0 \leq k \leq d$.

We will want to assemble these basis vectors into a matrix A in order to compute the Weingarten matrix with respect to this basis. But we have gotten a bit ahead of ourselves–we have specified a basis for our finite-dimensional vector space, but we have not explained its ordering. We will do so now. The basis $\{e_S\}_{S \subseteq [d]}$ of $\bigwedge \mathbb{C}^d$ has a straightforward graded-lex ordering, where we first order the sets $S \subseteq [d]$ into classes by cardinality, then order the elements of each cass lexicographically. For example, if $d = 3$, we have the ordering

$$\emptyset < \{1\} < \{2\} < \{3\} < \{1,2\} < \{1,3\} < \{2,3\} < \{1,2,3\}.$$

To order the $\{\mathfrak{e}_{S,T}\}_{S,T \subseteq [d]}$, we can first assume the graded-lex ordering on the subsets of $[d]$, then add another layer of lexicographic ordering on the tuples $(S,T) \in \mathcal{P}([d])^2$. For instance, at $d = 2$, we would order the tuples like so:

$$(\emptyset, \emptyset) < (\emptyset, \{1\}) < (\emptyset, \{2\}) < (\emptyset, \{1,2\}) < (\{1\}, \emptyset) < (\{1\}, \{1\}) < \ldots$$

This resembles counting from 0 to $2^{2d} - 1$ in base 2^d. We will apply this ordering to the basis $\{\mathfrak{e}_{S,T}\}$.

Now that we know this, let's try writing down the matrix A, listing the basis of $(\operatorname{End} \bigwedge \mathbb{C}^d)^G$. Because the first column corresponds to $I_{\binom{d}{0}}$, the second column corresponds to $I_{\binom{d}{1}}$, and so on, we will index the columns of A by the integers 0 through d. Observe

$$I_{\binom{d}{0}} = \sum_{S \in \binom{[d]}{0}} \mathfrak{e}_{S,S} = \mathfrak{e}_{\emptyset,\emptyset}.$$

In our chosen basis, this vector has a 1 in its first component, and 0s in its remaining $2^{2d}-1$ components.

The second column corresponds to the basis element

$$I_{\binom{d}{1}} = \sum_{S\in\binom{[d]}{1}} \mathfrak{e}_{S,S} = \sum_{i=1}^{d} \mathfrak{e}_{\{i\},\{i\}}.$$

The vector $\mathfrak{e}_{\{i\},\{i\}}$ is $(i2^d+i)$th in our chosen order, so $I_{\binom{d}{1}}$ is associated to the vector containing 1s in its (2^d+1)th entry, $(2\cdot 2^d+2)$th entry, and so on, and containing 0s everywhere else.

The pattern continues. Each additional basis element $I_{\binom{d}{k}}$ corresponds to a new vector, whose entries consist of 1s distributed across $\binom{d}{k}$ of the 2^{2d} components of $\operatorname{End}\bigwedge\mathbb{C}^d$. At each stage, identifying what specific entries are nonzero becomes a more and more involved clerical task. But this is a clear enough description that we can now establish the entries of the $d+1\times d+1$ matrix

$$A^*A.$$

Observe that

$$\begin{aligned}(A^*A)_{i,j} &= \sum_{l=1}^{2^{2d}} (A^*)_{i,l}A_{l,j}\\ &= \sum_{l=1}^{2^{2d}} A_{l,i}A_{l,j}\end{aligned}$$

This sum counts the number of times $A_{l,i}$ and $A_{l,j}$ are simultaneously nonzero as l ranges from 1 to 2^{2d}. Each column of A corresponds to a distinct basis vector $\sum_{S\in\binom{[d]}{r}}\mathfrak{e}_{S,S}$, where r corresponds uniquely to that column. If an entry $A_{l,i}$ is simultaneously nonzero with $A_{l,j}$, then the lth basis vector $\mathfrak{e}_{S,S}$ of $\operatorname{End}\bigwedge\mathbb{C}^d$ contributes to both the jth and ith columns of A. If $\mathfrak{e}_{S,S}$ contributes to $I_{\binom{d}{i}}$, then $|S|=i$. Thus

$$(A^*A)_{i,j} = \begin{cases}\langle I_{\binom{d}{i}}|I_{\binom{d}{i}}\rangle & \text{if } i=j,\\ 0 & \text{otherwise}\end{cases}.$$

Therefore, A^*A is a diagonal matrix. Since $\langle I_{\binom{d}{i}}|I_{\binom{d}{i}}\rangle = \binom{d}{i}$, the Weingarten matrix is easy to write down:

$$W_{i,j} = ((A^*A)^{-1})_{i,j} = \binom{d}{i}^{-1}[i=j].$$

Finally, by the fundamental theorem of Weingarten calculus,

$$\begin{aligned}
[\rho(g)]_{S,T}\overline{[\rho(g)]_{V,U}} &= [\rho(g)]_{S,T}[\rho(g)^*]_{U,V} \\
&= (\operatorname{Adj} \bigwedge \rho(g))_{(V,S),(U,T)} \\
\mathbb{E}[[\rho(g)]_{S,T}[\rho(g)^*]_{U,V}] &= \int_G (\operatorname{Adj} \bigwedge \rho(g))_{(V,S),(U,T)}\, dg \\
&= (AWA^*)_{(V,S),(U,T)} \\
&= \sum_{k,l=0}^{d} A_{(V,S),k} W_{k,l} A^*_{l,(U,T)} \\
&= \sum_{k=0}^{d} A_{(V,S),k} W_{k,k} A^*_{k,(U,T)} \qquad (W \text{ is diagonal}) \\
&= \sum_{k=0}^{d} \binom{d}{k}^{-1} A_{(V,S),k} A_{(U,T),k} \\
&= \frac{1}{\binom{d}{|S|}} [S=V][T=U]
\end{aligned}$$

The final equality says that if $(AWA^*)_{(V,S),(U,T)}$ is nonzero, then it must be equal to $\binom{d}{k}^{-1}$, with $A_{(V,S),k} = A_{(U,T),k} = 1$. But $A_{(V,S),k} = 1$ only under the circumstance that $\mathfrak{e}_{V,S}$ contributes to $I_{\binom{d}{k}}$, which is true only if $V = S$ and $|S| = k$. By the same reasoning, it must be true that $T = U$ and $|T| = k$. These conditions on S, T, U and V are handled by the terms $[S = V]$ and $[T = U]$.

In conclusion, $\rho(g)$ is minor-orthogonal. $\square$

The following fact is a theorem in [27], but here we derive it as a corollary of the previous theorem.

Corollary 7.2.4 ([27])**.** Any random matrix which is Haar-distributed over $\mathcal{U}(d)$ is minor-orthogonal.

Proof. Let $\rho : \mathcal{U}(d) \to \operatorname{Aut} \mathbb{C}^d$ be the tautological representation $g \mapsto g$. Since the exterior powers of ρ are irreducible and distinct by 5.5.5, theorem 7.2.3 applies. $\square$

Theorem 7.2.5 ([27])**.** Suppose that Q is a minor-orthogonal random matrix, and that Q is compactly supported over the unitary group. If A and B are self-adjoint, then

$$\mathbb{E}[\chi(A + QBQ^*)] = \sum_{i,j=0}^{d} x^{d-i-j} \frac{(d-i)!(d-j)!}{d!(d-i-j)!} \sigma_i(A)\sigma_j(B).$$

Proof. First, we will assume without loss of generality that A and B are diagonal. This is permissible because self-adjoint matrices are diagonalizable by unitaries, and by 7.1.5, conjugating A and B by unitaries does not change the value of our expression.

Now, by 2.4.12,

$$\sigma_k(A + QBQ^*) = \sum_{S \in \binom{[d]}{k}} [A + QBQ^*]_{S,S}.$$

And by 2.4.9

$$\begin{aligned}
[A + QBQ^*]_{S,S} &= \sum_{i=0}^{k} \sum_{U,V \in \binom{[k]}{i}} (-1)^{\|U\|_1 + \|V\|_1} [A]_{U(S),V(S)} [QBQ^*]_{\overline{U}(S),\overline{V}(S)} \\
&= \sum_{i=0}^{k} \sum_{U \in \binom{[k]}{i}} [A]_{U(S),U(S)} [QBQ^*]_{\overline{U}(S),\overline{U}(S)},
\end{aligned}$$

since A is diagonal. Again by 2.4.9,

$$\begin{aligned}
[QBQ^*]_{\overline{U}(S),\overline{U}(S)} &= \sum_{W,X \in \binom{[d]}{d-|U|}} [Q]_{\overline{U}(S),W} [B]_{W,X} [Q^*]_{X,\overline{U}(S)} \\
&= \sum_{W \in \binom{[d]}{d-|U|}} [Q]_{\overline{U}(S),W} [B]_{W,W} [Q]_{W,\overline{U}(S)},
\end{aligned}$$

since B is diagonal.

Combining these identities together, and using minor-orthogonality,

$$\begin{aligned}
&\mathbb{E}[\sigma_k(A+QBQ^*)] \\
&= \sum_{S\in\binom{[d]}{k}} \sum_{i=0}^{k} \sum_{U\in\binom{[k]}{i}} \sum_{W\in\binom{S}{k-i}} [A]_{U(S),U(S)}[B]_{W,W}\mathbb{E}[[Q]_{\overline{U}(S),W}[Q^*]_{W,\overline{U}(S)}] \\
&= \sum_{S\in\binom{[d]}{k}} \sum_{i=0}^{k} \sum_{U\in\binom{[k]}{i}} \sum_{W\in\binom{S}{k-i}} [A]_{U(S),U(S)}[B]_{W,W}\binom{d}{k-i}^{-1} \\
&= \sum_{S\in\binom{[d]}{k}} \sum_{i=0}^{k} \sum_{U\in\binom{[k]}{i}} [A]_{U(S),U(S)}\binom{d}{k-i}^{-1}\sigma_{k-i}(B) \\
&= \sum_{i=0}^{k} \sum_{U\in\binom{[d]}{i}} \binom{d-i}{k-i}[A]_{U,U}\binom{d}{k-i}^{-1}\sigma_{k-i}(B) \\
&= \sum_{i=0}^{k} \binom{d-i}{k-i}\binom{d}{k-i}^{-1}\sigma_i(A)\sigma_{k-i}(B) \\
&= \sum_{i=0}^{k} \frac{(d-i)!(k-i)!(d-k+i)!}{(k-i)!(d-i-k+i)!d!}\sigma_i(A)\sigma_{k-i}(B) \\
&= \sum_{i+j=k} \frac{(d-i)!(d-j)!}{d!(d-i-j)!}\sigma_i(A)\sigma_j(B).
\end{aligned}$$

Thus

$$\mathbb{E}[\chi(A+QBQ^*)] = \sum_{i+j\le d} x^{d-i-j}(-1)^{i+j}\frac{(d-i)!(d-j)!}{d!(d-i-j)!}\sigma_i(A)\sigma_j(B).$$

□

The expression on the right-hand side, seen as an operation combining $\chi(A)$ and $\chi(B)$, is known as the *d-finite free convolution* of those polynomials, which we write $\chi(A) \boxplus_d \chi(B)$. There are strong constraints on the roots of $p \boxplus_d q$ if the roots of p and q are known. For example, if p and q are real-rooted, then so is $p \boxplus_d q$ [27].

The knowledge of this expected value is a valuable contribution to the theory of finite free probability, but there is still a great deal of information in the distribution of $\chi(A+QBQ^*)$ which this expectation does not summarize. In the conclusion of [32], Mirabelli opened up the question of whether this result could be extended to analyze *covariance matrix* of $\chi(A+QBQ^*)$, an object closely related to the quadratic mixed moments of the vector. The answer is yes, but the precise formula which one eventually derives depends much more strongly on the distribution of Q. We will make this answer more precise in the following chapters.

Chapter 8

Weingarten calculus for tensor powers of $\mathcal{U}(d)$

In this chapter, we will develop a more sophisticated form of Weingarten calculus capable of integrating expressions like $\int g_{1,2}g_{2,2}\overline{g_{1,1}g_{1,2}}$, where g is Haar-distributed over $\mathcal{U}(2)$. This will require some more advanced representation theory than what we have seen before, in particular the notion of *Schur-Weyl duality.* In its most famous iteration, Schur-Weyl duality centers on the fact that the general linear group and the symmetric group have commuting actions on $(\mathbb{C}^d)^{\otimes n}$, which ends up having enormous consequences for the representation theory of the two groups. Here, we will instead consider a version of Schur-Weyl duality implicating $\mathcal{U}(d)$ with $\mathfrak{S}_n$. But readers familiar with the story of $\mathrm{GL}(d)$ can extrapolate almost all of their intuitions to this setting, due to the close relationship between it and $\mathcal{U}(d)$.

There are a number of papers proving the formula for Weingarten integrals over the unitary group. For instance, [33] proves it using special symmetric functions known as *Jack polynomials.* The authors of [30] and [9] use gadgets known as *Jucys-Murphy elements*, special operators in the algebra $\mathbb{C}\mathfrak{S}_n$. My primary source is [10], which approaches the situation via Schur-Weyl duality. I also found invaluable the way [23] rearticulated certain arguments from [10].

8.1 Schur-Weyl duality

My primary references for this section are [15] and [10]. I will be presenting the main facts without proof, since it takes a fair amount of space to develop them, and because these facts are well-established in the literature.

Definition 8.1.1. If λ is a tuple of positive integers such that $\sum_i \lambda_i = n$, we say that λ is a composition of n.

Definition 8.1.2. If λ is a weakly decreasing composition of n, then we say it is a partition of n. We write $\lambda \vdash n$ to denote this.

Definition 8.1.3. The length $l(\lambda)$ of a partition λ is that integer for which λ may be regarded as a vector in $\mathbb{Z}^{l(\lambda)}$.

Proposition 8.1.4 ([13])**.** The isoclasses of simple $\mathfrak{S}_n$-modules are in bijection with the partitions $\lambda \vdash n$ for which $l(\lambda) \leq d$. Thus we may denote by S^λ a representative of the isoclass corresponding to λ.

It is difficult to find a citation for the following claim in this exact form. It is more common for representation-theoretic literature in this area to treat groups such as the special linear, general linear or special unitary group than it is for this literature to treat the unitary group. The citation I've given here is a treatment of the special unitary group, which may be presented as $\mathcal{U}(d)/\mathcal{U}(1)$. The basic ideas described in this source also hold for the unitary group $\mathcal{U}(d)$, but proving this would probably require me to introduce the theory of weights, which I want to avoid for brevity's sake.

Proposition 8.1.5 ([19])**.** The isoclasses of simple $\mathcal{U}(d)$-modules are in bijection with the partitions $\lambda \vdash n$ for all $n \leq d$. Thus we may denote by U^λ a representative of the isoclass corresponding to λ.

Theorem 8.1.6 (Schur-Weyl, [45])**.**

$$(\mathbb{C}^d)^{\otimes n} \cong \bigoplus_{\lambda \vdash n, l(\lambda) \leq d} U^\lambda \otimes S^\lambda,$$

Definition 8.1.7. Let $\lambda \vdash n$. We define a Young tableau T of shape λ to be a doubly-indexed family $T_{i,j}$ of integers λ, such that $T_{i,\cdot}$ is of length λ_i for each i. If $T_{i,j}$ is a strictly increasing in j for all i, and is weakly increasing in i for all j, then we call T a semistandard Young tableau.

Proposition 8.1.8 ([43])**.** The dimension of U^λ for $\lambda \vdash n$ is the number of semistandard Young tableau of shape λ. We will denote this quantity by symbol

$$s_\lambda(1^n),$$

to allude to the objects s_λ known to algebraic combinatorialists as *Schur polynomials*, but we will leave Schur polynomials undefined. This quantity $s_\lambda(1^n)$ is also the multiplicity of S^λ in $(\mathbb{C}^d)^{\otimes n}$.

8.2 A lemma about projectors

This argument may be found in [13].

Definition 8.2.1. Let V be a $\mathbb{C}G$-module, and let W be a simple $\mathbb{C}G$-module corresponding to a representation $\gamma : G \to \operatorname{Aut} W$. We call the maximal $\mathbb{C}G$-submodule of V which is decomposable into a direct sum of copies of W the *W-isotypic component*, or alternatively the *γ-isotypic* component.

Lemma 8.2.2 ([13])**.** Let $\rho : G \to \operatorname{Aut} V$ be a unitary representation of a finite group. Let $\gamma : G \to \operatorname{Aut} W$ be a fixed irreducible unitary representation of G. Then the projection of V onto its γ-isotypic component may be given by

$$\psi_\gamma = \frac{\dim W}{|G|} \sum_{g \in G} \overline{\chi_\gamma(g)} g \in \operatorname{End} V.$$

Proof. Notice that ψ_γ is a $\mathbb{C}G$-module homomorphism. That is,

$$\begin{aligned}
\psi_\gamma(hv) &= \frac{\dim W}{|G|} \sum_{g \in G} \overline{\chi_\gamma(g)} g.hv \\
&= \frac{\dim W}{|G|} \sum_{hgh^{-1} \in G} \overline{\chi_\gamma(hgh^{-1})} hgh^{-1}.hv \\
&= \frac{\dim W}{|G|} \sum_{hgh^{-1} \in G} \overline{\chi_\gamma(hgh^{-1})} hg.v \\
&= h \frac{\dim W}{|G|} \sum_{hgh^{-1} \in G} \overline{\chi_\gamma(hgh^{-1})} g.v \\
&= h\psi_\gamma(v).
\end{aligned}$$

Suppose for a moment that V is irreducible. Then ψ is zero in the case that $V \not\cong W$, or else $\psi = cI$ for some $c \in \mathbb{C}$. In the latter case, $c = \frac{\operatorname{Tr} \psi_\gamma}{\dim W}$. Since

$$\operatorname{Tr} \psi_\gamma = \dim W \langle \chi_\gamma | \chi_\gamma \rangle = \dim W,$$

we know $c = 1$, so $\psi_\gamma = I$.

Now suppose that V may not be irreducible, so $V = \bigoplus_{i=1}^m V_i$, where each V_i is irreducible. Then ψ_γ restricts to a $\mathbb{C}G$-module homomorphism from each component V_i. For each V_i, we have that the trace of the restriction to each summand of the isotypic component is

$$\operatorname{Tr}(\psi_\gamma \restriction V_i) = \dim W \langle \chi_\gamma | \chi_i \rangle,$$

where χ_i is the character of V_i. As we have already seen, this forces these restrictions to either be zero in the case $V_i \not\cong W$, or else they are the identity function when $V_i \cong W$. Thus ψ_γ is the projection of V onto its γ-isotypic component. $\square$

8.3 Weingarten calculus for $(\mathbb{C}^d)^{\otimes n}$

The theory of Weingarten calculus for $(\mathbb{C}^d)^{\otimes n}$ is historically prior to the general theory used in this thesis, but here I will present it as a special case of the more general theory.

I am under the impression that the argument I that will occupy the rest of this chapter was originally developed by Benoît Collins and Piotr Śniady in [10], but I am primarily relying on Georg Kostenberger's presentation of the argument in [23].

Definition 8.3.1. Define

$$\mathcal{U}(d)^{\otimes n} = \{u^{\otimes n} \mid u \in \mathcal{U}(d)\}.$$

I will begin with a lemma that is a direct consequence of the translation-invariance of the Haar measure.

Lemma 8.3.2 ([23])**.** Let A be an element of $\mathrm{End}(\mathbb{C}^d)^{\otimes n}$. Then

$$\int Q^{\otimes n} A (Q^*)^{\otimes n} \, d\mu(Q),$$

commutes with every element of $\mathcal{U}(d)^{\otimes n}$, where μ is the Haar measure over $\mathcal{U}(d)$.

Proof. Let $U \in \mathcal{U}(d)$. Then

$$\begin{aligned} U^{\otimes n} \int Q^{\otimes n} A (Q^*)^{\otimes n} \, d\mu(Q) &= \int (UQ)^{\otimes n} A (Q^*)^{\otimes n} \, d\mu(Q) \\ &= \int Q^{\otimes n} A (Q^* U)^{\otimes n} \, d\mu(Q) \\ &= \int Q^{\otimes n} A (Q^*)^{\otimes n} \, d\mu(Q) U^{\otimes n}. \end{aligned}$$

□

We will embed the algebra $\mathbb{C}\mathfrak{S}_n$ in $\mathrm{End}(\mathbb{C}^d)^{\otimes n}$ via the representation

$$\rho(\sigma)(v_1 \otimes v_1 \otimes \ldots \otimes v_n) = v_{\sigma^{-1}(1)} \otimes v_{\sigma^{-1}(2)} \otimes \ldots \otimes v_{\sigma^{-1}(n)}.$$

I present the following claim without proof, because the proof is of moderate length and involves several concepts not directly relevant to the new ideas developed in this thesis.

Proposition 8.3.3 ([15])**.** The commutant of $\mathcal{U}(d)^{\otimes n}$ in $\mathrm{End}(\mathbb{C}^d)^{\otimes n}$ is $\rho(\mathbb{C}\mathfrak{S}_n)$.

Corollary 8.3.4 ([23], [10])**.** For each $A \in \mathrm{End}(\mathbb{C}^d)^{\otimes n}$, let

$$P(A) = \int Q^{\otimes n} A (Q^*)^{\otimes n} \, d\mu(Q).$$

Then $P(A) \in \rho(\mathbb{C}\mathfrak{S}_n)$.

Proof. This is a consequence of the fact that P commutes with $\mathcal{U}(d)^{\otimes n}$. □

Proposition 8.3.5 ([23])**.** The map P is surjective, i.e.,

$$\operatorname{im} P = \rho(\mathbb{C}\mathfrak{S}_n).$$

Proof. Let $A \in \rho(\mathbb{C}\mathfrak{S}_n)$. Then

$$\begin{aligned}
P(A) &= \int Q^{\otimes n} A (Q^*)^{\otimes n} \, d\mu(Q) \\
&= \int A Q^{\otimes n} (Q^*)^{\otimes n} \, d\mu(Q) \\
&= A \int I \, d\mu(Q) \\
&= A.
\end{aligned}$$

□

By similar applications of translation-invariance and 8.3.3, one can show the following as well.

Corollary 8.3.6 ([23])**.** *P is self-adjoint and idempotent.*

Thus P, as a linear endomorphism of a finite-dimensional vector space, has the same properties as the matrix product AWA^* did in our derivation of Weingarten calculus. In particular, I claim we can read off integrals of products of matrix entries of Q from the entries of P, construed as a matrix. Like in the chapter introducing Weingarten calculus, we will regard $(\mathbb{C}^d)^{\otimes n}$ to be generated by the basis $\{e_i\}$, where i ranges over all functions $[n] \to [d]$. Furthermore, we will view every pair of such functions i, i' as indexing a basis element $e_{i,i'}$ of $(\mathbb{C}^d)^{\otimes n} \otimes ((\mathbb{C}^d)^{\otimes n})^*$ with the property that $(e_{i,i'})_{k,l} = [i = k][i' = l]$. Since this tensor product is isomorphic to the endomorphism algebra $\mathrm{End}(\mathbb{C}^d)^{\otimes n}$, we may see the entries of P as indexed by quadruples $(i, i'), (j, j')$. With our notation set, we may express the following useful calculation.

Proposition 8.3.7 ([23], [10])**.**

$$P_{(i,i'),(j,j')} = \int Q_{i_1,j_1} Q_{i_2,j_2} ... Q_{i_n,j_n} (Q^*)_{i'_1,j'_1} (Q^*)_{i'_2,j'_2} ... (Q^*)_{i'_n,j'_n} \, d\mu(Q).$$

Proof.

$$\begin{aligned}
P_{(i,i'),(j,j')} &= \langle e_{i,i'} | P e_{j,j'} \rangle \\
&= \operatorname{Tr} e_{i,i'}^* \int Q^{\otimes n} e_{j,j'} (Q^*)^{\otimes n} \, d\mu(Q) \\
&= \int \operatorname{Tr} e_{i',i} Q^{\otimes n} e_{j,j'} (Q^*)^{\otimes n} \, d\mu(Q) \\
&= \int \sum_{k:[n]\to[k]} \left(e_{i',i} Q^{\otimes n} e_{j,j'} (Q^*)^{\otimes n} \right)_{k,k} \, d\mu(Q) \\
&= \int \sum_{k,l:[n]\to[d]} (e_{i',i})_{k,l} (Q^{\otimes n} e_{j,j'} (Q^*)^{\otimes n})_{l,k} \, d\mu(Q) \\
&= \int (Q^{\otimes n} e_{j,j'} (Q^*)^{\otimes n})_{i,i'} \, d\mu(Q) \\
&= \int \sum_{k:[n]\to[d]} (Q^{\otimes n})_{i,k} (e_{j,j'} (Q^*)^{\otimes n})_{k,i'} \, d\mu(Q) \\
&= \int \sum_{k,l:[n]\to[d]} (Q^{\otimes n})_{i,k} (e_{j,j'})_{k,l} ((Q^*)^{\otimes n})_{l,i'} \, d\mu(Q) \\
&= \int (Q^{\otimes n})_{i,j} ((Q^*)^{\otimes n})_{j',i'} \, d\mu(Q) \\
&= \int Q_{i_1,j_1} Q_{i_2,j_2} \ldots Q_{i_n,j_n} \overline{Q_{i'_1,j'_1} Q_{i'_2,j'_2}} \ldots \overline{Q_{i'_n,j'_n}} \, d\mu(Q).
\end{aligned}$$

□

Now that we know how to find the integrals we want in the entries of P, it is time to determine the values of those entries. This will involve some algebra.

Definition 8.3.8. Let $\Phi : \operatorname{End}(\mathbb{C}^d)^{\otimes n} \to \mathbb{C}\mathfrak{S}_n$ be given by

$$\Phi(A) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{Tr}(A\rho(\sigma^{-1}))\sigma.$$

Lemma 8.3.9 ([23], [10])**.** Φ is a $\mathbb{C}\mathfrak{S}_n$-$\mathbb{C}\mathfrak{S}_n$-bimodule homomorphism in the sense that

$$\Phi(A\rho(\sigma)) = \Phi(A)\sigma$$

and

$$\Phi(\rho(\sigma)A) = \sigma\Phi(A).$$

Proof. First,

$$\begin{aligned}\Phi(A\rho(\tau)) &= \sum_{\sigma\in\mathfrak{S}_n} \mathrm{Tr}(A\rho(\tau)\rho(\sigma^{-1}))\sigma \\ &= \sum_{\sigma\in\mathfrak{S}_n} \mathrm{Tr}(A\rho(\tau\sigma^{-1}))\sigma \\ &= \sum_{\sigma\in\mathfrak{S}_n} \mathrm{Tr}(A\rho(\sigma^{-1}))\sigma\tau \\ &= \Phi(A)\tau.\end{aligned}$$

Second,

$$\begin{aligned}\Phi(\rho(\tau)A) &= \sum_{\sigma\in\mathfrak{S}_n} \mathrm{Tr}(\rho(\tau)A\rho(\sigma^{-1}))\sigma \\ &= \sum_{\sigma\in\mathfrak{S}_n} \mathrm{Tr}(\rho(\tau)^{-1}\rho(\tau)A\rho(\sigma^{-1})\rho(\tau))\sigma \\ &= \sum_{\sigma\in\mathfrak{S}_n} \mathrm{Tr}(A\rho(\sigma^{-1}\tau))\sigma \\ &= \sum_{\sigma\in\mathfrak{S}_n} \mathrm{Tr}(A\rho(\sigma^{-1}))\tau\sigma \\ &= \tau\sum_{\sigma\in\mathfrak{S}_n} \mathrm{Tr}(A\rho(\sigma^{-1}))\sigma \\ &= \tau\Phi(A).\end{aligned}$$

□

In the following, we will reuse the notation ψ_λ from 8.2.2 to denote the projector of $\rho(\mathbb{C}\mathfrak{S}_n)$ onto the S^λ-isotypic component of $(\mathbb{C}^d)^{\otimes n}$. Recall that $s_\lambda(1^d)$ is notation for the number of semistandard Young tableaux possessing d cells. I continue to follow [23]'s presentation of the argument.

Lemma 8.3.10 ([23], [10])**.**

$$\Phi(I) = n! \sum_{\lambda\vdash n, l(\lambda)\le d} \frac{s_\lambda(1^d)}{\chi_\lambda(e)}\psi_\lambda.$$

Proof.

$$\begin{aligned}
\Phi(I) &= \sum_{\sigma\in\mathfrak{S}_n} \mathrm{Tr}(\rho(\sigma^{-1}))\sigma \\
&= \sum_{\sigma\in\mathfrak{S}_n} \chi_\rho(\sigma^{-1})\sigma \\
&= \sum_{\sigma\in\mathfrak{S}_n} \chi_\rho(\sigma^{-1})\sigma \\
&= \sum_{\sigma\in\mathfrak{S}_n} \overline{\chi_\rho(\sigma)}\sigma \\
&= \sum_{\sigma\in\mathfrak{S}_n} \sum_{\lambda\vdash n, l(\lambda)\le d} s_\lambda(1^d)\overline{\chi_\lambda}(\sigma)\sigma && \text{(via 5.4.3)} \\
&= \sum_{\lambda\vdash n, l(\lambda)\le d} s_\lambda(1^d) \sum_{\sigma\in\mathfrak{S}_n} \overline{\chi_\lambda}(\sigma)\sigma \\
&= \sum_{\lambda\vdash n, l(\lambda)\le d} \frac{n! s_\lambda(1^d)}{\chi_\lambda(e)}\psi_\lambda && \text{(via 8.2.2)} \\
&= n! \sum_{\lambda\vdash n, l(\lambda)\le d} \frac{s_\lambda(1^d)}{\chi_\lambda(e)}\psi_\lambda
\end{aligned}$$

□

Definition 8.3.11. Define

$$\mathrm{Wg}(\sigma) = \frac{1}{(n!)^2} \sum_{\lambda\vdash n, l(\lambda)\le d} \frac{\chi_\lambda(e)^2}{s_\lambda(1^d)}\chi_\lambda(\sigma).$$

We will call this function the Weingarten function for the unitary group. Extending by linearity on the generators σ, we regard it as an element of $\rho(\mathbb{C}\mathfrak{S}_n) \subseteq \mathrm{End}(\mathbb{C}^d)^{\otimes n}$.

Lemma 8.3.12 ([23], [10])**.** Denote by $\Phi(I)^{-1}$ the multiplicative inverse of $\Phi(I)$ in $\mathrm{End}(\mathbb{C}^d)^{\otimes n}$, i.e., such that $\Phi(I)\Phi(I)^{-1} = I$. Then

$$\mathrm{Wg} = \Phi(I)^{-1}.$$

Proof. Note that $\Phi(I)$ is a direct sum of scalar operators, one for each ψ_λ in the sum. Therefore, we may invert $\Phi(I)$ componentwise by taking the reciprocals of the coefficients of the maps ψ_λ. Each summand may be inverted as follows, due to the fact that ψ_λ is the identity on its own image.

$$n!\frac{s_\lambda(1^d)}{\chi_\lambda(e)}\psi_\lambda \mapsto \frac{\chi_\lambda(e)}{s_\lambda(1^d)n!}\psi_\lambda$$

Then we can expand $\psi_\lambda = (\chi_\lambda(e)/n!)\chi_\lambda$, acquiring

$$\frac{1}{(n!)^2}\frac{\chi_\lambda(e)^2}{s_\lambda(1^d)}\chi_\lambda.$$

This establishes the claim. $\square$

Lemma 8.3.13 ([23], [10]). For $A \in \mathrm{End}(\mathbb{C}^d)^{\otimes n}$,

$$\Phi(A) = P(A)\Phi(I).$$

Proof. First I'll point out that $\mathrm{Tr}(A) = \mathrm{Tr}(P(A))$. This can be seen by the calculation

$$\begin{aligned}
\mathrm{Tr}(P(A)) &= \mathrm{Tr}\left(\int_{\mathcal{U}(d)} Q^{\otimes n} A (Q^*)^{\otimes n}\, d\mu(Q)\right) \\
&= \int_{\mathcal{U}(d)} \mathrm{Tr}\left(Q^{\otimes n} A (Q^*)^{\otimes n}\right)\, d\mu(Q) \\
&= \int_{\mathcal{U}(d)} \mathrm{Tr}\left(A (Q^*)^{\otimes n} Q^{\otimes n}\right)\, d\mu(Q) \\
&= \int_{\mathcal{U}(d)} \mathrm{Tr}(A)\, d\mu(Q) \\
&= \mathrm{Tr}(A).
\end{aligned}$$

Using this fact, we can observe that

$$\begin{aligned}
\Phi(A) &= \sum_{\sigma} \mathrm{Tr}(A\rho(\sigma^{-1}))\rho(\sigma) \\
&= \sum_{\sigma} \mathrm{Tr}(P(A\rho(\sigma^{-1})))\rho(\sigma) \\
&= \sum_{\sigma} \mathrm{Tr}(P(A)\rho(\sigma^{-1}))\rho(\sigma) \\
&= \Phi(P(A)) \\
&= \Phi(P(A)I) \\
&= P(A)\Phi(I).
\end{aligned}$$

$\square$

Lemma 8.3.14 ([23], [10]). Let $X, Y \in \rho(\mathbb{C}\mathfrak{S}_n)$ and let $A \in \mathrm{End}(\mathbb{C}^d)^{\otimes n}$. Then

$$P(XAY) = XP(A)Y.$$

Proof. By 8.3.3,

$$\begin{aligned}
P(XAY) &= \int Q^{\otimes n} XAY (Q^*)^{\otimes n}\, d\mu(Q) \\
&= X \int Q^{\otimes n} A (Q^*)^{\otimes n}\, d\mu(Q) Y \\
&= XP(A)Y.
\end{aligned}$$

$\square$

Lemma 8.3.15 ([23], [10])**.** For $A, B \in \rho(\mathbb{C}\mathfrak{S}_n)$,

$$\Phi(AP(B)) = \Phi(A)\Phi(B)\,\mathrm{Wg}\,.$$

Proof.

$$\begin{aligned}
\Phi(AP(B)) &= P(AP(B))\Phi(I) && \\
&= P(A)P(P(B))\Phi(I) && \text{By 8.3.14} \\
&= P(A)P(B)\Phi(I) && P \text{ is idempotent} \\
&= \Phi(A)\Phi(I)^{-1}\Phi(B)\Phi(I)^{-1}\Phi(I) && \text{By 8.3.13} \\
&= \Phi(A)\,\mathrm{Wg}\,\Phi(B) && \\
&= \Phi(A)\Phi(B)\,\mathrm{Wg} && \mathrm{Wg} \text{ is a scalar element of } \mathbb{C}\mathfrak{S}_n.
\end{aligned}$$

□

Theorem 8.3.16 ([23], [10])**.**

$$\begin{aligned}
&\int Q_{i_1,j_1}Q_{i_2,j_2}...Q_{i_n,j_n}(Q^*)_{i'_1,j'_1}(Q^*)_{i'_2,j'_2}...(Q^*)_{i'_n,j'_n}\,d\mu(Q) \\
&= \sum_{\sigma,\tau\in\mathfrak{S}_n}[i = i'\circ\sigma][j = j'\circ\tau]\,\mathrm{Wg}(\tau\sigma^{-1}).
\end{aligned}$$

Proof. Recall from the argument in 8.3.7 that we can make the following substitution.

$$\begin{aligned}
&\int Q_{i_1,j_1}Q_{i_2,j_2}...Q_{i_n,j_n}(Q^*)_{i'_1,j'_1}(Q^*)_{i'_2,j'_2}...(Q^*)_{i'_n,j'_n}\,d\mu(Q) \\
&= \langle e_{i,i'} \mid Pe_{j,j'}\rangle \\
&= \mathrm{Tr}(e_{i',i}P(e_{j,j'}))
\end{aligned}$$

Recall that

$$\Phi(A) = \sum_{\sigma}\mathrm{Tr}(A\rho(\sigma^{-1}))\sigma.$$

Then $\mathrm{Tr}(e_{i',i}P(e_{j,j'}))$ is equal to the coefficient of e in $\Phi(e_{i',i}P(e_{j,j'}))$. We can calculate this latter as

$$\Phi(e_{i',i}P(e_{j,j'})) = (\Phi(e_{i',i})\Phi(e_{j,j'})\Phi(I)^{-1})$$

via 8.3.15. We can compute each term of this expression as follows.

$$\begin{aligned}
\Phi(e_{i',i}) &= \sum_{\sigma} \operatorname{Tr}(e_{i',i}\rho(\sigma^{-1}))\sigma \\
&= \sum_{\sigma} e_i^* \rho(\sigma^{-1}) e_{i'} \sigma \\
&= \sum_{\sigma} [i = \sigma^{-1} \circ i']\sigma \\
\Phi(e_{j,j'}) &= \sum_{\sigma} \operatorname{Tr}(e_{j,j'}\rho(\sigma^{-1}))\sigma \\
&= \sum_{\sigma} e_{j'}^* \rho(\sigma^{-1}) e_j \sigma \\
&= \sum_{\sigma} [j' = \rho(\sigma^{-1}) \circ j]\sigma.
\end{aligned}$$

By reindexing as convenient, we get

$$\begin{aligned}
\Phi(e_{i',i})\Phi(e_{j,j'}) \operatorname{Wg} &= \sum_{\sigma} [i = i' \circ \sigma]\sigma \sum_{\tau} [j = j' \circ \tau]\tau^{-1} \sum_{\kappa} \operatorname{Wg}(\kappa)\kappa \\
&= \sum_{\sigma,\tau,\kappa} [i = i' \circ \sigma][j = j' \circ \tau] \operatorname{Wg}(\kappa)\sigma\tau^{-1}\kappa.
\end{aligned}$$

Since we are looking for the coefficient of e in this sum, we are looking to restrict the above sum to the expressions in which $\kappa = (\sigma\tau^{-1})^{-1}$, acquiring

$$\sum_{\sigma,\tau} [i = i' \circ \sigma][j = j' \circ \tau] \operatorname{Wg}(\tau\sigma^{-1})$$

as expected. $\square$

Chapter 9

Second-order finite free probability

In this chapter, we will assume that $\mathcal{G}$ is a compact subgroup of $\mathcal{U}(d)$, containing the permutation matrices, that Q is Haar-distributed over $\mathcal{G}$, and that A and B are matrices which may be diagonalized by elements of $\mathcal{G}$. Our paradigmatic example is $\mathcal{G} = \mathcal{U}(d)$, where A and B are self-adjoint. We will continue to study the distribution of $\chi(A + QBQ^*)$ by investigating the mixed second moment

$$X_{e,f} = \mathbb{E}[\sigma_e(A + QBQ^*)\sigma_f(A + QBQ^*)].$$

In particular, we will develop a formula for $X_{e,f}$.

Let's take a moment to reflect on what the purpose of a formula is. After all, isn't $\mathbb{E}[\sigma_e(A + QBQ^*)\sigma_f(A + QBQ^*)]$ already a formula for $X_{e,f}$? The definition of $X_{e,f}$ tells the reader that it is a quantity that may be found by integrating a particular function. The integral does not in general specify a method of exact evaluation, which gives the expression a certain opacity to the reader. This is especially true when the Haar measure over $\mathcal{G}$ is specified by its characterizing properties, rather than a concrete density function. (Although, density functions are known for some important groups, for example see [40]). To the contrary, finite sums and combinatorical operations are transparent, in the sense that exact evaluation of such expressions have obvious (though perhaps inefficient) exact computer implementations. This transparency also often lends itself to qualitative developments. For instance, the real-rootedness of the average polynomial $\mathbb{E}[\chi(A + QBQ^*)]$ was established based on the summation formula which Marcus, Spielman and Srivastava developed for it. I aim in this chapter to develop an analogous summation formula for $X_{e,f}$, in order to make possible qualitative observations about it.

In the following derivation, we will denote the eigenvalues of A and B as $(a_i)_{i=1}^d$ and $(b_i)_{i=1}^d$, respectively, but readers may take advantage of 7.1.8 by ignoring the precise ordering of these sequences. We will also assume that $\chi(A+QBQ^*)$ has real coefficients on the support of Q, but the approach featured in this chapter would extend to the complex case by considering second *-moments also.

9.1 An initial simplification

Recall that the notation $[M]_S$ denotes the matrix minor $[M]_{S,S}$ of the S,S-principal submatrix of M.

The first rewriting we will apply to X, via 2.4.12, is

$$\sigma_e(A+QBQ^*) = \sum_{E\in\binom{[d]}{e}} [A+QBQ^*]_E.$$

Thus

$$X_{e,f} = \sum_{E\in\binom{[d]}{e}} \sum_{F\in\binom{[d]}{f}} \mathbb{E}[[A+QBQ^*]_E[A+QBQ^*]_F].$$

Next, via 2.4.8, we can rewrite

$$[A+QBQ^*]_E = \sum_{i=0}^{e} \sum_{G,H\in\binom{[e]}{i}} (-1)^{\|G\|_1+\|H\|_1}[A]_{G(E),H(E)}[QBQ^*]_{\overline{G}(E),\overline{H}(E)},$$

where $\overline{G}(E)$ denotes $E \setminus G(E) = ([e]\setminus G)(E)$. Just as we did in the chapter on first-order finite free probability, we will assume that A and B are diagonal matrices without loss of generality. Thus $[A]_{G(E),H(E)} = 0$ unless $G(E) = H(E)$, i.e., $G = H$. So we can apply 2.4.9, acquiring

$$[A+QBQ^*]_E = \sum_{i=0}^{e} \sum_{G\in\binom{[e]}{i}} [A]_{G(E)}[QBQ^*]_{\overline{G}(E)}.$$

Thus

$$X_{e,f} = \sum_{E\in\binom{[d]}{e}} \sum_{F\in\binom{[d]}{f}} \sum_{i=0}^{e}\sum_{j=0}^{f} \sum_{G\in\binom{[e]}{i}} \sum_{H\in\binom{[f]}{j}} [A]_{G(E)}[A]_{H(F)}\mathbb{E}[[QBQ^*]_{\overline{G}(E)}[QBQ^*]_{\overline{H}(F)}].$$

Next, we can rewrite

$$\begin{aligned}
[QBQ^*]_{\overline{G}(E)} &= \sum_{K\in\binom{[d]}{e-i}} [Q]_{\overline{G}(E),K}[BQ^*]_{K,\overline{G}(E)} \\
&= \sum_{K\in\binom{[d]}{e-i}} \sum_{L\in\binom{[d]}{e-i}} [Q]_{\overline{G}(E),K}[B]_{K,L}[Q^*]_{L,\overline{G}(E)}
\end{aligned}$$

by repeated application of 2.4.16. Then, by the diagonality of B, we can again apply 2.4.9 to get

$$[QBQ^*]_{\overline{G}(E)} = \sum_{K\in\binom{[d]}{e-i}} [Q]_{\overline{G}(E),K}[B]_K[Q^*]_{K,\overline{G}(E)}.$$

Let's apply the new rewrite rule:

$$X_{e,f} = \sum_{E\in\binom{[d]}{e}} \sum_{F\in\binom{[d]}{f}} \sum_{i=0}^{e}\sum_{j=0}^{f} \sum_{G\in\binom{[e]}{i}} \sum_{H\in\binom{[f]}{j}} \sum_{K\in\binom{[d]}{e-i}} \sum_{L\in\binom{[d]}{f-j}} [A]_{G(E)}[A]_{H(F)}[B]_K[B]_L$$
$$\cdot\, \mathbb{E}[[Q]_{\overline{G}(E),K}[Q^*]_{K,\overline{G}(E)}[Q]_{\overline{H}(F),L}[Q^*]_{L,\overline{H}(F)}]$$
$$= \sum_{E\in\binom{[d]}{e}} \sum_{F\in\binom{[d]}{f}} \sum_{i=0}^{e}\sum_{j=0}^{f} \sum_{G\in\binom{[e]}{i}} \sum_{H\in\binom{[f]}{j}} \sum_{K\in\binom{[d]}{e-i}} \sum_{L\in\binom{[d]}{f-j}} [A]_{G(E)}[A]_{H(F)}[B]_K[B]_L$$
$$\cdot\, \mathbb{E}[|[Q]_{\overline{G}(E),K}|^2|[Q]_{\overline{H}(F),L}|^2]$$

Now, since $\mathcal{G}$ contains the permutation matrices, the distribution of Q is equal to the distributions of QP and PQ for all permutation matrices P. Recall that left multiplication by P permutes the rows of Q and right multiplication permutes the columns. Thus for all permutations σ, τ acting on $[d]$, the expectation in the previous expression is equal to

$$\mathbb{E}[|[Q]_{\overline{G}(E)\circ\sigma,K\circ\tau}|^2|[Q]_{\overline{H}(F)\circ\sigma,L\circ\tau}|^2]$$

So the value of the expression on the previous line must only depend on the cardinalities

$$|\overline{G}(E)|, |\overline{H}(F)|, |\overline{G}(E)\cap\overline{H}(F)| \text{ and } |K\cap L|.$$

Therefore, we will define the following.

Definition 9.1.1. Let e', f', z and w be given as nonnegative integers. Let $o = e'+1-z, u = o+f', s = e'+1-w, t = s+f'$. Then define

$$q_d^{\mathcal{G}}(e', f', z, w) = \mathbb{E}[|[Q]_{[e'],[f']}|^2|[Q]_{[o,u],[s,t]}|^2],$$

where $[m,n]$ is defined as the set of integers from m to n, inclusive. Note that $q_d^{\mathcal{G}}$ is a partial function. We will take care to only use it where its defining expression is well-defined. We will typically suppress the superscript and subscript and write $q(e', f', z, w)$.

We can also rewrite

$$[A]_{G(E)}[A]_{H(F)}[B]_K[B]_L = \prod_{x\in G(E)} a_x \prod_{x\in H(F)} a_x \prod_{y\in K} b_y \prod_{y\in L} b_y.$$

Finally, we can exchange the order of summation to instantiate the variables E and F at the innermost two sums we perform in the iteration.

The following proposition summarizes all we currently know. This will be the first of the main results of this thesis.

Proposition 9.1.2. Let $\mathcal{G}$ be a compact, permuting subgroup of $\mathcal{U}(d)$, and let A and B be $\mathcal{G}$-diagonal matrices. Let Q be Haar-distributed over $\mathcal{G}$. Denote our quantity of interest by

$$X_{e,f} = \mathbb{E}[\sigma_e(A+QBQ^*)\sigma_f(A+QBQ^*)],$$

and denote the eigenvalues of A and B as $(a_i)_{i=1}^d$ and $(b_i)_{i=1}^d$, respectively. Then

$$X_{e,f} = \sum_{i=0}^{e}\sum_{j=0}^{f}\sum_{G\in\binom{[d]}{i}}\sum_{H\in\binom{[d]}{j}}\sum_{K\in\binom{[d]}{e-i}}\sum_{L\in\binom{[d]}{f-j}}\prod_{x\in G}a_x\prod_{x\in H}a_x\prod_{y\in K}b_y\prod_{y\in L}b_y$$
$$\sum_{G\subseteq E\in\binom{[d]}{e}}\sum_{H\subseteq F\in\binom{[d]}{f}} q(e-i, f-j, |(E\setminus G)\cap(F\setminus H)|, |K\cap L|).$$

We have reached a critical stage in the derivation. Instead of trying to compute the entire polynomial at once, we can reframe the question. If you give me a multidegree over the indeterminates $a_1, a_2, ..., a_d, b_1, b_2, ...b_d$, can I give you back the coefficient corresponding to the monomial of that multidegree in X?

In particular, suppose we are looking for the coefficient of a given monomial $\mathfrak{m} = \prod_{r=1}^d a_r^{m_r}\prod_{s=1}^d b_s^{n_s}$ in $X_{e,f}$. Consider any indeterminate a_r; the power m_r of this indeterminate is determined by its multiplicity in $\prod_{x\in G}a_x\prod_{x\in H}a_x$, which means that $m_r\in\{0,1,2\}$. The same is true of each n_s. Let $R=\sum_r m_r$. Also, for $i\in\{0,1,2\}$, let

$$R_i = \{r\in[d] \mid m_r = i\}$$

and

$$S_i = \{s\in[d] \mid n_s = i\}.$$

We will sometimes write R, R_i and S_i as functions of $\mathfrak{m}$, but we will typically suppress this in our notation. We will try to find the indices i, j, G, H, K, L, E', F' contributing to this $\mathfrak{m}$. We can write the coefficient of $\mathfrak{m}$ formed over these indices as

$$\sum_{i+j=R}\ \sum_{\substack{(G,H)\in\binom{[d]}{i}\times\binom{[d]}{j}\\ G\cap H=R_2, G\triangle H=R_1}}\ \sum_{\substack{(K,L)\in\binom{[d]}{e-i}\times\binom{[d]}{f-j}\\ K\cap L=S_2, K\triangle L=S_1}}$$
$$\sum_{E'\in\binom{[d]\setminus G}{e-i}}\sum_{F'\in\binom{[d]\setminus H}{f-j}} q(e-i, f-j, |E'\cap F'|, |S_2|).$$

It would be convenient to be able to fix $|E'\cap F'|$ and count how many pairs (E', F') satisfy this intersection in the above sum. Then we could delete the indices E' and F' and sum instead over $c = |E'\cap F'|$. We will develop a combinatorial argument in order to do so. For the duration of this argument, we fix i, j, G and H.

We will imagine ourselves distributing indices among urns, and then painting them, in order to construct E' and F' under these constraints. Let $D = [d]\setminus(G\cup H)$. Suppose that we have d white balls labelled by the integers $[d]$, distributed among urns labelled D, $H\setminus G$, $G\setminus H$, and $G\cap H$. The sets labelling these urns partition $[d]$. We fix a nonnegative integer c, which corresponds to our fixed value $|E'\cap F'|$ from before. We have both blue and red paint, and painting a ball both blue and red makes it purple. We want to apply red paint to $e-i$ balls (announcing that these indices lie in E'), we want to apply blue paint to $f-j$ balls (announcing that these indices lie in F'), and we will have c of the balls purple (announcing that these indices lies in $E' \cap F'$). Only balls in D and $H \setminus G$ may be painted red, and only balls in D and $G \setminus H$ may be painted blue. How many ways are there to satisfy this system of constraints? This corresponds to the problem of finding a pair of sets (E', F') satisfying the constraints for $\mathfrak{m}$.

We may immediately conclude that all c of the balls to be painted purple must belong to urn D, since urn D is the only one where we may apply both red and blue paint. So there are $\binom{|D|}{c}$ choices of purple balls which we can immediately make, expending c coats of both red and blue paint. The problem now is to allocate all the remaining coats of red and blue paint, without accidentally painting any more balls purple.

Let's begin by painting balls red. We have $e-i-c$ remaining coats of red paint to allocate, which we can split among the urn labelled D and the urn labelled $H \setminus G$. Let x be the number of coats of red paint we allocate among the $|D|-c$ white balls in urn D. Then the remaining $e-i-c-x$ coats of red paint must go to $H \setminus G$. Then there are

$$\sum_{x=0}^{e-i-c} \binom{|H \setminus G|}{e-i-c-x} \binom{|D|-c}{x}$$

ways of allocating our red paint, if we set $\binom{n}{k}$ to zero when $k<0$ or $k>n$.

After we have expended the red coats of paint, we must paint $f-j-c$ balls blue. The argument is similar for red balls, but we now have x fewer white balls in urn D which we can paint. Combining the sums for red and blue balls, we have

$$\begin{aligned}
&\sum_{x=0}^{e-i-c} \binom{|H \setminus G|}{e-i-c-x} \binom{|D|-c}{x} \sum_{y=0}^{f-j-c} \binom{|G \setminus H|}{f-j-c-y} \binom{|D|-c-x}{y} \\
&= \sum_{x=0}^{e-i-c} \sum_{y=0}^{f-j-c} \binom{|H \setminus G|}{e-i-c-x} \binom{|D|-c}{x} \binom{|G \setminus H|}{f-j-c-y} \binom{|D|-c-x}{y} \\
&= \sum_{x=0}^{e-i-c} \sum_{y=0}^{f-j-c} \binom{|H \setminus G|}{e-i-c-x} \binom{|G \setminus H|}{f-j-c-y} \binom{|D|-c}{x, y}
\end{aligned}$$

ways of satisfying the constraints, if we set the multinomial coefficient $\binom{n}{k_1, k_2}$ to 0 when $k_1 < 0, k_2 < 0$, or $k_1 + k_2 > n$.

Let

$$M(e', f', G, H, c) = \sum_{x=0}^{e'-c} \sum_{y=0}^{f'-c} \binom{|H \setminus G|}{e'-c-x} \binom{|G \setminus H|}{f'-c-y} \binom{|D|-c}{x, y}.$$

Then the previous argument tells us we can write the coefficient of $\mathfrak{m}$ as

$$\sum_{i+j=R} \sum_{\substack{(G,H)\in\binom{[d]}{i}\times\binom{[d]}{j} \\ G\cap H=R_2, G\triangle H=R_1}} \sum_{\substack{(K,L)\in\binom{[d]}{e-i}\times\binom{[d]}{f-j} \\ K\cap L=S_2, K\triangle L=S_1}} \sum_{c=0}^{d} M(e-i,f-j,G,H,c)q(e-i,f-j,c,|S_2|).$$

We can apply a similar, but simpler transformation to sum over pairs (K, L) satisfying $K\triangle L = S_1$. We need to pick which of these elements belong to K and which belong to L. There are $e - i - |S_2|$ available positions in K after allocating indices for S_2. After we have chosen these elements for K, the rest must go to L, so we have no more choices to make. So we can add a coefficient of $\binom{|S_1|}{e-i-|S_2|}$ and eliminate the sum over (K, L). We can summarize this as the following new result.

Proposition 9.1.3. The coefficient of $\mathfrak{m}$ in $X_{e,f}$ is

$$\sum_{i+j=R(\mathfrak{m})} \sum_{\substack{(G,H)\in\binom{[d]}{i}\times\binom{[d]}{j} \\ G\cap H=R_2(\mathfrak{m}), G\triangle H=R_1(\mathfrak{m})}} \binom{|S_1(\mathfrak{m})|}{e-i-|S_2(\mathfrak{m})|}$$
$$\sum_{c=0}^{d} M(e-i,f-j,G,H,c)q(e-i,f-j,c,|S_2(\mathfrak{m})|).$$

9.2 Specialized to the unitary group

The previous expression

$$\sum_{i+j=R} \sum_{\substack{(G,H)\in\binom{[d]}{i}\times\binom{[d]}{j} \\ G\cap H=R_2, G\triangle H=R_1}} \binom{|S_1|}{e-i-|S_2|} \sum_{c=0}^{d} M(e-i,f-j,G,H,c)q_d^{\mathcal{G}}(e-i,f-j,c,|S_2|)$$

for the coefficient of $\mathfrak{m}$ in $X_{e,f}$ is free in the group $\mathcal{G}$. We will now specialize $\mathcal{G} = \mathcal{U}(d)$ and evaluate this formula using the results from the previous chapter.

Let $\mathbf{e} = [e-i], \mathbf{k} = [e-i], \mathbf{f} = [e-i+1-c, e-i+1-c+f-j]$, and

$\mathbf{l} = [e-i+1-|S_2|, e-i+1-|S_2|+f-j]$. Then

$$\begin{aligned}
&q_d(e-i, f-j, c, |S_2|) \\
=&\mathbb{E}[|[Q]_{\mathbf{e},\mathbf{k}}|^2 |[Q]_{\mathbf{f},\mathbf{l}}|^2] \\
&\mathbb{E}\left[\left|\sum_{\sigma\in\mathfrak{S}_{e-i}} \prod_{m=1}^{e-i} \operatorname{sgn}(\sigma)(Q_{\mathbf{e},\mathbf{k}})_{m,\sigma(m)}\right|^2 \left|\sum_{\tau\in\mathfrak{S}_{f-j}} \operatorname{sgn}(\tau) \prod_{n=1}^{f-j} (Q_{\mathbf{f},\mathbf{l}})_{n,\tau(n)}\right|^2\right] \\
=&\mathbb{E}\Bigg[\sum_{\sigma,\tau\in\mathfrak{S}_{e-i}} \left(\operatorname{sgn}(\sigma) \prod_{m=1}^{e-i} (Q_{\mathbf{e},\mathbf{k}})_{m,\sigma(m)}\right) \left(\prod_{n=1}^{e-i} \operatorname{sgn}(\tau) \overline{(Q_{\mathbf{e},\mathbf{k}})_{n,\tau(n)}}\right) \\
&\sum_{\gamma,\delta\in\mathfrak{S}_{f-j}} \left(\operatorname{sgn}(\gamma) \prod_{o=1}^{f-j} (Q_{\mathbf{f},\mathbf{l}})_{o,\gamma(o)}\right) \left(\operatorname{sgn}(\delta) \prod_{p=1}^{e-j} \overline{(Q_{\mathbf{f},\mathbf{l}})_{p,\delta(p)}}\right)\Bigg] \\
=&\sum_{\sigma,\tau\in\mathfrak{S}_{e-i}} \sum_{\gamma,\delta\in\mathfrak{S}_{f-j}} \operatorname{sgn}(\sigma\tau)\operatorname{sgn}(\gamma\delta) \\
&\mathbb{E}\left[\prod_{m=1}^{e-i} (Q_{\mathbf{e},\mathbf{k}})_{m,\sigma(m)} \prod_{n=1}^{e-i} \overline{(Q_{\mathbf{e},\mathbf{k}})_{n,\tau(n)}} \prod_{o=1}^{f-j} (Q_{\mathbf{f},\mathbf{l}})_{o,\gamma(o)} \prod_{p=1}^{f-j} \overline{(Q_{\mathbf{f},\mathbf{l}})_{p,\delta(p)}}\right]
\end{aligned}$$

If $f : [m] \to [n], g : [m'] \to [n]$ are two functions, denote by $f \sqcup g$ the function $[m+m'] \to [n]$ given by

$$f \sqcup g(i) = \begin{cases} f(i) & \text{if } 1 \le i \le m \\ g(i-m) & \text{if } m+1 \le i \le m+m' \end{cases}.$$

We will also regard each of $\mathbf{e}, \mathbf{f}, \mathbf{k}$ and $\mathbf{l}$ as functions from $[e-i]$ or $[f-j]$ into $[d]$, where, for instance, $\mathbf{e}(k)$ is the kth element of $\mathbf{e}$.

Then we can express

$$\begin{aligned}
q_d(e-i, f-j, c, |S_2|) = &\sum_{\sigma,\tau\in\mathfrak{S}_{e-i}} \sum_{\gamma,\delta\in\mathfrak{S}_{f-j}} \operatorname{sgn}(\sigma\tau)\operatorname{sgn}(\gamma\delta) \\
&\mathbb{E}\left[\prod_{m=1}^{e+f-i-j} Q_{\mathbf{i}(m),\mathbf{j}(m)} \prod_{o=1}^{e+f-i-j} \overline{Q_{\mathbf{i}'(o),\mathbf{j}'(o)}}\right],
\end{aligned}$$

where $\mathbf{i} = \mathbf{e} \sqcup \mathbf{f}$, $\mathbf{j} = (\mathbf{k} \circ \sigma) \sqcup (\mathbf{l} \circ \gamma)$, $\mathbf{i}' = \mathbf{e} \sqcup \mathbf{f}$, and $\mathbf{j}' = (\mathbf{k} \circ \tau) \sqcup (\mathbf{l} \circ \delta)$. In this form, we can apply the fundamental theorem of Weingarten calculus for the unitary group, acquiring

$$\begin{aligned}
q_d(e-i, f-j, c, |S_2|) = &\sum_{\sigma,\tau\in\mathfrak{S}_{e-i}} \sum_{\gamma,\delta\in\mathfrak{S}_{f-j}} \operatorname{sgn}(\sigma\tau)\operatorname{sgn}(\gamma\delta) \\
&\left[\sum_{\eta,\zeta\in\mathfrak{S}_{e+f-i-j}} [\mathbf{i} = \mathbf{i}' \circ \eta][\mathbf{j} = \mathbf{j}' \circ \zeta] \operatorname{Wg}(\eta^{-1}\zeta)\right].
\end{aligned}$$

Thus we have an expression for $X_{e,f}$, breaking it up into a sum over the functions M and q, the first of which is independent of the group $\mathcal{G}$, and the second of which requires some development of the Weingarten calculus with respect to $\mathcal{G}$. In the case $\mathcal{G} = \mathcal{U}(d)$, we have a formula which involves only summation and combinatorical objects.

9.3 Prospects

In this chapter, we have seen the development of a summation formula for

$$\mathbb{E}[\sigma_e(A + QBQ^*)\sigma_f(A + QBQ^*)],$$

and a specialization of that formula to the coefficient of each individual monomial $\mathfrak{m}$ in the eigenvalues of A and B. The naive procedure which corresponds to the formula is complex. It requires, among other things, the use of the character table for the symmetric group. Loosely speaking, computer scientists conventionally consider an algorithm whose runtime grows as a polynomial in the size of its input to be efficient. The number of partitions of n grows faster than any polynomial in n, so the computation of the Weingarten function Wg cannot be efficient as we grow n and d together. Since the formulae I derived depend on Wg, the naive procedure corresponding to the formula is not efficient in general.

However, this dependency on Wg began only at the final step, when we determined a summation formula for the function q. The rest of the formula is relatively manageable from a combinatorial point of view, as the rest of it is essentially just a sum over binomial and multinomial coefficients. The general technique exhibited in this derivation shows that the problem may be split into these two parts, the combinatorial part (which is relatively mundane) and a representation-theoretic part, which is apparently hard. The author is of the opinion that this would not be too hard to repeat for higher moments of $\chi(A+QBQ^*)$, and while the combinatorial and representation-theoretic aspects would get more and more complex, it would remain possible to split the problem into these two parts using the same formulae applied in this derivation.

Thus it would be very interesting if the apparent difficulty of the representation-theoretic part of the problem turned out to be spurious. In this derivation, we derived a formula for

$$\mathbb{E}[|[Q]_{S,T}|^2|[Q]_{U,V}|^2]$$

by implicating the integrand with the representation corresponding to $\bigotimes^n \mathbb{C}^d$. The resulting Weingarten calculus is extremely powerful, and extremely general, due to the fact that it can compute the integral of any product of unitary matrix coefficients whatsoever. What if a less powerful Weingarten calculus, designed for this specific task, turned out also to involve easier computations? It is, for example, be possible to embed the above integrand into

$$V = \left[\bigoplus_{k=0}^{d} \operatorname{End} \bigwedge^{k} \mathbb{C}^d\right]^{\otimes 2}.$$

It is conceivable that, due to the greater specificity of this representation, the corresponding Weingarten matrix will be less general, and thus easier to compute with. This certainly turned out to be the case for the first moments, in which case the entries were either of the form $\binom{d}{k}^{-1}$, or they were zero. That fact could have been deduced from the same formula as I used for q in this chapter, as Campbell and Yin did in [6]. But if the more specialized Weingarten matrix were easier to compute with, then the formula developed for q in this chapter could be replaced by something more manageable, and the rest of the derivation could remain. I have left it outside of the scope of this thesis to pursue this line of work further, but it may be the case that qualitative observations about $\chi(A + QBQ^*)$ would be more forthcoming given a formula that easier to work with.

Chapter 10

Conclusion

10.1 Future improvements

There are a few ways in which this research could be expanded upon.

10.1.1 Use a more specialized representation

In this thesis, I focused on two different applications of Weingarten integration. For the first, I considered the representation $\operatorname{End} \bigwedge \mathbb{C}^d$ in order to evaluate integrals of the form $\mathbb{E}[[Q]_{S,T}[Q^*]_{U,V}]$. For the second, I considered a representation which amounted to $\operatorname{End}[(\mathbb{C}^d)^{\otimes n}]$ in order to evaluate integrals of the form $\mathbb{E}[|[Q]_{S,T}|^2|[Q]_{U,V}|^2]$.

However, it is possible to embed that second integrand into a much more specific representation. In particular, the representation

$$V = \left[\bigoplus_{k=0}^{d} \operatorname{End} \bigwedge^{k} \mathbb{C}^d\right]^{\otimes 2}$$

is fully capable of hosting this integrand as a matrix entry. While this representation may appear more "complicated" than $\operatorname{End}(\mathbb{C}^d)^{\otimes n}$, I would rather say it is more "specialized". It is a general rule of thumb in math that the more general your formula is, the less useful it is for performing computations. The Weingarten matrix for $\operatorname{End}[(\mathbb{C}^d)^{\otimes n}]$ allows one to compute a strictly greater variety of integrals than the Weingarten matrix for V. Therefore, I expect that the final Weingarten matrix for V will have a simpler structure, or, at least, it will encode information about the integrals $\mathbb{E}[|[Q]_{S,T}|^2|[Q]_{U,V}|^2]$ more concisely than the general formula does.

Applying the fundamental theorem of Weingarten calculus to this latter representation requires a substantially more technical argument about the structure of $\mathcal{U}(d)$-modules, which I decided to cut from the scope of this thesis. However, my preliminary findings suggest that it will be possible to find a general expression for this Weingarten matrix for V over generic dimension d.

10.1.2 Reiterate for higher $*$-moments and for other groups

One could easily reiterate the arguments of the last chapter to study the higher $*$-moments of the characteristic polynomial of $A + QBQ^*$. Recall that the argument had two main phases. First, we used algebra to pull summation expressions and the matrices A and B outside of the expectation operator, and isolated an expression in terms of the random Haar-distributed matrix Q inside the expectation brackets. I used standard combinatorical arguments to deal with the sums outside of the expectation, and I used the representation theory of the unitary group to evaluate the expectation expression itself.

This basic move could be imitated with higher $*$-moments. The resulting formulae would involve more summation indices and more complicated combinatorics. I would not recommend doing this unless someone had developed a better way of interpreting the resulting expressions, or if they had an application which needed these numerical results for some reason.

One could also easily swap out the unitary group for another group whose Weingarten calculus is well-understood, such as the orthogonal group.

10.1.3 Connect with infinite free probability

Adam Marcus's paper [28] explained how the finite d-free convolution formula

$$\mathbb{E}[\chi(A + QBQ^*)] = \sum_{i+j\leq d} x^{d-i-j}(-1)^{i+j}\frac{(d-i)!(d-j)!}{d!(d-i-j)!}\sigma_i(A)\sigma_j(B)$$

connects with (infinite) free probability. (Recall that this formula expresses the linear $*$-moments of $\chi(A + QBQ^*)$.) In [28], Marcus introduced an interesting power series device called the d-finite K-transform, which relates that finite d-free convolution formula to some power series devices which were developed by the progenitor of free probability, Dan Voiculescu. It would be interesting to know if the formulae I have developed in this thesis for the covariance of the characteristic polynomial could be related to a collection of power series transforms, and thereby associated with specific phenomena occurring in infinite-dimensional free probability.

Bibliography

[1] Sheldon Axler. “Down With Determinants!” In: *The American Mathematical Monthly* 102.2 (1995), pp. 139–154. DOI: 10.1080/00029890.1995.11990548. eprint: https://doi.org/10.1080/00029890.1995.11990548. URL: https://doi.org/10.1080/00029890.1995.11990548.

[2] Sheldon Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer, 1996. DOI: 10.1007/978-3-031-41026-0.

[3] Sheldon Axler. *Measure, Integration and Real Analysis*. Graduate Texts in Mathematics. Springer Cham, 2020. DOI: 10.1007/978-3-030-33143-6.

[4] Patrick Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. Wiley, 1995.

[5] Daniel Bump. *Lie Groups*. Graduate Texts in Mathematics. Springer, 2004. DOI: 10.1007/978-1-4614-8024-2.

[6] Jacob Campbell and Zhi Yin. “Finite free convolutions via Weingarten calculus”. In: *Random Matrices: Theory and Applications* 10.04 (Nov. 2020). DOI: 10.1142/s2010326321500386. URL: https://doi.org/10.1142%2Fs2010326321500386.

[7] *Carleman theorem*. Accessed November 8 from the Encyclopedia of Mathematics. First published February 2011, last edited June 2020. URL: https://encyclopediaofmath.org/wiki/Carleman_theorem.

[8] *Characteristic polynomial*. Accessed November 7 from the Encyclopedia of Mathematics. First published February 2011, last edited December 2020. URL: https://encyclopediaofmath.org/wiki/Characteristic_polynomial.

[9] Benoit Collins, Sho Matsumoto, and Jonathan Novak. *The Weingarten Calculus*. 2021. DOI: 10.48550/ARXIV.2109.14890. URL: https://arxiv.org/abs/2109.14890.

[10] Benoıt Collins and Piotr Śniady. “Integration with Respect to the Haar Measure on Unitary, Orthogonal and Symplectic Group”. In: *Communications in Mathematical Physics* 264.3 (Mar. 2006), pp. 773–795. DOI: 10.1007/s00220-006-1554-3. URL: https://doi.org/10.1007%2Fs00220-006-1554-3.

[11] Joe Diestel and Angela Spalsbury. *The Joys of Haar Measure*. Graduate Studies in Mathematics. American Mathematics Society, 2014. DOI: 10.1090//gsm/150.

[12] David S. Dummit and Richard M. Foote. *Abstract Algebra*. Wiley, 2003.

[13] William Fulton and Joe Harris. *Representation Theory: A First Course*. Graduate Texts in Mathematics. Springer-Verlag, 1991. DOI: 10.1007/978-1-4612-0979-9.

[14] Paul Garrett. *Projectors, projection maps, orthogonal projections*. Course notes for "Functional Analysis" at the University of Minnesota. 2017. URL: https://www-users.cse.umn.edu/~garrett/m/fun/notes_2016-17/projectors.pdf.

[15] Roe Goodman and Nolan R. Wallach. *Symmetry, Representations, and Invariants*. Graduate Texts in Mathematics. Springer, 2009. DOI: 10.1007/978-0-387-79852-3.

[16] Darij Grinberg. *Notes on the combinatorial fundamental of algebra*. 2022. URL: http://www.cip.ifi.lmu.de/~grinberg/primes2015/sols.pdf.

[17] Chris Hall, Doron Puder, and William F. Sawin. "Ramanujan coverings of graphs". In: *Advances in Mathematics* 323 (Jan. 2018), pp. 367–410. DOI: 10.1016/j.aim.2017.10.042. URL: https://doi.org/10.1016%2Fj.aim.2017.10.042.

[18] Thomas W. Hungerford. *Algebra*. Graduate Texts in Mathematics. Springer-Verlag, 1974. DOI: 10.1007/978-1-4612-6101-8.

[19] C. Itzykson and M. Nauenberg. "Unitary Groups: Representations and Decompositions". In: *Rev. Mod. Phys.* 38 (1 Jan. 1966), pp. 95–120. DOI: 10.1103/RevModPhys.38.95. URL: https://link.aps.org/doi/10.1103/RevModPhys.38.95.

[20] Marcel de Jeu. "Determinate multidimensional measures, the extended Carleman theorem and quasi-analytic weights". In: (2001). DOI: 10.48550/ARXIV.MATH/0111019. URL: https://arxiv.org/abs/math/0111019.

[21] David Joyce. *Eigenvalues, eigenvectors, and eigenspaces of linear operators*. Course notes for "Math 130 Linear Algebra" at Clark University. 2013. URL: https://www2.clarku.edu/faculty/djoyce/ma130/eigen.pdf.

[22] Joel Kamnitzer. *Lectures on geometric constructions of the irreducible representations of GL_n*. 2010. arXiv: 0912.0569 [math.RT].

[23] Georg Köstenberger. *Weingarten Calculus*. 2021. arXiv: 2101.00921 [math.PR].

[24] Joseph P. S. Kung, Gian-Carlo Rota, and Catherine H. Yan. *Combinatorics: The Rota Way*. Cambridge Mathematical Library. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511803895.

[25] Amos Lapidoth. *A Foundation in Digital Communication*. ETH Zurich, 2009.

[26] Adam Marcus, Daniel A Spielman, and Nikhil Srivastava. *Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem*. 2014. arXiv: 1306.3969 [math.CO].

[27] Adam Marcus, Daniel A. Spielman, and Nikhil Srivastava. *Finite free convolutions of polynomials*. 2015. DOI: 10.48550/ARXIV.1504.00350. URL: https://arxiv.org/abs/1504.00350.

[28] Adam W. Marcus. *Polynomial convolutions and (finite) free probability*. 2021. DOI: 10.48550/ARXIV.2108.07054. URL: https://arxiv.org/abs/2108.07054.

[29] Marvin Marcus. "Determinants of Sums". In: *The College Mathematics Journal* 21.2 (1990), pp. 130–135. DOI: 10.1080/07468342.1990.11973297. eprint: https://doi.org/10.1080/07468342.1990.11973297. URL: https://doi.org/10.1080/07468342.1990.11973297.

[30] Sho Matsumoto and Jonathan Novak. *Jucys-Murphy Elements and Unitary Matrix Integrals*. 2012. arXiv: 0905.1992 [math.CO].

[31] James A. Mingo and Roland Speicher. *Free Probability and Random Matrices*. Fields Institute Monographs. Springer. DOI: 10.1007/978-1-4939-6942-5.

[32] Benjamin (Benno) Pine Mirabelli. "Hermitian, Non-Hermitian and Multivariate Finite Free Probability". PhD thesis. Princeton University, 2021.

[33] Marcel Novaes. *Elementary derivation of Weingarten functions of classical Lie groups*. 2014. DOI: 10.48550/ARXIV.1406.2182. URL: https://arxiv.org/abs/1406.2182.

[34] Gert K. Pederson. *Analysis Now*. Graduate Texts in Mathematics. Springer, 1989. DOI: 10.1007/978-1-4612-1007-8.

[35] Claudio Procesi. *Lie Groups: An Approach through Invariants and Representations*. Universitext. Springer, 2007.

[36] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, 1974.

[37] Bruce E. Sagan. *The Symmetric Group*. Graduate Texts in Mathematics. Springer, 2001. DOI: 10.1007/978-1-4757-6804-6.

[38] *Schur's Lemma*. Accessed 23 May 2023 from MathWorld–A Wolfram Web Resource. 2004. URL: https://mathworld.wolfram.com/SchursLemma.html.

[39] Sasha Sodin. *The classical moment problem*. Course notes for "LTTC 2017: The classical moment problem" at Queen Mary University of London. 2019. URL: https://webspace.maths.qmul.ac.uk/a.sodin/teaching/moment/clmp.pdf.

[40] Christoph Spengler, Marcus Huber, and Beatrix C. Hiesmayr. "Composite parameterization and Haar measure for all unitary and special unitary groups". In: *Journal of Mathematical Physics* 53.1 (Jan. 2012), p. 013501. DOI: 10.1063/1.3672064. URL: https://doi.org/10.1063%2F1.3672064.

[41] David E. Speyer. *The formula for the orthogonal projection*. Course notes for "Math 417: Matrix Algebra" at the University of Michigan. 2013. URL: https://dept.math.lsa.umich.edu/~speyer/417/OrthoProj.pdf.

[42] Terence Tao. *254A, Notes 5: Free probability*. Accessed 18 May 2023. 2010. URL: https://terrytao.wordpress.com/2010/02/10/245a-notes-5-free-probability/.

[43] Dmitri Vvedensky. *Group Theory*. Course notes for "Group Theory - 2001" at Imperial College London. 2001. URL: http://cmth.ph.ic.ac.uk/people/d.vvedensky/groups/Chapter9.pdf.

[44] Eric W. Weisstein. *Iverson Bracket*. Accessed 23 May 2023 from MathWorld–A Wolfram Web Resource. URL: https://mathworld.wolfram.com/IversonBracket.html.

[45] Lin Zhang. *Matrix integrals over unitary groups: An application of Schur-Weyl duality*. 2015. arXiv: 1408.3782 [quant-ph].