MATH 595
REPRESENTATION-THEORETIC METHODS IN QUANTUM INFORMATION

SUJEET BHALERAO

ABSTRACT. These are lecture notes based on the course Math 595: Representation-theoretic methods in quantum information taught by Prof. Felix Leditzky. Any mistakes in these notes are my own.

CONTENTS

1. Basics from representation theory 2
1.1. Representations 2
1.2. Irreducible representations and decompositions 2
2. Basics from representation theory (contd.) 4
2.1. Tensor and dual representations, hom spaces 4
2.2. Group algebra and characters 4
2.3. From finite and compact groups 6
3. Schur-Weyl duality 6
3.1. Representations of direct product groups 6
3.2. Double commutant theorem 7
3.3. Schur-Weyl duality 8
4. Irreps of symmetric and unitary groups 10
4.1. Minimal projections and irreducible representations 10
4.2. Conjugacy classes of the symmetric group 10
4.3. Constructing the irreps of S_n and U_d 11
5. Mathematics of finite dimensional quantum information theory 13
5.1. Quantum systems and quantum states 13
5.2. Measurements 14
5.3. Composite systems and entanglement 14
5.4. Distance measures 16
6. Invariant states 16
6.1. Werner states 16
6.2. Isotropic states 20
7. The de Finetti theorem 21
7.1. Extendability of quantum states 21
7.2. A de Finetti theorem for pure symmetric states 21
7.3. Extension to permutation-invariant mixed states 23
8. Approximate cloning 24
8.1. The no-cloning theorem 24
8.2. Approximate cloning machines 25
8.3. Further results on approximate cloning 26
9. Spectrum estimation 27
9.1. Problem setup 27
9.2. Symmetries of spectrum estimation 27
9.3. Weak Schur sampling 28
9.4. Asymptotics of spectrum estimation 30
10. Quantum state tomography 31
10.1. Warm-up: pure state estimation 31
10.2. Symmetries of state tomography 32
10.3. Error analysis of our tomography protocol 33
11. Universal Quantum Source Compression 33
11.1. Classical sources and entropy 34
11.2. Compressing a classical source 34
11.3. Strong typicality and universal compression 35
11.4. Quantum source compression 36
11.5. Universal compression based on Schur-Weyl duality 38
References 38

Date: Nov 11, 2022.

1. Basics from representation theory

1.1. Representations

Definition 1. A group $(G, \cdot)$ is a set G together with a binary operation $\cdot : G \times G \to G$ satisfying:

- Associativity: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in G$.
- Identity element: there exists an element $e \in G$ with $e \cdot g = g \cdot e = g$ for all $g \in G$.
- Inverse: for all $g \in G$, there exists $h \in G$ with $g \cdot h = h \cdot g = e$. Such an element h is unique; it is called the *inverse* of g, and is denoted by g^{-1}.

Example 1. If $(\mathbb{F}, +, \cdot)$ is a field, then $(\mathbb{F}, +)$ is a group. The collection of bijections from the set $\{1, 2, ..., n\}$ to itself is the *symmetric group S_n*. The set of invertible linear maps from a vector space V to itself is the *general linear group* $\mathrm{GL}(V)$.

Representation theory is the study of groups via group actions on vector spaces.

Definition 2. An action of a group G on a set X is a map $\varphi : G \times X \to X$ such that for all $x \in X$ and $g, h \in G$ it holds that $\varphi(e, x) = x$ and $\varphi(g, \varphi(h, x)) = \varphi(gh, x)$.

Definition 3. A representation (φ, V) of a group G on a vector space V (over a field $\mathbb{F}$) is a group homomorphism $\varphi : G \to \mathrm{GL}(V)$.

A representation always satisfies $\varphi(e) = \mathbb{1}_V$ and $\varphi(g^{-1}) = \varphi(g)^{-1}$. The *dimension* or *degree* of a representation (φ, V) is the dimension of V. This course only deals with finite-dimensional representations. Rename?

Example 2. Let G be a cyclic group of order d generated by g. Let $V = \mathbb{C}^d$ with basis $|0\rangle, |1\rangle, ..., |d-1\rangle$. Consider a linear operator X on V defined by $X|i\rangle = |i + 1 \mod d\rangle$ for all i. Then the map $g \mapsto X$ determines a representation (φ, V) of G. Another representation (φ', V) is defined the map $g \mapsto Z$, where $Z|j\rangle = w^j|j\rangle$ for a primitive d-th root of unity.

The two representations in the example above are essentially the same, a notion which we now make precise:

Definition 4. Let G be a group. Two representations (φ, V) and (φ', V') of G are said to be *isomorphic* or *similar* if there exists a vector space isomorphism $\psi : V \to V'$ such that $\varphi'(g) = \psi \circ \varphi(g) \circ \psi^{-1}$ for all $g \in G$.

For example, the matrix X corresponding to the shift $|i\rangle \mapsto |\, [i-1] \pmod{d}\rangle$ of $|0\rangle, ..., |d-1\rangle$ has eigenvalues $e^{2\pi i k/d}$ for $k = 0, 1, \ldots d-1$. Hence, if $w = e^{2\pi i/d}$, a primitive root of unity, then the unitary U diagonalizing X satisfies $\varphi' = U \circ \varphi \circ U^{\dagger}$.
Here are some examples of representations for any group G:

Example 3. The trivial representation: $\varphi(g) = \mathbb{1}_{\mathbb{F}}$ for all $g \in G$, where $\mathbb{F}$ is some field.

Example 4. The regular representation of a (finite) group G: Let $n = |G|$ and $V \cong \mathbb{C}^n$ with basis $\{|g\rangle\}_{g \in G}$. The linear extension of the map $\varphi(g) : |h\rangle \mapsto |gh\rangle$ to all of V is called the *regular representation*. Conversely, let (ψ, W) be a representation such that there exists $w \in W$ so that $\{\psi(g)(w)\}_{g \in G}$ is a basis of W. Then ψ is isomorphic to the regular representation.

Example 5. The permutation representation: Let X be a finite set and G be a group acting on X. Consider the free vector space generated by X, i.e., $V \cong \mathbb{C}^m$, where $m = |X|$, and $\{|x\rangle\}_{x \in X}$ is a basis for V. Then the linear extension of the map $\varphi(g) : |x\rangle \mapsto |gx\rangle$ defines the *permutation representation* of G.

Note that the regular representation of G is the permutation representation of G that results from G acting on itself by left multiplication.

1.2. Irreducible representations and decompositions

Definition 5. Let (φ, V) be a representation of a group G. A subspace $W \subset V$ is called *invariant* or *stable* if $\varphi(g)|w\rangle \in W$ for all $|w\rangle \in W$ and $g \in G$. The restriction $\varphi|_W$ of φ onto W is called a *subrepresentation.*

Example 6. Let G be a finite group with $n = |G|$ and $(\varphi, \mathbb{C}^n)$ be the regular representation. Let $W = \mathrm{span}(\sum_{g \in G} |g\rangle)$. Then $(\varphi|_W, W)$ is a subrepresentation of $(\varphi, \mathbb{C}^n)$.

Note $\{0\}$ and V are always invariant subspaces.

Definition 6. A representation (φ, V) is called *irreducible* if $\{0\}$ and V are the only invariant subspaces of V.

A one-dimensional representation is always irreducible. For example, the one-dimensional subspace W in example 6 is irreducible. A goal of representation theory is to find all irreducible representations of a group G.

Definition 7. Let (φ_1, V_1) and (φ_2, V_2) be a representation of a group G. Then the direct sum $V_1 \oplus V_2$ affords the representation $[(\varphi_1 \oplus \varphi_2)(g)](v_1 \oplus v_2) := [\varphi_1(g)](v_1) \oplus [\varphi_2(g)](v_2)$ of G; this is called the *direct sum* of the representations (φ_1, V_1) and (φ_2, V_2).

Definition 8. A representation is called *completely reducible* if it decomposes as a direct sum of irreducible representations.

Proposition 1.1. Let (φ, V) be a representation of a finite group G, where V is a vector space over a field whose characteristic does not divide the order of G. Then every G-invariant subspace W has a G-invariant complement W', i.e., $V = W \oplus W'$ (as vector spaces and as representations).

Proof sketch. Let P_W be the projection onto W and define

$$Q_W = \frac{1}{|G|} \sum_{g \in G} \varphi(g) \circ P_W \circ \varphi(g)^{-1}.$$

Then one can check that $\operatorname{Im} Q_W = W$ and $W' := \ker Q_W$ is the desired G-invariant complement. ■

Alternate proof using Weyl's unitarity trick. Let (φ, V) be a representation over $\mathbb{C}$ and let $\langle \cdot | \cdot \rangle : V \times V \to \mathbb{C}$ be an inner product on V. Define a new inner product by

$$\langle v | w \rangle_G := \frac{1}{|G|} \sum_{g \in G} \langle \varphi(g) v | \varphi(g) w \rangle .$$

Then for every G-invariant subspace W the orthogonal complement $W^\perp$ (taken w.r.t. $\langle \cdot | \cdot \rangle_G$) is G-invariant as well, and $V = W \oplus W^\perp$ as representations. Moreover, (φ, V) is a *unitary representation* w.r.t $\langle \cdot | \cdot \rangle_G$, that is, $\varphi(G) \subset U(V)$ and $\varphi(g^{-1}) = \varphi(g)^{-1} = \varphi(g)^\dagger$.
For general unitary representations (φ, V) and an invariant subspace $W \subset V$, the orthogonal complement $W^\perp$ is again invariant. ■

Proposition 1.2 (Maschke's theorem)**.** Every finite-dimensional representation of a finite group G over a field with characteristic not dividing $|G|$ is completely reducible.

Proof. Use induction on $\dim V$ and the preceding proposition. ■

The proposition above is referred to as *Maschke's theorem*, and when the representation is over C, it says that for a finite group G and a finite-dimensional representation V of G over $\mathbb{C}$, we can write $V = V_1 \oplus \cdots \oplus V_m$ with each V_i irreducible. Is this decomposition unique?

Proposition 1.3 (Schur's Lemma)**.** Let (φ_1, V_1) and (φ_2, V_2) be irreducible representations of a group G, and let $f : V_1 \to V_2$ be a G-equivariant linear map, that is, f satisfies $f \circ \varphi_1(g) = \varphi_2(g) \circ f$ for all $g \in G$. Then:

(1) Either f is invertible (and hence $V_1 \cong V_2$) or $f = 0$.

(2) If $V_1 = V_2$ is finite-dimensional over an algebraically closed field $\mathbb{F}$ (for example $\mathbb{F} = \mathbb{C}$), then $f = \lambda \mathbb{1}_{V_1}$ for some $\lambda \in \mathbb{F}$.

Proof. (1) Suppose $f \neq 0$. Then $\ker f \neq V_1$ is a G-invariant subspace of V_1, so $\ker f = \{0\}$ by irreducibility of V_1. Likewise, $\operatorname{Im} f \neq \{0\}$ is a G-invariant subspace of V_2, so $\operatorname{Im} f = V_2$ by irreducibility of V_2. This proves f is invertible.

(2) $\mathbb{F}$ being algebraically closed guarantees that the linear map f has an eigenvalue, say $\lambda \in \mathbb{F}$. The map $f' = f = \lambda \mathbb{1}_{V_1}$ is G-equivariant, and it is not invertible since its kernel has a non-zero eigenvector of f. By (1), it follows that $f' = 0$, and so $f = \lambda \mathbb{1}_{V_1}$. ■

Corollary 1. Let G be an abelian group. Then any complex irreducible representation of G is one-dimensional.

Definition 9 (Isotypical component)**.** Let V be a finite-dimensional representation of a finite group G over $\mathbb{C}$, and consider a decomposition $V = \oplus_k V_k$, where each V_k is the direct sum of n_k copies of an irreducible representation W_k of G, i.e., $V_k = W_k \oplus \cdots \oplus W_k = W_k^{\oplus n_k} = W_k \otimes \mathbb{C}^{n_k}$, such that $W_k \ncong W_j$ for $j \neq k$. Then V_k is called an *isotypical component*.

An application of Schur's lemma shows:

Proposition 1.4. The decomposition $V = \oplus_k V_k$ of a representation V into isotypical components V_k is unique, and so are the multiplicities n_k of W_k in V_k.

Proof. See [Tel05]. ■

2. Basics from representation theory (contd.)

2.1. Tensor and dual representations, hom spaces

Definition 10. Let (φ, V) and (ψ, W) be representations of a group G. Then $(\varphi \otimes \psi)(g) := \varphi(g) \otimes \psi(g)$ defines a representation on $V \otimes W$ called the *tensor representation*.

Note: even if V and W are irreducible representations, the representation $V \otimes W$ is in general reducible.

Example 7. Let (φ, V) be a representation of G and consider the tensor representation $W = V \otimes V$. Let $\mathbb{F} : V \otimes V \to V \otimes V$ be the swap operator defined as the linear extension of the map $\mathbb{F}(|x\rangle \otimes |y\rangle) = |y\rangle \otimes |x\rangle$ for $|x\rangle, |y\rangle \in V$. Then we have the decomposition $V \otimes V = \mathrm{Sym}^2(V) \oplus \mathrm{Alt}^2(V)$, where $\mathrm{Sym}^2(V) := \{|z\rangle \in V \otimes V : \mathbb{F}|z\rangle = |z\rangle\}$ and $\mathrm{Alt}^2(V) := \{|z\rangle \in V \otimes V : \mathbb{F}|z\rangle = -|z\rangle\}$. $\mathrm{Sym}^2(V)$ and $\mathrm{Alt}^2(V)$ are both G-invariant subspaces called the *symmetric* and *antisymmetric square* respectively.

Definition 11. Let (φ, V) be a representation of G. Write V^* for the dual space of V, i.e., V^* is the set of linear maps from V to $\mathbb{C}$. The *dual* representation (φ^*, V^*) is defined as $\varphi^*(g)(L) := L \circ \varphi(g)^{-1}$ for $g \in G$ and $L \in V^*$.

The dual representation satisfies for all $g \in G$, $|v\rangle \in V$, $\langle w| \in V^*$ that $\varphi^*(g) = \varphi(g^{-1})^T$ and $(\varphi^*(g)\langle w|)(\varphi(g)|v\rangle) = \langle w|\varphi^*(g)^T\varphi(g)|v\rangle = \langle w|v\rangle$. It also holds that (φ^*, V^*) is irreducible iff (φ, V) is, and if (φ, V) is unitary then $\varphi^*(g) = \overline{\varphi(g)}$ (i.e. $\varphi^*(g)$ is the complex conjugate of $\varphi(g)$).

Definition 12. Let V, W be two vector spaces over $\mathbb{F}$. Then $\mathrm{Hom}(V, W)$ is the vector space of linear maps from V to W. Now let (φ, V), (ψ, W) be two representations of a group G. Then G acts on $\mathrm{Hom}(V, W)$ by sending $f : V \to W$ to $\psi(g) \circ f \circ \varphi(g)^{-1}$, which turns $\mathrm{Hom}(V, W)$ into a representation of G.

Note that setting $W = \mathbb{C}$ in the definition above with the trivial action of G recovers the dual representation of G. We also have

1. $\mathrm{Hom}(V, W) \cong V^* \otimes W$ as vector spaces and representations.
2. The set of vectors in V invariant under the action of G make up the set $V^G := \{|v\rangle \in V : \varphi(g)|v\rangle = |v\rangle \text{ for all } g \in G\}$. With this notation we have $\mathrm{Hom}_G(V, W) := \mathrm{Hom}(V, W)^G = (V^* \otimes W)^G$.
3. If $V = \oplus_i V_i$ is an isotypical decomposition with isotypical components $V_i = W_i^{n_i}$ for pairwise inequivalent irreducible representations W_i of G, then $n_i = \dim \mathrm{Hom}_G(V, V_i) = \dim(V^* \otimes V_i)^G$.

2.2. Group algebra and characters

Recall that in defining the regular representation of a group G we denoted by V the free vector space on the elements $\{|g\rangle\}_{g \in G}$ of G, that is, V is the set of formal linear combination of elements of G. The group multiplication endows V with the structure of an algebra, with multiplication defined by

$$\left[\sum_{g \in G} c_g |g\rangle\right] \cdot \left[\sum_{h \in G} d_h |h\rangle\right] = \sum_{g,h \in G} c_g d_h |gh\rangle = \sum_{g \in G} f_g |g\rangle,$$

with $f_g = \sum_{h \in G} c_{gh^{-1}} d_h$. This multiplication on V is associative, has the group identity element e as the multiplicative identity, and satisfies distributivity over addition. Thus $(V, +, \cdot)$ is an algebra, called the *group algebra*, and is denoted by $\mathbb{C}[G]$ (alternatively $\mathbb{C}G, \mathcal{A}_{\mathbb{C}}(G)$).

A representation of an algebra $\mathcal{A}$ over a field $\mathbb{F}$ is an algebra homomorphism $\mathcal{A} \to \mathrm{End}_{\mathbb{F}}(V)$ into the algebra of endomorphisms on an $\mathbb{F}$-vector space V with multiplication defined by composition of linear operators on V.

For $\mathcal{A} = \mathbb{C}[G]$, any representation (φ, V) of G can be extended to a representation $(\widetilde{\varphi}, V)$ of $\mathbb{C}[G]$ by setting $\widetilde{\varphi}(|g\rangle) = \varphi(g)$ and linearly extending to all of $\mathbb{C}[G]$. Conversely, any t $(\widetilde{\varphi}, V)$ of $\mathbb{C}[G]$ yields a t of G by restricting $\widetilde{\varphi}$ to $\{|g\rangle\}$. Therefore representations of G correspond exactly to representations of $\mathbb{C}[G]$.

Another interpretation of elements of $\mathbb{C}[G]$ is as functions $f : G \to \mathbb{C}$: the element $\sum_{g\in G} c_g|g\rangle$ can be thought of as the function that maps $g \in G$ to c_g. Functions from G to $\mathbb{C}$ that are constant on conjugacy classes of G, i.e. they satisfy $f(g) = f(hgh^{-1})$ for all $g, h \in G$, are called *class functions.* The set of class functions is exactly the center $Z(\mathbb{C}[G]) = \{f \in \mathbb{C}[G] : fg = gf \text{ for all } g \in \mathbb{C}[G]\}$ of the group algebra.

Definition 13. Let (φ, V) be a representation of G. The *character* $\chi = \chi_V$ of (φ, V) is the class function defined by $\chi(g) = \mathrm{tr}(\varphi(g))$.

Here we list basic properties of the character of a representation. The character of a representation evaluated at the identity of the group gives the dimension of the representation. If (φ, V) is unitary, then $\chi(g^{-1}) = \overline{\chi(g)}$. The character $\chi_{V\oplus W}$ of a direct sum of two representations V and W is the sum of the individual characters χ_V and χ_W. The character $\chi_{V\otimes W}$ of a tensor product of two representations V and W is the product of the individual characters χ_V and χ_W.

In the following proposition we will use an inner product structure on $\mathbb{C}[G]$, which we now define. For $x = \sum_{g\in G} x_g|g\rangle, y = \sum_{g\in G} y_g|g\rangle$ in $\mathbb{C}[G]$, set $(x, y) := \frac{1}{|G|}\sum_{g\in G} \overline{x_g}y_g$.

Proposition 2.1. Let W_i for $i = 1, ..., k$ be pairwise independent irreducible representations of a group G, and denote by χ_i the corresponding characters. Then $(\chi_i, \chi_j) = \delta_{ij}$. Moreover, any class function orthogonal to all $\chi_i's$ is identically 0. Hence, $\{\chi_i\}_{i=1}^k$ is an orthonormal basis of the set of class functions.

Proof. See [Ser77], [Tel05] for a proof. ■

Corollary 2. Proposition 2.1 has the following corollaries.

(1) The multiplicity of an irreducible representation W in some representation V is (χ_V, χ_W).
(2) V is irreducible iff $(\chi_V, \chi_V) = 1$.
(3) Two representations are isomorphic iff they have the same character.
(4) The number of distinct irreducible representations of a finite group G is equal to the number of conjugacy classes.

We now determine the character of the regular representation $R(G)$ of a finite group G. Recall that $R(G)$ has basis $\{|g\rangle\}_{g\in G}$ and G acts by left multiplication $\varphi(g) : |h\rangle \mapsto |gh\rangle$. Therefore $\chi_{R(G)}(g) = \mathrm{tr}\, \varphi(g) = |G|\delta_{g,e}$.

Corollary 3. Combining the previous corollary with the character of the regular representation yields:

(1) The multiplicity of any irreducible representation in the regular representation is equal to its dimension.
(2) Let $W_1, ..., W_k$ be a complete list of irreducible representations of G. Then every W_i appears in $R(G)$ and

$$\sum_{i=1}^{k}(\dim W_i)^2 = |G|.$$

Proposition 2.2. Let (φ, V) be a representation of G and W be a fixed irreducible representation with character χ_W. Then the projection onto the isotypical component of W in V is

$$P_W = \frac{\dim W}{|G|}\sum_{g\in G}\overline{\chi_W(g)}\varphi(g).$$

In particular, $P = \frac{1}{|G|}\sum_{g\in G}\varphi(g)$ projects onto $V^G = \{|v\rangle \in V : \varphi(g)|v\rangle = |v\rangle \text{ for all } g \in G\}$.

2.3. From finite and compact groups

Definition 14. A *topological group* is a group G endowed with a topology such that group multiplication and inversion are continuous. A *compact group* is a topological group that is compact, that is, every open cover of G has a finite subcover. Closed subgroups of a compact group are also compact groups.

Definition 15. A *representation* (φ, V) of a topological group G on a normed, finite-dimensional vector space V is a continuous group homomorphism $\varphi : G \to \mathrm{GL}(V)$.

Recall that the averaging operation $\frac{1}{|G|}\sum_{g\in G}$ over a finite group was essential for proving Maschke's theorem, character formulas etc.
For compact groups we can replace this discrete averaging by a suitable integral to recover many of the previous results for finite group also for compact groups.

Proposition 2.3. Let G be a compact group. There exists a unique measure $\mathrm{d}g$ on G, called the *Haar measure*, satisfying:

(1) Invariance: for every continuous function $f : G \to \mathbb{C}$ and every $h \in G$,

$$\int_G f(g)\,\mathrm{d}g = \int_G f(gh)\,\mathrm{d}g = \int_G f(hg)\,\mathrm{d}g\,.$$

(2) Normalization: $\int_G 1\,\mathrm{d}g = 1$.

Example 8. Every finite group with the discrete topology is a compact group, and $\mathrm{d}g = \frac{1}{|G|}$, $\int_G \mathrm{d}g \approx \frac{1}{|G|}\sum_{g\in G}$.

Example 9. The circle group $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\} = \{\exp(i\theta) : \theta \in [0, 2\pi)\}$ has Haar measure $\mathrm{d}g = \frac{1}{2\pi}\,\mathrm{d}\theta$.

Using the Haar measure, one can prove analogous statements about finite-dimensional representations of compact groups, e.g.:

(1) Every G-invariant subspace has a G-invariant complement.
(2) Every representation decomposes as a sum of irreducible representations.
(3) Most aspects of character theory also carry over to the compact case (note however that if G is an infinite compact group, then expressions involving $|G|$ may no longer be valid).

The regular representation of a compact group G is defined as the Hilbert space $L^2(G)$ of square integrable functions on G, with the action of G given by $\varphi(g)(f)(h) = f(g^{-1}h)$. If $|G| = \infty$, then $\dim L^2(G) = \infty$. We have the following theorem about the decomposition of the regular representation for compact groups.

Proposition 2.4. Let G be a compact group.

(1) The linear span of all matrix coefficients of the irreducible unitary representations of G is dense in $L^2(G)$.
(2) Every irreducible unitary representation of G is finite-dimensional.
(3) The regular representation (which has infinite dimension if G is not finite) $L^2(G)$ decomposes into a direct sum of the irreducible unitary representations of G, each occurring with multiplicity equal to its dimension. The matrix coefficients of the complete set of irreps form an orthonormal basis of $L^2(G)$.

Proof. See Theorem 1.12, [Kna16]. ■

3. Schur-Weyl duality

3.1. Representations of direct product groups

Definition 16. Let G and H be groups. The *direct product* $G \times H$ of G and H is a group; the underlying set is $G \times H = \{(g, h) : g \in G, h \in H\}$ with multiplication defined as $(g_1, h_1) \cdot (g_2, h_2) = (g_1 g_2, h_1 h_2)$ for all $g_1, g_2 \in G$ and $h_1, h_2 \in H$.

Definition 17. Let (φ, V) and (ψ, W) be representations of groups G and H respectively. Then $V \widehat{\otimes} W$ affords the *external product representation* of the direct product $G \times H$ by defining $(\varphi \widehat{\otimes} \psi)(g, h) := \varphi(g) \otimes \psi(h)$, where the notation $V \widehat{\otimes} W$ is used to distinguish it from the tensor representation $V \otimes W$ from Section 2.1.

Remark.

(1) If (φ, V) and (ψ, W) are irreducible, then so is $(\varphi \widehat{\otimes} \psi, V \widehat{\otimes} W)$.

(2) Every irreducible representation of $G \times H$ arises this way.

3.2. Double commutant theorem

Definition 18. Let $\mathcal{A}$ be a subset of an algebra $\mathcal{C}$. The commutant $\mathcal{A}'$ of $\mathcal{A}$ is the collection of all elements in $\mathcal{C}$ commuting with all of cA:

$$\mathcal{A}' = \{b \in \mathcal{C} : ab = ba \text{ for all } a \in A\}.$$

For a vector space V, the algebra of operators acting on V is denoted $\mathrm{End}(V)$.

Lemma 3.1. Let V and W be finite-dimensional complex vector spaces. The commutant of $\mathrm{End}(V) \otimes \mathbb{1}_W$ in $\mathrm{End}(V \otimes W) \cong \mathrm{End}(V) \otimes \mathrm{End}(W)$ is $\mathbb{1}_W \otimes \mathrm{End}(W)$.

Proof. Set $\mathcal{A} = \mathrm{End}(V) \otimes \mathbb{1}_W$ and $\mathcal{B} = \mathbb{1}_V \otimes \mathrm{End}(W)$. Clearly, an element $\mathbb{1}_V \otimes b \in \mathcal{B}$ commutes with every element $a \otimes \mathbb{1}_W \in \mathcal{A}$, and hence $\mathcal{B} \subset \mathcal{A}'$.

Suppose now that $a \otimes \mathbb{1}_W \in \mathcal{A}$ and $\widetilde{a} \in \mathcal{A}'$ are arbitrary. Let $\dim W = n$ and write

$$a \otimes \mathbb{1}_W = \begin{pmatrix} a & 0 & \cdots & 0 \\ 0 & a & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & a \end{pmatrix} \quad \text{and} \quad \widetilde{a} = \begin{pmatrix} \widetilde{a_{11}} & \widetilde{a_{12}} & \cdots & \widetilde{a_{1n}} \\ \widetilde{a_{21}} & \widetilde{a_{22}} & \cdots & \widetilde{a_{2n}} \\ \vdots & & \ddots & \vdots \\ \widetilde{a_{n1}} & \widetilde{a_{n2}} & \cdots & \widetilde{a_{nn}} \end{pmatrix}.$$

Then

$$(a \otimes \mathbb{1}_W)\widetilde{a} = \begin{pmatrix} a\widetilde{a_{11}} & a\widetilde{a_{12}} & \cdots & a\widetilde{a_{1n}} \\ a\widetilde{a_{21}} & a\widetilde{a_{22}} & \cdots & a\widetilde{a_{2n}} \\ \vdots & & \ddots & \vdots \\ a\widetilde{a_{n1}} & a\widetilde{a_{n2}} & \cdots & a\widetilde{a_{nn}} \end{pmatrix} \tag{1}$$

$$= \begin{pmatrix} \widetilde{a_{11}}a & \widetilde{a_{12}}a & \cdots & \widetilde{a_{1n}}a \\ \widetilde{a_{21}}a & \widetilde{a_{22}}a & \cdots & \widetilde{a_{2n}}a \\ \vdots & & \ddots & \vdots \\ \widetilde{a_{n1}}a & \widetilde{a_{n2}}a & \cdots & \widetilde{a_{nn}}a \end{pmatrix} \tag{2}$$

$$= \widetilde{a}(a \otimes \mathbb{1}_W). \tag{3}$$

Hence, for fixed i, j, we have $[a, \widetilde{a_{ij}}] = 0$ for all $a \in \mathrm{End}(V)$, and so $\widetilde{a_{ij}} = \lambda_{ij}\mathbb{1}_A$ for some $\lambda_{ij} \in \mathbb{C}$. Let $b \in \mathrm{End}(W)$ be defined by $(b)_{ij} = \lambda_{ij}$, then $\widetilde{a} = \mathbb{1}_W \otimes b \in \mathbb{1}_A \otimes \mathrm{End}(W) = \mathcal{B}$, and thus $\mathcal{A}' \subset \mathcal{B}$. ■

We can now the prove the double commutant theorem.

Proposition 3.2. Let (φ, V) be a representation of a finite group G with decomposition $V = \oplus_\alpha V_\alpha \otimes \mathbb{C}^{n_\alpha}$ into pairwise inequivalent irreducible representations V_α with multiplicity n_α. Let $\mathcal{A} \subset \mathrm{End}(V)$ be the subalgebra generated by φ, and set $\mathcal{B} = \mathcal{A}'$. Then:

(1) $\mathcal{A} \cong \oplus_\alpha \mathrm{End}(V_\alpha) \otimes \mathbb{1}_{\mathbb{C}^{n_\alpha}}$

(2) $\mathcal{B} \cong \oplus_\alpha \mathbb{1}_{V_\alpha} \otimes \mathrm{End}(\mathbb{C}^{n_\alpha})$

(3) $\mathcal{B}' = (\mathcal{A}')' = \mathcal{A}$

Proof. Let $(\varphi_\alpha, V_\alpha)$ be the irreducible representations appearing in (φ, V) and set $d_\alpha = \dim V_\alpha$.

(1) An application of Schur's lemma ([Ser77], Sec 2.2) shows

$$E_{ij}^{(\alpha)} \otimes \mathbb{1}_{\mathbb{C}^{n_\alpha}} = d_\alpha \sum_{g \in G} \overline{\varphi_\alpha(g)_{ij}} \varphi(g) \in \mathcal{A},$$

where $\varphi_\alpha(g)_{ij}$ is the (i,j)-elementary matrix in $\mathrm{End}(V_\alpha)$. Since the E^α_{ij} is the (i,j)-elementary matrix in $\mathrm{End}(V_\alpha)$. Since the E^α_{ij} are a basis of $\mathrm{End}(V_\alpha)$, we have

$$\oplus_\alpha \mathrm{End}(V_\alpha) \otimes \mathbb{1}_{\mathbb{C}^{n_\alpha}} \subset \mathcal{A}.$$

The reverse inclusion follows by the decomposition of V into isotypical components $V_\alpha \otimes \mathbb{C}^{n_\alpha}$, and hence we have equality.

(2) First we show

$$\mathcal{B} \subset \oplus_\alpha \mathbb{1}_{V_\alpha} \otimes \mathrm{End}(\mathbb{C}^{n_\alpha}).$$

Let P_α be the projection onto V_α, that is, $P_\alpha \mathcal{A} = V_\alpha \otimes \mathbb{C}^{n_\alpha}$. Then every $b \in \mathcal{B}$ commutes with P_α by definition, and hence $b = \mathbb{1}_{cA} b = \sum_\alpha P_\alpha b = \sum_\alpha P_\alpha b P_\alpha = \sum_\alpha b_\alpha$, where $b_\alpha \in \mathrm{End}(V_\alpha \otimes \mathbb{C}^{n_\alpha})$. By the preceding lemma, $b_\alpha = \mathbb{1}_{V_\alpha} \otimes \widetilde{b_\alpha}$ for some $\widetilde{b_\alpha} \in \mathrm{End}(\mathbb{C}^{n_\alpha})$, as was to be shown. The other inclusion holds since for $b_\alpha \in \mathrm{End}(\mathbb{C}^{n_\alpha})$, $\oplus_\alpha \mathbb{1}_{V_\alpha} \otimes b_\alpha$ commutes with any $\oplus_\alpha a_\alpha \otimes \mathbb{1}_{\mathbb{C}^{n_\alpha}} \in \mathcal{A}$.

(3) This follows by a similar argument to (2). ■

3.3. Schur-Weyl duality

We now focus on the following two groups:

- the symmetric group S_n, which is the set of bijections from $\{1, ..., n\}$ to itself.
- the unitary group $\mathrm{U}_d = \{U \in \mathcal{L}(\mathbb{C}^d) : U^\dagger U = UU^\dagger = \mathbb{1}_d\}$.

The symmetric group has a representation on $(\mathbb{C}^d)^{\otimes n}$ with the action of $\pi \in S_n$ given by the map φ defined on a basis of $(\mathbb{C}^d)^{\otimes n}$ as:

$$\varphi(\pi)(|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle) = |\psi_{\pi^{-1}(1)}\rangle \otimes \cdots \otimes |\psi_{\pi^{-1}(n)}\rangle.$$

The unitary group also has a representation on $(\mathbb{C}^d)^{\otimes n}$ with the action of $U \in \mathrm{U}_d$ given by the map ω defined on a basis of $(\mathbb{C}^d)^{\otimes n}$ as:

$$\omega(U)(|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle) = |\psi_{\pi^{-1}(1)}\rangle \otimes \cdots \otimes |\psi_{\pi^{-1}(n)}\rangle.$$

Definition 19. The *symmetric subspace* $\mathrm{Sym}^n(V)$, also called the *n-th symmetric power* of V, is the subspace

$$\mathrm{Sym}^n(V) = (V^{\otimes n})^{S_n} = \{|v\rangle \in V^{\otimes n} : \varphi(\pi)|v\rangle = |v\rangle \text{ for all } \pi \in S_n\}.$$

With $P = \frac{1}{n!}\sum_{\pi \in S_n} \varphi(\pi)$, we have $\mathrm{Sym}^n(V) = PV^{\otimes n}$.

Lemma 3.3. $\mathrm{Sym}^n(V) = \mathrm{span}\{|v\rangle^{\otimes n} : |v\rangle \in V\}$.

Proof. Let $\{|e_i\rangle\}_{i=1}^d$ be an orthonormal basis for V, with $d = \dim V$. By definition $\mathrm{Sym}^n(V)$ is spanned by the vectors

$$|v_{i_1\cdots i_n}\rangle := \sum_{\pi \in S_n} \varphi(\pi)(|e_{i_1}\rangle \otimes \cdots \otimes |e_{i_n}\rangle) \tag{4}$$

$$= \sum_{\pi \in S_n} |e_{i_{\pi^{-1}(1)}}\rangle \otimes \cdots \otimes |e_{i_{\pi^{-1}(n)}}\rangle \tag{5}$$

for indices $i_j \in \{1, ..., d\}$ $j = 1, ..., n$. This proves $\mathrm{span}\{|v\rangle^{\otimes n} : |v\rangle \in V\} \in \mathrm{Sym}^n(V)$.
To show the other inclusion, we rewrite the vectors $|v_{i_1\cdots i_n}\rangle$ using derivatives as

$$|v_{i_1\cdots i_n}\rangle = \partial_{\lambda_2} \cdots \partial_{\lambda_n} (|e_{i_1}\rangle + \sum_{j=2}^{n} \lambda_j |e_{i_j}\rangle)^{\otimes n}\Big|_{\lambda_2 = \cdots = \lambda_n = 0}.$$

Using the definition of the derivative, we have

$$\partial_{\lambda_j}(|e_1\rangle + \lambda_j|e_j\rangle)^{\otimes n}\Big|_{\lambda_j = 0} = \lim_{\lambda_j \to 0} \frac{(|e_1\rangle + \lambda_j|e_j\rangle)^{|\otimes n\rangle} - |e_1\rangle^{\otimes n}}{\lambda_j}.$$

The $|v_{i_1\cdots i_n}\rangle$ are limits of elements in $W = \mathrm{span}\{|v\rangle^{\otimes n} : |v\rangle \in V\}$, and since W is finite-dimensional and thus closed in $\mathrm{Sym}^n(V)$, we have $|v_{i_1\cdots i_n}\rangle \in W$ for all indices $i_1, ..., i_n$. It follows that $\mathrm{Sym}^n(V) \subset W$. ■

Corollary 4. Let $C \in \mathrm{End}(V^{\otimes n})$ be such that $\varphi(\pi) \subset \varphi(\pi)^\dagger = C$ for all $\pi \in S_n$. Then $C \in \mathrm{span}\{X^{\otimes n} : X \in \mathrm{End}(V)\}$.

Proof. Let $W = \mathrm{End}(V^{\otimes n}) \cong \mathrm{End}(V)^{\otimes n}$ and let $\{|e_i\rangle\}_{i=1}^d$ be a fixed basis of V. Consider the basis $\{E_{ij}\}_{i,j=1}^d$ of $\mathrm{End}(V)$, where $E_{ij} : |e_k\rangle \mapsto \delta_{jk}|e_i\rangle$. Denote by $\varphi : S_n \to \mathrm{GL}(V^{\otimes n})$ then tensor representation of S_n on $V^{\otimes n}$ and by $\widetilde{\varphi} : S_n \to \mathrm{GL}(W)$ the analogous tensor representation of S_n on $W = \mathrm{End}(V)^{\otimes n}$. Then $\widetilde{\varphi}(\pi)$ acting on $X \in \mathrm{End}(V^{\otimes n})$ has the matrix representation $\varphi(\pi)X\varphi(\pi)^{-1}$. The claim then follows from the preceding lemma applied to $(\widetilde{\varphi}, W)$. ■

In what follows we view $\omega : X \mapsto X^{\otimes n}$ as a representation of $\mathrm{GL}(V) = \{X \in \mathrm{End}(V) : X \text{ is invertible}\}$.

Proposition 3.4. A representation of U_d ($d = \dim V$) is irreducible if and only if the corresponding representation of $\mathrm{GL}(V)$ is irreducible.

Proof. For a proof, see [Alc18]. ■

Proposition 3.5. S_n and $\mathrm{GL}(V)$ span each other's commutants in $\mathrm{End}(V^{\otimes n})$.

Proof. Let $\mathcal{A} \subset \mathrm{End}(V^{\otimes n})$ be the subalgebra generated by $\varphi(\pi), \pi \in S_n$ and let $\mathcal{B} \subset \mathrm{End}(V^{\otimes n})$ be the subalgebra generated by $\omega(g), g \in \mathrm{GL}(V)$. Since $\varphi(\pi)$ and $\omega(U)$ commute for all $\pi \in S_n, U \in \mathrm{U}_d$, we have $\mathcal{B} \subset \mathcal{A}'$. The previous corollary shows that $\mathcal{A}' = \mathrm{span}\{X^{\otimes n} : X \in \mathrm{End}(V)\}$. Let $X \in \mathrm{End}(V)$, then $X + t\mathbb{1}$ is invertible for all but finitely many t, and so $(X + t\mathbb{1})^{\otimes n} \in \mathcal{B}$ for all but finitely many t. But $(X + t\mathbb{1})^{\otimes n}$ is a polynomial in t of degree n, and by Lagrange's interpolation theorem determined by any $n + 1$ distinct points. Hence, $(X + t\mathbb{1})^{\otimes n} \in \mathcal{B}$ for all t, in particular for $t = 0$. It follows that $\mathcal{A}' = \mathrm{span}\{X^{\otimes n} : X \in \mathrm{End}(V)\} \in \mathcal{B}$, hence $\mathcal{A}' = \mathcal{B}$. The Double Commutant theorem now implies $\mathcal{B}' = \mathcal{A}$, which concludes the proof. ■

Proposition 3.6. Let $V = \mathbb{C}^d$ and $(\varphi, V^{\otimes n})$ and $(\omega, V^{\otimes})$ be the tensor representations of S_n and $\mathrm{GL}(V)$ defined above. As a representation of $S_n \times \mathrm{GL}(V)$, the space $V^{\otimes n}$ decomposes as

$$V^{\otimes n} = \bigoplus_\lambda V_\lambda \otimes U_\lambda,$$

where $(\varphi_\lambda, V_\lambda)$ and $(\omega_\lambda, U_\lambda)$ are inequivalent irreducible representations of S_n and $\mathrm{GL}(V)$ respectively and

$$\varphi(\pi) = \bigoplus_\lambda \varphi_\lambda(\pi) \otimes \mathbb{1}_{U_\lambda} \qquad \text{for } \pi \in S_n \tag{6}$$

$$\omega(g) = \bigoplus_\lambda \mathbb{1}_{V_\lambda} \otimes \omega_\lambda(g) \qquad \text{for } g \in \mathrm{GL}(V). \tag{7}$$

The same assertion holds when $\mathrm{GL}(V)$ is replaced with U_d.

Proof. The decomposition of $V^{\otimes n}$ follows from the Double Commutant Theorem and the fact that S_n and $\mathrm{GL}(V)$ span each other's commutant. It remains to show that $U_\lambda \cong \mathrm{Hom}_{S_n}(V_\lambda, V^{\otimes n})$ is an irreducible representation of $\mathrm{GL}(V)$ (or U_d). By Schur's lemma, this is equivalent to showing that

$$\mathrm{End}_{\mathrm{GL}(V)}(U_\lambda) := \mathrm{Hom}_{\mathrm{GL}(V)}(U_\lambda, U_\lambda) \cong \mathbb{C}.$$

We have $Z(\mathrm{End}(U_\lambda)) \cong \mathbb{C}$. Schur's lemma and the above decomposition show that

$$\mathrm{End}_{S_n}(V^{\otimes n}) \cong \oplus_\lambda \mathrm{End}(U_\lambda) \tag{8}$$

$$\mathrm{End}_{\mathrm{GL}(V)\times S_n}(V^{\otimes n}) \cong \oplus_\lambda \mathrm{End}_{\mathrm{GL}(V)}(U_\lambda). \tag{9}$$

Since $\mathrm{End}_{S_n}(V^{\otimes n}) = \mathrm{span}\{X^{\otimes n} : X \in \mathrm{GL}(V)\}$, we have

$$\mathrm{End}_{\mathrm{GL}(V)\times S_n}(V^{\otimes n}) \subset Z(\mathrm{End}_{S_n}(V^{\otimes n})),$$

and hence also $\mathrm{End}_{\mathrm{GL}(V)}(V^{\otimes n}) \subset Z(\mathrm{End}(U_\lambda)) \cong \mathbb{C}$. ■

Summary. Schur-Weyl duality says that

$$V^{\otimes n} \cong \bigoplus_{\lambda} V_\lambda \otimes U_\lambda$$

as a representation of $S_n \times \mathrm{U}_d$, with V_λ and U_λ being irreps of S_n and U_d respectively.
The next chapter is a discussion of the index λ and the irreps V_λ, U_λ.

4. Irreps of symmetric and unitary groups

4.1. Minimal projections and irreducible representations

Recall that for a given finite group G, the group algebra $\mathbb{C}[G]$ was defined as the $\mathbb{C}$-vector space with basis $\{|g\rangle\}_{g\in G}$ and multiplication

$$\left(\sum_{g\in G} c_g|g\rangle\right) \cdot \left(\sum_{[} h \in G]d_h|h\rangle\right) = \sum_{g,h\in G} c_g d_h |gh\rangle.$$

Definition 20. A *projection* in $\mathbb{C}[G]$ is an element $p \in \mathbb{C}[G]$ with $p^2 = p$. A non-zero projection p is called *minimal*, if there are no non-zero projections q, r such that $p = q + r$. Two projections p and q are equivalent if there are invertible elements $x, y \in \mathbb{C}[G]$ such that $xpy = q$, and disjoint if $pzq = 0$ for all $z \in \mathbb{C}[G]$.

Definition 21. A *central projection* in $\mathbb{C}[G]$ is a projection in $Z(\mathbb{C}[G]) = \{x \in \mathbb{C}[G] : xy = yx \text{ for all } y \in \mathbb{C}[G]\}$. A non-zero central projection is called minimal if it cannot be written as a sum of non-zero central projections.

Proposition 4.1. Let G be a finite group with group algebra $\mathcal{A} = \mathbb{C}[G]$. Irreducible representations of G are in one-to-one correspondence with:

- equivalence classes of minimal projections in cA.
- minimal central projections in $\mathcal{A}$.

Let $(\varphi_\alpha, V_\alpha)$ be an irreducible representation of G with character $\chi_\alpha(g) = \operatorname{tr} \varphi_\alpha(g)$. Then

$$P_\alpha = \frac{\dim V_\alpha}{|G|} \chi_\alpha$$

is the minimal central projection corresponding to $(\varphi_\alpha, V_\alpha)$.

Proof idea. Use the fact that $\mathbb{C}[G] \cong \oplus_\alpha \operatorname{End}(\mathbb{C}^{d_\alpha})$, where α runs through the irreducible representations, and that the centre $\operatorname{End}(\mathbb{C}^{d_\alpha})$ is 1-dimensional and spanned by χ_α. ■

Corollary 5. Let (φ, V) be a representation of a finite group G with isotypical decomposition $V \cong \oplus_\alpha V_\alpha$ and $V_\alpha = W_\alpha \oplus \cdots \oplus W_\alpha$ for inequivalent irreducible representations W_α of G. Let χ_α be the character of W_α. Then

$$\pi_\alpha = \frac{\dim W_\alpha}{|G|} \sum_{g\in G} \overline{\chi_\alpha(g)} \varphi(g)$$

projects onto the isotypical component V_α of V.

4.2. Conjugacy classes of the symmetric group

Recall from character theory that for a finite group G, the number of irreducible representations of G is equal to the number of conjugacy classes of G.
The relation $\sim$ on G defined as: $g \sim h$ if and only if there exists $s \in G$ such that $g = shs^{-1}$ is an equivalence relation. The equivalence classes $C_1, \ldots, C_k$ are called *conjugacy classes* and these form a partition of G, so $G = \sqcup_{i=1}^k C_i$.
Facts about permutations:

(1) Every permutation $\pi \in S_n$ can be written uniquely as a product of disjoint cycles, e.g., $\pi = (13)(2)(465) \in S_6$. The *cycle type* of a permutation $\pi \in S_n$ is the tuple of cycle lengths in non-increasing order. For example $\pi = (14)(236)(58)(7)$ has cycle type $(3, 2, 2, 1)$.

(2) Cycle types $(\lambda_1, ..., \lambda_d)$ of a permutation $\pi \in S_n$ form an ordered partition of n: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$ and $\sum_{i=1}^d \lambda_i = n$.
We use the notation $\lambda \vdash_d n$ for an ordered partition of n into at most d parts.
Note: If $d < n$ then not all possible partitions or cycle types appear.

(3) Two permutations π, π' are conjugate iff they have the same cycle type: let $(i_1, ..., i_k)$ be a cycle of length k and $\sigma \in S_n$ be arbitrary $(k \leq n)$, then $\sigma(i_1, ..., i_k)\sigma^{-1} = (\sigma(i_1), ..., \sigma(i_k))$.

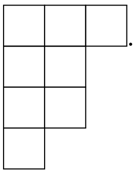
(4) It follows from $(1) - (3)$ above that the conjugacy classes of S_n, and hence its irreducible representations, are indexed by the ordered partitions of n into n parts.

4.3. Constructing the irreps of S_n and U_d

Irreducible representations of S_n are in bijection with ordered partitions of n.

Definition 22. Let $\lambda \vdash_d n$ be a partition of n into at most d parts. The *Young diagram* corresponding to $\lambda \vdash_d n$ is an arrangement of n boxes into d rows such that the i-th row has length λ_i.

For example $\lambda = (3, 2, 2, 1) \vdash_4 8$ has Young diagram:



A *Young tableau* is a Young diagram where boxes are labeled with numbers $\{1, ..., N\}$ where $N \neq n$ in general.
A *standard Young tableau* is a Young tableau with $N = n$ where the labels are increasing along rows (left to right) and along columns (top to bottom).
A *semistandard Young tableau* is a Young tableau whose labels are non-decreasing along rows and increasing along columns.

Example 10. Standard Young tableaux:

$$\begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4&5\\\cline{1-2}\end{array}\qquad\begin{array}{|c|c|c|}\hline 1&2&4\\\hline 3&5\\\cline{1-2}\end{array}\qquad\begin{array}{|c|c|c|}\hline 1&2&5\\\hline 3&4\\\cline{1-2}\end{array}\qquad\begin{array}{|c|c|c|}\hline 1&3&4\\\hline 2&5\\\cline{1-2}\end{array}\qquad\begin{array}{|c|c|c|}\hline 1&3&5\\\hline 2&4\\\cline{1-2}\end{array}$$

Semistandard Young tableaux with numbering $\{1, 2\}$:

$$\begin{array}{|c|c|c|}\hline 1&1&1\\\hline 2&2\\\cline{1-2}\end{array}\qquad\begin{array}{|c|c|c|}\hline 1&1&2\\\hline 2&2\\\cline{1-2}\end{array}$$

Schur-Weyl duality gives

$$\left(\mathbb{C}^d\right)^{\otimes n} = \bigoplus_{\lambda \vdash_d n} V_\lambda \otimes U_\lambda,$$

where

- the irrep V_λ of S_n has an orthonormal basis index by the set of standard Young tableaux of shape $\lambda \vdash_d n$.
- the irrep U_λ of U_d has an orthonormal basis indexed by the set of semistandard Young tableaux of shape $\lambda \vdash_d n$ and numbering $\{1, ..., d\}$.

Recall that every permutation $\pi \in S_n$ can be written as a product of at most $n - 1$ transpositions (jk) with $1 \leq j < k \leq n$.

Definition 23. Write $\pi = \tau_1 \cdots \tau_k \in S_n$ for transpositions τ_i. The *sign* of π is defined as $\mathrm{sgn}(\pi) = (-1)^k$.

Let T be a standard Young tableau of shape $\lambda \vdash_d n$. Define two subgroups R_T, C_T of S_n as

$$R_T := \{\pi \in S_n : \pi \text{ permutes integers within rows of } T\}$$

$$C_T := \{\pi \in S_n : \pi \text{ permutes integers within columns of } T\}.$$

Example 11. The table below lists standard Young tableau T of shape $\lambda \vdash_3 6$ and the corresponding groups R_T, C_T.

T	$\begin{ytableau} 1 & 2 & 3 \end{ytableau}$	$\begin{ytableau} 1 & 2 \\ 3 \end{ytableau}$	$\begin{ytableau} 1 & 3 \\ 2 \end{ytableau}$	$\begin{ytableau} 1 \\ 2 \\ 3 \end{ytableau}$
R_T	S_3	$\{e, (12)(3)\} \cong S_2$	$\{e, (13)(2)\} \cong S_2$	$\{e\} \cong S_1$
C_T	$\{e\} \cong S_1$	$\{e, (13)(2)\} \cong S_2$	$\{e, (12)(3)\} \cong S_2$	S_3

We define two elements in $\mathbb{C}[S_n]$:

$$r_T := \sum_{\pi \in R_T} \pi \qquad c_T := \sum_{\pi \in C_T} \operatorname{sgn}(\pi)\pi.$$

Definition 24. For given standard Young tableaux T of shape $\lambda \vdash n$, the *Young symmetrizer* e_T is defined as $e_T := r_T c_T$.

Example 12. For $\lambda = (n) \vdash n$, we have $c_T = \{e\}$, $R_T = S_n$ and $e_T = \sum_{\pi \in S_n} \pi$. For $\lambda = (1, ..., 1)$, we have $e_T = \sum_{\pi \in S_n} \operatorname{sgn}(\pi)\pi$.

Proposition 4.2. Let T be a Young tableau of shape $\lambda \vdash n$, and let e_T be the corresponding Young symmetrizer. Then $f_T := \frac{d_\lambda}{n!} e_T$ is the minimal projection in $\mathbb{C}[S_n]$ corresponding to the irreducible representation V_λ of S_n, that is, $V_\lambda \cong \mathbb{C}[S_n]e_T$. Here,

$$d_\lambda := \dim V_\lambda \frac{n!}{\prod_{(i,j) \in \lambda} h(i,j)}$$

where for a box (i, j) in row i and column j of λ we define the *hook length*

$$\begin{aligned} h(i,j) = {} & \text{number of boxes to the right of } (i,j) && (10) \\ & + \text{number of boxes below } (i,j) && (11) \\ & + 1. && (12) \end{aligned}$$

The V_λ are called *Specht modules*. Every irreducible representation of S_n is isomorphic to a Specht module V_λ for some $\lambda \vdash n$, and $V_\lambda \ncong V_{\lambda'}$ for $\lambda \neq \lambda'$.

Proof. See [Chr06] or [Alc18]. ■

Example 13. For the Young diagram λ given by

$$\ydiagram{4,2,2,1},$$

the hook length of the box at $(1, 2)$ is $h(1, 2) = 3 + 2 + 1 = 6$. The dimension of the corresponding irreducible representation V_λ is $d_\lambda = \frac{10!}{8 \cdot 6 \cdot 3 \cdot 2 \cdot 4 \cdot 2 \cdot 3} = 25 \cdot 21 = 525$.

Proposition 4.3. Let $|v\rangle \in V^{\otimes n}$, and for a standard Young tableau T of shape $\lambda \vdash n$. Consider the Young symmetrizer e_T. Let p be the number of parts of the partition λ (or the number of non-zero rows of the Young diagram λ).

- If $p \leq d = \dim V$, then $\mathbb{C}[S_n]e_T|v\rangle$ is an irreducible representation of S_n isomorphic to the Specht module V_λ.
- If $p \leq d$, then $e_T V^{\otimes n}$ is an irreducible representation of $\mathrm{GL}(V)$ (or U_d) on $V^{\otimes n}$. These are inequivalent Young tableaux of different shape.

- Using the above, we have the Schur-Weyl decomposition of $V^{\otimes n}$ with $d = \dim V$ as an $S_n \times \mathrm{U}_d$ representation:

$$V^{\otimes n} = \bigoplus_{\lambda \vdash_d n} V_\lambda \otimes U_\lambda.$$

Proof. See [Chr06]. ■

The dimensions of V_λ and U_λ are given by:

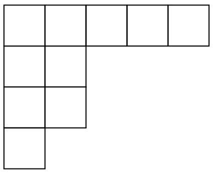
$$d_\lambda = \dim V_\lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h(i,j)}$$

$$m_\lambda = \dim U_\lambda = \prod_{1\leq i<j\leq d} \frac{\lambda_i - \lambda_j + j - i}{j - i}.$$

The Schur-Weyl decomposition shows that

$$d^n = \sum_{\lambda \vdash_d n} d_\lambda m_\lambda.$$

Example 14. If λ is the Young diagram



then

$$d_\lambda = 525$$

and

$$m_\lambda = \frac{3+1}{1} \cdot \frac{3+2}{2} \cdot \frac{4+3}{3} \cdot \frac{0+1}{1} \cdot \frac{1+2}{2} \cdot \frac{1+1}{1} = 70.$$

5. Mathematics of finite dimensional quantum information theory

5.1. Quantum systems and quantum states

A *quantum system* is a physical system with one or more quantum-mechanical degrees of freedom that are either discrete or continuous:

- position and momentum of a particle
- spin of a particle (e.g. spin along z-axis of an electron)
- polarization of a photon

The motivating example we will use is that of the spin of an electron. There are two possible “basis states”: spin up (↑) and spin down (↓). Each of these is assigned a vector in the *state space* $\mathbb{C}^2$:

$$|{\uparrow}\rangle = \begin{pmatrix}1\\0\end{pmatrix} \qquad |{\downarrow}\rangle = \begin{pmatrix}0\\1\end{pmatrix}.$$

The *superposition principle* states that a quantum state can be prepared in a state $|\psi\rangle = \alpha|{\uparrow}\rangle + \beta|{\downarrow}\rangle$, where $\alpha, \beta \in \mathbb{C}$ satisfy $|\alpha|^2 + |\beta|^2 = 1$. The probabilities of finding electron in spin-up or spin-down are given $\Pr(\uparrow) = |\langle \uparrow |\psi\rangle|^2 = |\alpha|^2$ and $\Pr(\downarrow) = |\langle \downarrow |\psi\rangle|^2 = |\beta|^2$.

More formally:

(1) The *state space* describing a quantum system is given by a *Hilbert space*, a complex inner-product space that is complete. We restrict our attention to finite-dimensional Hilbert spaces $\mathcal{H} = \mathbb{C}^d$.

(2) Observable quantities are represented by *Hermitian operators* $\mathcal{A} \in \{X \in \mathcal{L}(\mathcal{H}) : X^\dagger = X\}$. The real eigenvalues of A can be measured in an experiment.

(3) A state of a quantum system assigns an expectation value to observables, that is, it describes the expected measurement statistics of an observable in a quantum system.
We identify states with *density operators* $\rho \in \mathcal{L}(\mathcal{H})$ satisfying:

- positivity: $\rho \geq 0$, i.e., $\langle\varphi|\rho|\varphi\rangle \geq 0$ for all $|\varphi\rangle \in \mathcal{H}$.
- *normalization:* $\operatorname{tr}\rho = 1$.

The expectation of an observable A w.r.t. a state ρ is given by

$$\langle\varphi|\rho|\varphi\rangle\,.$$

The set of density matrices of a finite-dimensional Hilbert space is convex and compact. That is, if ρ_i are density matrices and λ_i probabilities then $\rho = \sum_i \lambda_i \rho_i$ is also a density matrix.

(4) A *pure state* is an extreme point in the convex set of density matrices, that is, it cannot be written non-trivially as $\rho = \sum_i \lambda_i \rho_i$. A pure density matrix has rank 1 and can be written as a projector $\rho = |\psi\rangle\langle psi|$ for some vector $|\psi\rangle \in \mathcal{H}$ with $\langle\psi|\psi\rangle = 1$, or equivalently, $\operatorname{tr}\rho = 1$. $|\psi\rangle$ is also often called a pure state or state state vector. A density matrix (state) that is not pure is called *mixed.*

(5) A collection of state vectors $\{|\psi_i\rangle\}_i$ with probabilities $\{p_i\}_i$ is called a *pure state ensemble* for a mixed state ρ if

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|.$$

Every mixed state has infinitely many pure-state ensembles realizing it. Every quantum state ρ has a *spectral decomposition*; one can write $\rho = \sum_i \lambda_i |v_i\rangle\langle v_i|$, where λ_i are the eigenvalues of ρ and $\{|v_i\rangle\}_i$ is an orthonormal basis of eigenvectors of ρ: $\rho|v_i\rangle = \lambda_i|v_i\rangle$.

(6) Because $\rho \geq 0$ and $\operatorname{tr}\rho = 1$, we have $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. Hence, the eigenvalues of a density matrix form a probability distribution, thus generalizing ”classical” states.

5.2. Measurements

Definition 25. Let A be an observable on a quantum system $\mathcal{H}$ in the state ρ. Consider the spectral decomposition

$$A = \sum_\alpha x_\alpha P_\alpha$$

where x_α are the eigenvalues of A and P_α are the orthogonal projectors onto the corresponding eigenspaces. They satisfy:

(1) $P_\alpha \geq 0$, in particular $P_\alpha^\dagger = P_\alpha$.
(2) $P_\alpha P_\beta = \delta_{\alpha\beta} P_\alpha$
(3) $\sum_\alpha P_\alpha = \mathbb{1}$.

$\{P_\alpha\}_\alpha$ is called a *projective measurement*, that gives the value x_α with probability $p_{[\alpha]} = \operatorname{tr}(\rho P_\alpha)$.

For the p_α to be probabilities, we only need (1) and (2) above.

Definition 26. A collection of operators $\{E_k\}_k$ with $E_k \geq 0$ and $\sum_k E_k = \mathbb{1}$ is called a *positive operator-valued measure (POVM)*. The E_k are often called *effect operators*. The outcome “k” is obtained with probability $p_k = \operatorname{tr}(\rho E_k)$.

5.3. Composite systems and entanglement

Consider two quantum systems A and B with associated Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. The joint system AB is described by the tensor product $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. A density matrix ρ_{AB} for the joint system lies in $\mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$, which is isomorphic to $\mathcal{L}(\mathcal{H}_A) \otimes \mathcal{L}(\mathcal{H}_B)$.

The *marginal state* ρ_A of a bipartite state ρ_{AB} is defined via

$$\operatorname{tr}(\rho_{AB}(X_A \otimes \mathbb{1}_B)) = \operatorname{tr}(\rho_A X_A) \tag{13}$$

for all $X_A \in \mathcal{L}(\mathcal{H}_A)$. This uniquely defines a linear map $\operatorname{tr}_B : \mathcal{L}(\mathcal{H}_{AB}) \to \mathcal{L}(\mathcal{H}_A)$ called the *partial trace.* If $\{|e_i\rangle_B\}_{i=1}$ is an orthonormal basis for $\mathcal{H}_B$, then

$$\operatorname{tr}_B X_{AB} = \sum_{i=1}^{\dim B} (\mathbb{1}_A \otimes \langle e_i|_B) X_{AB} (\mathbb{1}_A \otimes |e_i\rangle_B).$$

Equation 13 shows that the marginal ρ_A describes the *effective state* of system A when doing a local measurement.

We distinguish between types of correlations between A and B:

(1) *Product states*: $\rho_{AB} = \omega_A \otimes \sigma_B$ for states ω_A and σ_B. In a product state any local measurements do not depend on the other system, hence A and B are completely uncorrelated.
(2) *Separable states*: $\rho_{AB} = \sum_i p_i \omega_A^{(i)} \otimes \sigma_B^{(i)}$ for states $\left(\omega_A^{(i)}\right)$ and $\left(\sigma_B^{(i)}\right)$ and a probability distribution $(p_i)_i$. Separable states describe classical correlation between A and B corresponding to the index i. Conditioned on this value i, the state $\omega_A^{(i)} \otimes \sigma_B^{(i)}$ is uncorrelated.
(3) *Entangled states* are states that are not separable. They describe quantum correlations.

Example 15. Let $\{|0\rangle, |1\rangle\}$ be a basis for $\mathbb{C}^2$ and consider $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B)$ is called *EPR state, Bell state* or *maximally entangled state.*

$$\Phi^+ = |\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

is <u>not</u> separable.

Note that a pure separable state is automatically a product state.
It is NP-hard to decide where a given mixed state is separable. However for pure states there is an efficient criterion based on the singular value decomposition.

Proposition 5.1. Let $|\psi\rangle_{AB}$ be a pure bipartite quantum state. Then there are sets of orthonormal vectors $\{|e_i\rangle_A\}_{i=1}^r$ and $\{|f_j\rangle_B\}_{j=1}^r$ and strictly positive real numbers $(\lambda_i)_{i=1}^r$ such that

$$|\psi\rangle_{AB} = \sum_{i=1}^r \sqrt{\lambda_i} |e_i\rangle_A \otimes |f_i\rangle_B.$$

The *Schmidt coefficients* $(\lambda_i)_{i=1}^r$ satisfy $\sum_{i=1}^r \lambda_i = 1$, and are unique up to reordering. The integer r is called the *Schmidt rank* of $|\psi\rangle_{AB}$.
$|\psi\rangle_{AB}$ is entangled iff $r > 1$. The marginals of $|\psi\rangle_{AB}$ are given by

$$\rho_A = \operatorname{tr}_B \psi_{AB} = \sum_{i=1}^r \lambda_i |e_i\rangle\langle e_i|_A$$

$$\rho_B = \operatorname{tr}_A \psi_{AB} = \sum_{i=1}^r \lambda_i |f_i\rangle\langle f_i|_B.$$

These are spectral decompositions, that is ρ_A and ρ_B have the same spectrum given by the Schmidt coefficients, and the *Schmidt vectors* $\{|e_i\rangle_A\}$ and $\{|f_j\rangle_B\}$ can be completed to eigenbases of ρ_A and ρ_B respectively.

Proof sketch. Consider orthonormal bases $\{|v_i\rangle_A\}_{i=1}^{|A|}$ and $\{|w_j\rangle_B\}_{j=1}^{|B|}$, and expand $|\psi\rangle_{AB} = \sum_{i,j} x_{i,j} |v_i\rangle_A \otimes |w_j\rangle_B$. All claims now follow from the singular value decomposition of the matrix X with coefficients $x_{i,j}$. ■

Definition 27. Let ρ_A be a mixed quantum state. Any state $|\psi\rangle_{AR} \in \mathcal{H}_A \otimes \mathcal{H}_R$ satisfying $\operatorname{tr}_R \psi_{AR} = \rho_A$ where $\mathcal{H}_R$ is some auxiliary Hilbert space, is called a *purification* of ρ_A.

Proposition 5.2. Let ρ_A be a mixed quantum state.
(1) A purification of ρ exists on $\mathcal{H}_A \otimes \mathcal{H}_R$ where $\dim \mathcal{H}_R \geq \operatorname{rank} \rho_A$.
(2) Let $|\psi\rangle_{AR_1}$ and $|\varphi\rangle_{AR_2}$ be two purifications of ρ_A, and without loss of generality assume $\dim \mathcal{H}_{R_1} \leq \dim \mathcal{H}_{R_2}$. Then there exists an isometry $V : \mathcal{H}_{R_1} \to \mathcal{H}_{R_2}$ such that $|\varphi\rangle_{AR_2} = (\mathbb{1}_A \otimes V)\,|\psi\rangle_{AR_1}$.

Proof. (1) Consider a spectral decomposition $\rho_A = \sum_{i=1}^n \lambda_i |v_i\rangle\langle v_i|_A$, where $\lambda_i > 0$ such that $r = \operatorname{rank} \rho_A$. Take $\mathcal{H}_R = \mathbb{C}^r$ with orthonormal basis $\{|w_i\rangle_R\}_{i=1}^r$, then $|\psi\rangle_{AR} := \sum_{i=1}^r \sqrt{\lambda_i} |v_i\rangle_A \otimes |w_i\rangle_R$ is the desired purification.
(2) This follows from the Schmidt decomposition. ■

5.4. Distance measures

Approximations are quantified using measures of how close quantum states are. There are many such measures of closeness; here we focus on two: fidelity and trace norm.

Definition 28. The *trace norm* of a linear operator $X \in \mathcal{L}(\mathcal{H})$ is

$$\|X\|_1 = \operatorname{tr}\sqrt{X^\dagger X} = \sum_{i=1}^{d} S_i(X),$$

where $d = \dim \mathcal{H}$ and $S_i(X)$ are the singular values of X.

This defines a norm (in the usual sense) on $\mathcal{L}(\mathcal{H})$. In the special case when X is Hermitian with real eigenvalues λ_i, we have $\|X\|_1 = \sum_i^d |\lambda_i|$.

Definition 29. Let ρ and σ be quantum states on $\mathcal{H}$. Then their *trace distance* is defined as $D(\rho,\sigma) := \frac{1}{2}\|\rho - \sigma\|_1$.

Properties of the trace distance.

(1) $D(\cdot,\cdot)$ is a metric, that is, it is non-negative, symmetric and satisfies the triangle inequality.
(2) $0 \leq D(\rho,\sigma) \leq 1$ and $D(\rho,\sigma) = 0$ iff $\rho = \sigma$. With $\operatorname{supp} X := (\ker X)^\perp$, we also have $D(\rho,\sigma) = 1$ iff $\operatorname{supp}\rho \perp \operatorname{supp}\sigma$.
(3) $D(\rho,\sigma) = D(U\rho U^\dagger, U\sigma U^\dagger)$ for all unitaries U and $D(\rho_A,\sigma_A) \leq D(\rho_{AB},\sigma_{AB})$.
(4) $D(\rho,\sigma) = \sup\{\operatorname{tr}[P(\rho-\sigma)] : P \geq 0 \text{ and } \mathbb{1} - P \geq 0\}$.
(5) $D(\rho,\sigma)$ is related to the maximum probability of distinguishing ρ and σ.

Definition 30. The fidelity $F(\rho,\sigma)$ of quantum states ρ and σ is defined as

$$F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \operatorname{tr}\left(\sigma^{\frac{1}{2}} \cdot \rho \cdot \sigma^{\frac{1}{2}}\right)^{\frac{1}{2}}.$$

Properties of the fidelity.

(1) $0 \leq F(\rho,\sigma) \leq 1$ and $F(\rho,\sigma) = 1$ iff $\rho = \sigma$ and $F(\rho,\sigma) = 0$ iff $\operatorname{supp}\rho \perp \operatorname{supp}\sigma$.
(2) $F(\rho,\sigma) = F(\sigma,\rho)$, but F is not a metric.
(3) $F(\rho,\sigma) = F(U\rho U^\dagger, U\sigma U^\dagger)$ for all unitaries U, and $F(\rho_{AB},\sigma_{AB}) \leq F(\rho_A,\sigma_A)$.
(4) $F(\cdot,\cdot)$ is jointly concave: $F(\sum_i p_i\rho_i, \sum_i p_i\sigma_i) \geq \sum_i p_i F(\rho_i,\sigma_i)$.
(5) For pure states $|\psi\rangle$ and $|\varphi\rangle$, $F(\psi,\varphi) = |\langle\psi|\varphi\rangle|$.
(6) Uhlmann's theorem:

$$F(\rho,\sigma) = \max\{|\langle\psi^\rho|\varphi^\sigma\rangle| : |\psi^\rho\rangle \text{ purifies } \rho, |\varphi^\sigma\rangle \text{ purifies } \sigma\}.$$

Proposition 5.3 (Fuchs-van de Graaf inequalities)**.** For any two quantum states ρ and σ,

$$1 - F(\rho,\sigma) \leq D(\rho,\sigma) \leq \sqrt{1 - F(\rho,\sigma)^2}.$$

6. Invariant states

Detecting entanglement in arbitrary bipartite quantum systems is NP-hard. This is easier in the presence of symmetries.

6.1. Werner states

Definition 31. Let $\mathcal{H}_A = \mathcal{H}_B \cong \mathbb{C}^d$ be d-dimensional Hilbert spaces $d \geq 2$. A quantum state ρ_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$ is called a *Werner state* if

$$(U \otimes U)\rho_{AB}(U \otimes U)^\dagger = \rho_{AB}$$

for all $U \in \mathrm{U}_d$.

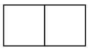
Recall that Schur-Weyl duality gives a decomposition

$$\left(\mathbb{C}^d\right)^{\otimes n} = \bigoplus_{\lambda \vdash_d n} V_\lambda \otimes W_\lambda,$$

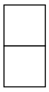
where

- the irrep V_λ of S_n has an orthonormal basis index by the set of standard Young tableaux of shape $\lambda \vdash_d n$, and
$$\dim V_\lambda = d_\lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h(i,j)}.$$
- the irrep W_λ of U_d has an orthonormal basis indexed by the set of semistandard Young tableaux of shape $\lambda \vdash_d n$ and numbering $\{1, ..., d\}$, and
$$\dim W_\lambda = m_\lambda = \prod_{1\le i<j\le d} \frac{\lambda_i - \lambda_j + j - i}{j-i}.$$

If $n = 2$, there are only two partitions of 2: $2 = 2 + 0$ and $2 = 1 + 1$, with corresponding Young diagrams



and



respectively. The dimension of the corresponding irreducible representations $V_{(2,0)}$ and $V_{(1,1)}$ of S_2 is 1, so Schur-Weyl duality in this case becomes

$$\left(\mathbb{C}^d\right)^{\otimes 2} = W_{(2,0)} \oplus W_{(1,1)}.$$

The representation space $W_{(2,0)}$ is the symmetric subspace $\mathrm{Sym}^2(\mathbb{C}^d) = \{|v\rangle \in (\mathbb{C}^d)^{\otimes 2} : \mathbb{F}|v\rangle = |v\rangle\}$ where $\mathbb{F}$ is the swap operator. The dimension of $W_{(2,0)} = \mathrm{Sym}^2(|C^d)$ is $m_{(2,0)} = \dim \mathrm{Sym}^2(\mathbb{C}^d) = \frac{d(d+1)}{2}$.
The representation space $W_{(1,1)}$ is the antisymmetric subspace $\mathrm{Alt}^2(\mathbb{C}^d) = \{|v\rangle \in (\mathbb{C}^d)^{\otimes 2} : \mathbb{F}|v\rangle = -|v\rangle\}$, and it has dimension $m_{(1,1)} = \dim \mathrm{Alt}^2(\mathbb{C}^d) = \frac{d(d-1)}{2}$.
Using Schur's lemma and $(U \otimes U)\rho_{AB}(U \otimes U)^\dagger = \rho_{AB}$ gives

$$\rho_{AB} = c_{(2,0)}\mathbb{1}_{W_{(2,0)}} \oplus c_{(1,1)}\mathbb{1}_{W_{(1,1)}}$$

for some $c_{(2,0)}, c_{(1,1)} \geq 0$ with

$$1 = c_{(2,0)}\frac{d(d+1)}{2} + c_{(1,1)}\frac{d(d-1)}{2}.$$

The Young symmetrizers for $(2, 0)$ and $(1, 1)$ are given by

$$e_{(2,0)} = \mathbb{1} + \mathbb{F}, \qquad e_{(1,1)} = \mathbb{1} - \mathbb{F}$$

and hence we have the projectors

$$P_{(2,0)} = \frac{1}{2}(\mathbb{1} + \mathbb{F}) \qquad \text{onto } V_{(2,0)} \otimes W_{(2,0)} = W_{(2,0)}$$

$$P_{(1,1)} = \frac{1}{2}(\mathbb{1} - \mathbb{F}) \qquad \text{onto } V_{(1,1)} \otimes W_{(1,1)} = W_{(1,1)}$$

with $\operatorname{tr} P_{(2,0)} = \frac{d(d+1)}{2}$ and $\operatorname{tr} P_{(1,1)} = \frac{d(d-1)}{2}$.

Proposition 6.1. A Werner state has the form $\rho_{AB} = x\frac{2}{d(d+1)}P_{(2,0)} + (1-x)\frac{2}{d(d-1)}P_{(1,1)}$ where $x \in [0, 1]$ and $P_{(2,0)} = \frac{1}{2}(\mathbb{1} + \mathbb{F})$, $P_{(1,1)} = \frac{1}{2}(\mathbb{1} - \mathbb{F})$.

Proof. The claim follows from $\operatorname{tr} P_{(2,0)} = \dim W_{(2,0)} = \frac{d(d+1)}{2}$ and $\operatorname{tr} P_{(1,1)} = \dim W_{(1,1)} = \frac{d(d-1)}{2}$. ■

A Werner state ρ_{AB} has an alternative parametrization using the *visibility* α, which is defined by $\alpha = \operatorname{tr}(\rho_{AB}\mathbb{F})$:

$$\rho_{AB} = \frac{1}{d(d^2-1)}\left[(d-\alpha)\mathbb{1} + (d\alpha - 1)\mathbb{F}\right].$$

Definition 32. For $X \in \mathcal{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$ we define the *twirling operation*

$$\mathcal{T}(X) = \int_{\mathrm{U}_d} \mathrm{d}U\,(U\otimes U)X(U\otimes U)^\dagger,$$

where $\mathrm{d}U$ denotes the Haar measure on U_d.

Proposition 6.2. Properties of Werner states:
(1) Every Werner state is invariant under $\mathcal{T}$.
(2) Let ρ_{AB} be an arbitrary state. Then $\mathcal{T}(\rho_{AB})$ is a Werner state of visibility $\alpha = \operatorname{tr}(\mathbb{F}\rho_{AB})$.

Proof. (1) If $(U\otimes U)\rho_{AB}(U\otimes U)^\dagger = \rho_{AB}$ for all $U \in \mathrm{U}_d$, then

$$\begin{aligned}\mathcal{T}(\rho_{AB}) &= \int_{\mathrm{U}_d} \mathrm{d}U\,(U\otimes U)\rho_{AB}(U\otimes U)^\dagger &(14)\\ &= \int_{\mathrm{U}_d} \mathrm{d}U\,\rho_{AB} &(15)\\ &= \rho_{AB}, &(16)\end{aligned}$$

by normalization of the Haar measure.
(2) We compute:

$$\begin{aligned}(U\otimes U)\mathcal{T}(\rho_{AB})(U\otimes U)^\dagger &= (U\otimes U)\left[\int_{\mathrm{U}_d} \mathrm{d}V\,(V\otimes V)\rho_{AB}(V\otimes V)^\dagger\right](U\otimes U)^\dagger &(17)\\ &= \int_{\mathrm{U}_d} \mathrm{d}V\,(UV\otimes UV)\rho_{AB}(UV\otimes UV)^\dagger &(18)\\ &= \mathcal{T}(\rho_{AB}), &(19)\end{aligned}$$

by left invariance of the Haar measure. Hence $\mathcal{T}(\rho_{AB})$ is a Werner state of visibility

$$\begin{aligned}\alpha &= \operatorname{tr}(\mathcal{T}(\rho_{AB})\mathbb{1}) &(20)\\ &= \int_{\mathrm{U}_d} \mathrm{d}U\,\operatorname{tr}\left[(U\otimes U)\rho_{AB}(U\otimes U)^\dagger\mathbb{F}\right] &(21)\\ &= \int_{\mathrm{U}_d} \mathrm{d}U\,\operatorname{tr}\left[\rho_{AB}(U\otimes U)^\dagger\mathbb{F}(U\otimes U)\right] &(22)\\ &= \operatorname{tr}(\rho_{AB}\mathbb{F}) \qquad \text{using } (U\otimes U)^\dagger\mathbb{F}(U\otimes U) = \mathbb{F} \text{ for all } U. \quad \blacksquare &(23)\end{aligned}$$

Lemma 6.3. Let σ_{AB} be a separable state. Then $\mathcal{T}(\sigma_{AB})$ is separable as well, and $\operatorname{tr}(\mathcal{T}(\sigma_{AB}\mathbb{F})) \geq 0$.

Proof. If σ_{AB} is separable, that is, $\sigma_{AB} = \sum_i p_i\sigma_A^{(i)} \otimes \sigma_B^{(i)}$, then clearly $(U\otimes U)\sigma_{AB}(U\otimes U)^\dagger$ is separable for all $U \in \mathrm{U}_d$, and a suitable approximation of the Haar integral using Riemann sums shows that $\mathcal{T}(\sigma_{AB}) = \int_{\mathrm{U}_d} \mathrm{d}U\,(U\otimes U)\sigma_{AB}(U\otimes U)^\dagger$ is a limit of a convex combination of separable states and hence is itself separable. For a product state $\rho_A \otimes \chi_B$, it can be shown that $\operatorname{tr}((\rho_A\otimes\chi_B)\mathbb{F}) = \operatorname{tr}(\rho_A\cdot\chi_B) \geq 0$, since $\rho_A, \chi_B \geq 0$. Hence,

$$\operatorname{tr}(\sigma_{AB}\mathbb{F}_{AB}) = \sum_i p_i \operatorname{tr}\left[(\sigma_A^{(i)}\otimes\sigma_B^{(i)})\mathbb{F}\right] \geq 0. \quad \blacksquare$$

Remark. (1) The identity

$$\operatorname{tr}\left[(X_A\otimes Y_B)\mathbb{F}\right] = \operatorname{tr}(X_A Y_B)$$

is often called the "swap trick."

(2) Let $\alpha \in [0,1]$ be arbitrary, and set $|\varphi\rangle = \sqrt{\alpha}|0\rangle + \sqrt{1-\alpha}|1\rangle$ for some orthonormal $|0\rangle, |1\rangle \in \mathbb{C}^d$. Then

$$\operatorname{tr}\left[(\varphi_A \otimes |0\rangle\langle 0|_B)\mathbb{F}\right] = \operatorname{tr}(|\varphi\rangle\langle\varphi|_A |0\rangle\langle 0|_A) = |\langle\varphi|0\rangle|^2 = \alpha,$$

and hence $\mathcal{T}(\varphi_A \otimes |0\rangle\langle 0|_B)$ is a separable Werner state of visibility α.

We now show that every Werner state ρ_{AB} with $\operatorname{tr}(\rho_{AB}\mathbb{F}) < 0$ is entangled. To this end, we will employ a useful criterion for entanglement based on the *partial transpose*

$$\vartheta_B := \mathrm{id}_A \otimes \vartheta,$$

where $\vartheta : X \mapsto X^T$ denotes the transpose. On product operators, we have $\vartheta_B(X_A \otimes Y_B) = X_A \otimes Y_B^T$.

Proposition 6.4 (PPT criterion.)**.** For every separable state σ_{AB}, $\vartheta_B(\sigma_{AB}) \geq 0$. Hence, if $\vartheta_B(\rho_{AB})$ has a negative eigenvalue, then ρ_{AB} is entangled.

Proof. Let $\sigma_{AB} = \sum_i p_i \sigma_A^{(i)} \otimes \sigma_B^{(i)}$ be separable. Since ϑ_B is linear and X is positive semidefinite iff X^T is positive semidefinite, we have

$$\vartheta_B(\sigma_{AB}) = \sum_i p_i \sigma_A^{(i)} \otimes \left(\sigma_B^{(i)}\right)^T \geq 0,$$

as a convex combination of positive semidefinite operators. ■

Lemma 6.5. A Werner state ρ_{AB} is entangled if $\operatorname{tr}(\rho_{AB}\mathbb{F}) < 0$.

Proof. We can parametrize ρ_{AB} with $\alpha = \operatorname{tr}(\rho_{AB}\mathbb{F}) < 0$ as

$$\rho_{AB} = \frac{1}{d(d^2-1)}\left[(d-\alpha)\mathbb{1} + (d\alpha - 1)\mathbb{F}\right].$$

We also have $\vartheta_B(\mathbb{1}_{AB}) = \mathbb{1}_{AB}$, and fixing an orthonormal basis $\{|i\rangle\}_{i=1}^d$ of $\mathbb{C}^d$ we can write

$$\mathbb{F}_{AB} = \sum_{i,j=1}^d |i\rangle\langle j|_A \otimes |j\rangle\langle i|_B,$$

and hence

$$\vartheta_B(\mathbb{F}_{AB}) = \sum_{i,j=1}^d |i\rangle\langle j|_A \otimes \left(|j\rangle\langle i|_B\right)^T \tag{24}$$

$$= \sum_{i,j=1}^d |i\rangle\langle j|_A \otimes |i\rangle\langle j|_B \tag{25}$$

$$= d|\Phi^+\rangle\langle\Phi^+|_{AB} \tag{26}$$

with $|\Phi^+\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |i\rangle_A|i\rangle_B$ is a maximally entangled state. Then $\vartheta_B(\rho_{AB}) \propto (d-\alpha)\mathbb{1} + d(d\alpha - 1)|\Phi^+\rangle\langle\Phi^+|$. Write X_{AB} for the operator on the right side of this equation. Let P_{AB} be the projection onto $\operatorname{span}\{|\Phi^+\rangle\}^\perp$, then

$$\mathbb{1}_{AB} = |\Phi^+\rangle\langle\Phi^+|_{AB} + P,$$

and hence X_{AB} has eigenvalues $\lambda_1 = d - \alpha + d^2\alpha - d = \alpha(d^2-1)$, $\lambda_2 = d - \alpha$. For $\alpha < 0$, we have $\lambda_1 = \alpha(d^2-1) < 0$, since $d \geq 2$. ■

The following proposition summarizes what we have proved so far.

Proposition 6.6. A Werner state ρ_{AB} is entangled iff $\operatorname{tr}(\rho_{AB}\mathbb{F}) < 0$.

The PPT criterion is generally only a *necessary* criterion for separability. There are entangled states ρ_{AB} with $\vartheta_V(\rho_{AB}) \geq 0$; these are called "bound entangled" states.
However, the PPT criterion is also sufficient in some special cases:

(1) A Werner state is separable iff it is PPT.
(2) Let A, B be two systems with $\dim A \cdot \dim B \leq 6$. Then a state is separable iff it is PPT.
(3) This result can be generalized to some low-rank states in higher dimensions.

We can generalize Werner states to the multipartite setting: Let $\mathcal{H}_{A_i} = \mathbb{C}^d$ for $i = 1, ..., n$. A state $\rho_{A_1 \cdots A_n}$ is called a *multipartite Werner state* if

$$U^{\otimes n}\rho_{A_1\cdots A_n}(U^\dagger)^{\otimes n} = \rho_{A_1\cdots A_n}$$

for all $U \in \mathrm{U}_d$.
Schur-Weyl duality for multipartite Werner states. Let $\mathcal{A} = \operatorname{span}\{U^{\otimes n} : U \in \mathrm{U}_d\}$ and $B = \operatorname{span}\{Q_\pi : \pi \in S_n\}$, where $Q_\pi := \varphi(\pi)$ where $\varphi(\pi)$ denotes the action of S_n. Then

$$U^{\otimes n}\rho_{A_1\cdots A_n}(U^\dagger)^{\otimes n} = \rho_{A_1\cdots A_n}$$

for all $U \in \mathrm{U}_d$ implies

$$\rho_{A_1\cdots A_n} \in \mathcal{A}' = \mathcal{B},$$

and thus

$$\rho_{A_1\cdots A_n} = \sum_{\pi \in S_n} c_\pi Q_\pi$$

for some $c_\pi \in \mathbb{C}$.
In the $n = 2$ case, we had $\rho_{A_1A_2} = \alpha\mathbb{1} + \beta\mathbb{F}$. This expression for $\rho_{A_1\cdots A_n}$ may not always be useful since the Q_π are not always positive semi-definite.
Alternatively, one can consider the decomposition

$$(\mathbb{C}^d)^{\otimes n} \cong \bigoplus_{\lambda \vdash_d n} V_\lambda \otimes W_\lambda.$$

Using Schur's lemma, $U^{\otimes n}$-invariance forces $\rho_{A_1\dots A_n}$ to be a scalar multiple of the identity $\mathbb{1}_{W_\lambda}$ on W_λ. Thus

$$\rho_{A_1\cdots A_n} = \bigoplus_{\lambda \vdash_d n} x_\lambda \rho_\lambda \otimes \frac{1}{m_\lambda \mathbb{1}_{W_\lambda}}$$

where $(x_\lambda)_{\lambda \vdash_d n}$ is a probability distribution, ρ_λ is a quantum state on V_λ for $\lambda \vdash_d n$, and $m_\lambda = \dim W_\lambda$.
If in addition $\rho_{A_1\dots A_n}$ is *symmetric*, that is, $Q_\pi \rho_{A_1\dots A_n} Q_\pi^\dagger = \rho_{A_1\dots A_n}$ for all $\pi \in S_n$, then

$$\rho_{A_1\cdots A_n} = \bigoplus_{\lambda \vdash_d n} x_\lambda \frac{1}{d_\lambda}\mathbb{1}_{V_\lambda} \otimes \frac{1}{m_\lambda}\mathbb{1}_{W_\lambda} = \sum_{\lambda \vdash_d n} x_\lambda \rho_\lambda,$$

where $\rho_\lambda = \frac{1}{d_\lambda m_\lambda} P_\lambda$ and P_λ is the projector onto $V_\lambda \otimes W_\lambda$.

6.2. Isotropic states

A state ρ_{AB} on systems AB with $\mathcal{H}_A \cong \mathcal{H}_B \cong \mathbb{C}^d$ is called *isotropic* if $(U \otimes \overline{U})\rho_{AB}(U \otimes \overline{U})^\dagger = \rho_{AB}$ for all $U \in \mathrm{U}_d$.
Clearly $(U \otimes \overline{U})\mathbb{1}_{AB}(U \otimes \overline{U})^\dagger = \mathbb{1}_{AB}$ for all $U \in \mathrm{U}_d$ as well. Using $\vartheta_B(\Phi^+_{AB}) = \frac{1}{d}\mathbb{F}_{AB}$, Schur-Weyl duality shows:

Proposition 6.7. An isotropic state ρ_{AB} can be written as $\rho_{AB} = (1-x)|\Phi^+\rangle\langle\Phi^+|_{AB} + x\frac{1}{d^2}\mathbb{1}_{AB}$ for $x \in \left[0, \frac{d^2}{d^2-1}\right]$.

Since $\operatorname{tr}_B \varphi_{AB} = \frac{1}{d}\mathbb{1}_A$, we have

$$\begin{aligned} \rho_{AB} &= (1-x)\Phi^+_{AB} + x\frac{1}{d^2}\mathbb{1}_A \otimes \mathbb{1}_B && (27) \\ &= (\mathrm{id}_A \otimes \mathcal{D})(\Phi^+_{AB}), && (28) \end{aligned}$$

where we defined the depolarizing channel

$$\mathcal{D}_X(\omega) := (1-x)\omega + x\,\mathrm{tr}(\omega)\frac{1}{d}\mathbb{1},$$

which is an important noise model satisfying $\mathcal{D}_X(U\omega U^\dagger) = U\mathcal{D}_X(\omega)U^\dagger$ for all $\omega \in \mathcal{L}(\mathbb{C}^d)$, $U \in \mathrm{U}_d$.

Proposition 6.8. Let $\rho_{AB} := (1-x)\Phi^+_{AB} + \frac{x}{d^2}\mathbb{1}_{AB}$ with $x \in [0, \frac{d^2}{d^2-1}]$.

(1) ρ_{AB} is separable iff $x \geq \frac{d}{d+1}$.

(2) Let σ_{AB} be arbitrary with $\beta := \mathrm{tr}\big(\sigma_{AB}\Phi^+_{AB}\big) = \langle\Phi^+|\sigma_{AB}|\Phi^+\rangle$. Then

$$\int_{\mathrm{U}_d} (U \otimes \overline{U})\sigma_{AB}(U \otimes \overline{U})^\dagger = \rho_{AB}(y)$$

where $y = \frac{d^2}{d^2-1}(1-\beta)$.

Proof. Exercise. ■

7. The de Finetti theorem

7.1. Extendability of quantum states

Consider a bipartite state ρ_{AB}. We call ρ_{AB} k-extendible if there exists a k-extension, a state $\rho_{AB_1\cdots B_k}$ with $B_i \cong B$ and $\rho_{AB_i} = \mathrm{tr}_{B_1\cdots B_{i-1}B_{i+1}\cdots B_k}\,\rho_{AB_1\cdots B_k} = \rho_{AB}$ for all $i = 1, ..., k$.

Lemma 7.1. Separable states are ∞-extendible.

Proof. Let $\sigma_{AB} = \sum_i p_i\sigma_A^{(i)} \otimes \sigma_B^{(i)}$ be separable, then $\sigma_{AB_1\cdots B_k} = \sum_i p_i\sigma_A^{(i)} \otimes \sigma_{B_1}^{(i)} \otimes \cdots \otimes \sigma_{B_k}^{(i)}$ defines a k-extension for arbitrary $k \in \mathbb{N}$. ■

Conversely one can show that for every entangled state ρ_{AB} exists a k_0 such that ρ_{AB} has no k-extension for $k \geq k_0$.

Example 16. Pure entangled states are not even 2-extendible.

This is usually called the *monogamy of entanglement*:

A quantum system cannot be entangled with a large number of other systems.

The De Finetti theorems provide a quantitative version of monogamy.

7.2. A de Finetti theorem for pure symmetric states

We will focus on pure states in the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^d) = \{|\psi\rangle \in (\mathbb{C}^d)^{\otimes n} : \varphi(\pi)|\psi\rangle = |\psi\rangle\}$. Note that $\dim \mathrm{Sym}^n(\mathbb{C}^d) = \binom{n+d-1}{n}$ by Weyl's dimension formula.

Lemma 7.2. Let $|\psi\rangle \in \mathbb{C}^d$ be arbitrary. Then

$$\pi_n = \binom{n+d-1}{n}\int_{\mathrm{U}_d} \mathrm{d}U\, (U|\varphi\rangle\langle\varphi|U^\dagger)^\dagger$$

is equal to the projector onto the symmetric subspace.

Proof. This follows from showing:

(1) $\pi_n \in \mathrm{End}(\mathrm{Sym}^n(\mathbb{C}^d))$

(2) U_d acts irreducibly on $\mathrm{Sym}^n(\mathbb{C}^d)$ via $U \to U^{\otimes n}$.

(3) π_n is invariant under this action and Schur's lemma.

Details are left as an exercise. ■

Proposition 7.3 (De Finetti theorem for pure symmetric states). Let $\mathcal{H}_{A_i} \cong \mathbb{C}^d$ and $|\psi\rangle_{A_1\cdots A_n} \in \mathrm{Sym}^n(\mathbb{C}^d)$. Then for any $k < n$,

$$D\left(\psi_{A_1\cdots A_k}, \int \mathrm{d}\psi\, p(\varphi)|\varphi\rangle\langle\psi|^{\otimes k}\right) \leq \sqrt{\frac{dk}{n-k}}.$$

Here $p(\varphi)$ is a probability density that depends on $|\psi\rangle$ and $\mathrm{d}\varphi$ is the measure on pure states induced by the Haar measure, that is, $|\varphi\rangle = U|\varphi_0\rangle$ for some fixed $|\varphi_0\rangle$, and $\mathrm{d}\varphi\, f(|\varphi\rangle) = \mathrm{d}U\, f(U|\varphi\rangle)$.

Proof. The main idea is the following: For $m = n-k$ interpret $\pi_m = \binom{m+d+1}{m}\int \mathrm{d}\varphi\, |\varphi\rangle\langle\varphi|^{\otimes m}$ as a continuous POVM with operators $\binom{m+d-1}{m}|\varphi\rangle\langle\varphi|^{\otimes m}$ measuring the last m systems that we trace out. Getting a specific outcome $|\psi\rangle \in \mathbb{C}^d$ should then imply that the first k systems are also in the state $|\psi\rangle^{\otimes l}$ on average, due to the permutation invariance of $|\psi\rangle_{A1\cdots A_n}$. Since $|\psi\rangle_{A1\cdots A_n}$ is symmetric under arbitrary permutations, in particular we have $|A_1\cdots A_n\rangle = (\mathbb{1}_k \otimes \pi_m)|\psi\rangle_{A_1\cdots A_n}$. Hence,

$$\begin{aligned}
\psi_{A_1\ldots A_k} &= \mathrm{tr}_{A_{k+1}\ldots A_n} |\psi\rangle\langle\psi|_{A_1\cdots A_n} && (29)\\
&= \mathrm{tr}_{A_{k+1}\ldots A_n}\left[(\mathbb{1}_k \otimes \pi_m)|\psi\rangle\langle\psi|_{A_1\cdots A_n}\right] && (30)\\
&= \binom{m+d-1}{m}\int \mathrm{d}U\,(\mathbb{1}_k \otimes \langle\varphi|^{\otimes n})|\psi\rangle\langle\psi|(\mathbb{1}_k \otimes |\varphi\rangle^{\otimes n}) && (31)
\end{aligned}$$

where the last equality uses the the partial cyclicity property

$$\mathrm{tr}_2((\mathbb{1}\otimes X_2)Y_{12}) = \mathrm{tr}_2(Y_{12}(\mathbb{1}\otimes X_2)).$$

We define

$$\sqrt{p(\varphi)}|e_\varphi\rangle = \binom{m+d-1}{m}^{\frac{1}{2}}(\mathbb{1}_k \otimes \langle\varphi|^{\otimes m})|\psi\rangle_{A_1\cdots A_n} \in (\mathbb{C}^d)^{\otimes k},$$

where $p(\varphi) \geq 0$ ensures that $\langle e_\varphi|e_\varphi\rangle = 1$. Note that $p(\varphi)$ is a probability density, that is, $\int \mathrm{d}\varphi\, p(\varphi) = 1$. Hence, we have $\psi_{A_1\cdots A_k} = \int \mathrm{d}\varphi\, p(\varphi)|e_\varphi\rangle\langle e_\varphi|$. We want to show

$$\int \mathrm{d}\varphi\, p(\varphi)|e_\varphi\rangle\langle e_\varphi| \approx \int \mathrm{d}\varphi\, p(\varphi)|\varphi\rangle\langle\varphi|^{\otimes k}.$$

First, we compute the average (squared) fidelity of $|e_\varphi\rangle$ and $|\varphi\rangle^{\otimes k}$:

$$\begin{aligned}
\int \mathrm{d}\varphi\, p(\varphi)F(|e_\varphi\rangle, |\varphi\rangle^{\otimes k})^2 &= \int \mathrm{d}\varphi\, p(\varphi)\, \langle e_\varphi|\varphi^{\otimes k}|e_\varphi\rangle && (32)\\
&= \binom{m+d-1}{m}\int \mathrm{d}\varphi\, \langle\psi|\varphi^{\otimes k+m}|\psi\rangle && (33)\\
&= \binom{m+d-1}{m}\cdot\binom{n+d-1}{n}^{-1}\langle\psi|\pi_{k+m}|\psi\rangle && (34)\\
&= \binom{m+d-1}{m}\cdot\binom{k+m+d-1}{k+m}^{-1} && (35)\\
&= \frac{(m+d-1)\cdots(m+1)}{(k+m+d-1)\cdots(k+m+1)} && (36)\\
&\geq \left(\frac{m+1}{k+m+1}\right)^{d-1} = \left(1-\frac{k}{k+m+1}\right)^{d-1} && (37)\\
&\geq 1-\frac{k(d-1)}{k+m+1} && (38)\\
&\geq 1-\frac{kd}{m}. && (39)
\end{aligned}$$

We are now prepared to finish the proof. Recall the Fuchs-van-de-Graaf inequality $D(\rho,\sigma) \leq \sqrt{1-F(\rho,\sigma)^2}$ (which is in fact an equality for pure states).

$$
\begin{aligned}
D\left(\psi_{A_1\cdots A_k}, \int \mathrm{d}\varphi\, p(\varphi)|\varphi\rangle\langle\varphi|^{\otimes k}\right) &= D\left(\int \mathrm{d}\varphi\, p(\varphi)|e_\varphi\langle e_\varphi|\rangle, \int \mathrm{d}\varphi\, p(\varphi)|\varphi\rangle\langle\varphi|^{\otimes k}\right) && (40)\\
&\leq \int \mathrm{d}\varphi\, p(\varphi) D(|e_\varphi\rangle\langle e_\varphi|, |\varphi\rangle\langle\varphi|^{\otimes k}) && \text{by convexity of norms} \quad (41)\\
&\leq \int \mathrm{d}\varphi\, p(\varphi)\sqrt{1 - F(e_\varphi, \varphi^{\otimes k})^2} && (42)\\
&\leq \left(\int \mathrm{d}\varphi\, p(\varphi)(1 - F(e_\varphi, \varphi^{\otimes k})^2)\right)^{\frac{1}{2}} && \text{by Jensen's inequality} \quad (43)\\
&= \left(1 - \int \mathrm{d}\varphi\, p(\varphi) F(e_\varphi, \varphi^{\otimes k})^2\right)^{\frac{1}{2}} && (44)\\
&\leq \sqrt{1 - (1 - \frac{kd}{m})} = \sqrt{\frac{kd}{m}}. \quad \blacksquare && (45)
\end{aligned}
$$

7.3. Extension to permutation-invariant mixed states

A state $\rho_{A_1\cdots A_n}$ is called *permutation-invariant* if

$$Q_\pi \rho_{A_1\cdots A_n} Q_\pi^\dagger = \rho_{A_1\cdots A_n}$$

for all $\pi \in S_n$, where $Q_\pi = \varphi(\pi)$.
Our goal is to prove the de Finetti theorem for permutation invariant ρ. Our strategy will be to use our theorem from Section 7.2 for pure states. But first, we need to relate permutation invariant states to pure states in $\mathrm{Sym}^n(\mathbb{C}^d)$. This relation is provided by the next lemma.

Lemma 7.4. Let $\mathcal{H}_{A_i} = \mathbb{C}^d$ for $i = 1, ..., n$ and $\rho_{A_1\cdots A_n}$ be permutation invariant. Then $\rho_{A_1\cdots A_n}$ has a purification $|\psi^\rho\rangle \in \mathrm{Sym}^n(\mathbb{C}^d \otimes \mathbb{C}^d)$.

Proof. Let $\rho_{A_1\cdots A_n} = \sum_{\lambda\in\mathrm{Spec}(\rho)} \lambda P_\lambda$ be a spectral decomposition where $\mathrm{Spec}(\rho)$ is the set of distinct eigenvalues of ρ with corresponding orthogonal projector P_λ into the eigenspace $\mathcal{H}_\lambda$. Since $\rho = Q_\pi \rho Q_\pi^\dagger$ for all $\pi \in S_n$, we have for any $\lambda \in \mathrm{Spec}(\rho)$ and $|\varphi\rangle \in \mathcal{H}_\lambda$ that

$$\lambda|\varphi\rangle = \rho|\varphi\rangle = Q_\pi \rho Q_\pi^\dagger |\varphi\rangle$$

and hence $Q_\pi^\dagger|\varphi\rangle \in \mathcal{H}_\lambda$ for all $\pi \in S_n$, that is, the eigenspaces are permutation-invariant too, and $P_\lambda Q_\pi = Q_\pi P_\lambda$ for all $\pi \in S_n$, $\lambda \in \mathrm{Spec}(\rho)$. Define $M = \sum_{\lambda\in\mathrm{Spec}(\rho)} \sqrt{\lambda} P_\lambda$, then we we have $Q_\pi M = M Q_\pi$ for all $\pi \in S_n$. Now let

$$|\varphi\rangle_{A1\cdots A_n R_1\cdots R_n} := \sum_{x=1}^{d^n} |x\rangle_{A^n} \otimes |x\rangle_{R^n}$$

where $\{|x\rangle\}_{x=1}^{d^n}$ is a product basis for $(\mathbb{C}^d)^{\otimes n}$, and set $|\psi^\rho\rangle = (M \otimes \mathbb{1}_{R^n})|\varphi\rangle$. Then

$$
\begin{aligned}
\mathrm{tr}_{R^n} \psi^\rho_{A^n R^n} &= M(\mathrm{tr}_{R^n} \varphi_{A^n R^n}) M^\dagger && (46)\\
&= MM^\dagger && (47)\\
&= \sum_{\lambda,\lambda'} \sqrt{\lambda\lambda'} P_\lambda P_{\lambda'} = \rho_{A^n}. && (\text{ since } P_\lambda P_{\lambda'} = \delta_{\lambda\lambda'} P_\lambda) \quad (48)
\end{aligned}
$$

Note that if S_n acts on both $A_1, ..., A_n$ and $R_1, ..., R_n$ via Q_π, then S_n acts on $A_1 \cdots A_n R_1 \cdots R_n \cong A_1 R_1 \cdots A_n R_n$ via $Q_\pi \otimes Q_\pi$.
The following calculation proves that $|\psi^\rho\rangle \in \mathrm{Sym}^n(\mathbb{C}^d \otimes \mathbb{C}^d)$, which completes the proof.

$$
\begin{aligned}
(Q_\pi \otimes Q_\pi)|\psi^\rho\rangle &= (Q_\pi \otimes Q_\pi)(M \otimes \mathbb{1})|\varphi\rangle && && (49)\\
&= (Q_\pi M \otimes \mathbb{1})(\mathbb{1} \otimes Q_\pi)|\varphi\rangle && && (50)\\
&= (Q_\pi M Q_\pi^T \otimes \mathbb{1})|\varphi\rangle && \text{by the transpose trick} && (51)\\
&= (M Q_\pi Q_\pi^T \otimes \mathbb{1})|\varphi\rangle && \text{since } [M, Q_\pi] = 0 && (52)\\
&= (M \otimes \mathbb{1})|\varphi\rangle && && (53)\\
&= |\psi^\rho\rangle && \text{for all } \pi \in S_n. \quad \blacksquare && (54)
\end{aligned}
$$

Proposition 7.5. *Let $\mathcal{H}_{A_i} = \mathbb{C}^d$ for $i = 1, ..., n$ and $\rho_{A_1\cdots A_n}$ be a permutation-invariant state. Then for any $k < n$,*

$$D(\rho_{A_1\cdots A_k}, \int \mathrm{d}\mu(\sigma)\, \sigma^{\otimes k}) \leq \sqrt{\frac{d^2 k}{n-k}},$$

where $\mathrm{d}\mu(\sigma)$ is a measure on the space of mixed states on $\mathbb{C}^d$ that depends on ρ.

Proof. Let $|\psi^\rho\rangle_{A^n R^n} \in \mathrm{Sym}^n(\mathbb{C}^d \otimes \mathbb{C}^d)$ be a symmetric purification of ρ. Applying the symmetric subspace de Finetti theorem shows

$$D\left(\psi^\rho_{A_1 R_1 \cdots A_k R_k}, \int \mathrm{d}\varphi\, p(\varphi) |\varphi\rangle\langle\varphi|^{\otimes k}_{AR}\right) \leq \sqrt{\frac{d^2 k}{n-k}}$$

for a suitable probability density $p(\varphi)$. The claim now follows from the monotonicity of $D(\cdot, \cdot)$ under partial trace:

$$
\begin{aligned}
D\left(\rho_{A_1\cdots A_k}, \int \mathrm{d}\varphi\, p(\varphi)\, \mathrm{tr}_R\, \varphi^{\otimes k}_{AR}\right) &\leq D\left(\psi^\rho, \int \mathrm{d}\varphi\, p(\varphi) |\varphi\rangle\langle\varphi|^{\otimes k}_{AR}\right) && (55)\\
&\leq \sqrt{\frac{d^2 k}{n-k}}. \quad \blacksquare && (56)
\end{aligned}
$$

8. Approximate cloning

Classical and quantum information are fundamentally different. Classical information can be cloned and thus replicated arbitrarily. This is impossible for quantum information, as the main theorem of the next section shows.

8.1. The no-cloning theorem

Theorem 8.1 (No-cloning theorem)**.** Let A, B be d-dimensional quantum systems. There is no unitary $U \in \mathrm{U}_d$ that achieves the transformation

$$U : |\psi\rangle_A \otimes |0\rangle_B \mapsto |\psi\rangle_A \otimes |\psi\rangle_B$$

for arbitrary $|\psi\rangle \in \mathcal{H}_A$. Here $|0\rangle_B$ is some reference state.

Proof. Let $|\psi\rangle, |\varphi\rangle \in \mathcal{H}_A$ be such that

$$
\begin{aligned}
U(|\psi\rangle_A \otimes |0\rangle_B) &= |\psi\rangle_A \otimes |\psi\rangle_B,\\
U(|\varphi\rangle_A \otimes |0\rangle_B) &= |\varphi\rangle_A \otimes |\varphi\rangle_B.
\end{aligned}
$$

Then

$$
\begin{aligned}
\langle\psi|\varphi\rangle^2 &= (\langle\psi| \otimes \langle\psi|)(|\varphi\rangle \otimes |\varphi\rangle) && (57)\\
&= (\langle\psi| \otimes \langle 0|) U^\dagger U (|\varphi\rangle \otimes |0\rangle) && (58)\\
&= \langle\psi|\varphi\rangle\,. && (59)
\end{aligned}
$$

This proves $\langle\psi|\varphi\rangle$ must be either 0 and 1,which proves there is no U that achieves $U(|\psi\rangle\otimes|0\rangle) = |\psi\rangle^{\otimes 2}$ for all $|\psi\rangle$. ■

8.2. Approximate cloning machines

Exact cloning is forbidden by the no-cloning theorem, but what about approximate cloning? We consider the scenario where we are given a Hilbert space of dimension d and N copies of a pure state $|\psi\rangle \in \mathcal{H}$. The goal is to produce an approximation of M copies of $|\psi\rangle\langle\psi|$ for some $M > N$. The *figure of merit* for this scenario is defined as follows. Let T be the approximate cloning map

$$T : \mathcal{L}(\mathcal{H}^{\otimes N}) \to \mathcal{L}(\mathcal{H}^{\otimes M}).$$

We require T to be a completely positive and trace-preserving linear map. We define the *worst case fidelity* of T by

$$F(T) = \inf_{|\psi\rangle} F(\psi^{\otimes M}, T(\psi^{\otimes N}))^2 \tag{60}$$

$$= \inf_{|\psi\rangle} \operatorname{tr}\Big(\psi^{\otimes M} T(\psi^{\otimes n})\Big). \tag{61}$$

The next lemma gives an upper bound for the worst case fidelity.

Lemma 8.2. Set $d_N := \dim \operatorname{Sym}^N(\mathcal{H}) = \binom{d+N-1}{N}$. For any approximate cloning map $T : \mathcal{L}(\mathcal{H}^{\otimes N}) \to \mathcal{L}(\mathcal{H}^{\otimes M})$,

$$F(T) \leq \frac{d_N}{d_M} = \binom{d+N-1}{N}\binom{d+M-1}{M}^{-1}.$$

Proof. For a given $T : \mathcal{L}(\mathcal{H}^{\otimes N}) \to \mathcal{L}(\mathcal{H}^{\otimes M})$ define a twirled version

$$\overline{T}(X) := \int_{\mathrm{U}_d} (U^\dagger)^{\otimes M} T(U^{\otimes N} X (U^\dagger)^{\otimes N}) U^{\otimes M} \, \mathrm{d}U$$

which satisfies $\overline{T}(U^{\otimes N} X (U^\dagger)^{\otimes N}) = U^{\otimes M} T(X) (U^\dagger)^{\otimes M}$ for all $U \in \mathrm{U}_d$. Let $|\varphi\rangle \in \mathcal{H}$ be arbitrary, then

$$\operatorname{tr}\Big(\varphi^{\otimes M}\overline{T}(\varphi^{\otimes N})\Big) = \int \mathrm{d}U \operatorname{tr}\Big[\varphi^{\otimes M}(U^\dagger)^{\otimes M} T(U^{\otimes N}\varphi^{\otimes N}(U^\dagger)^{\otimes N})U^{\otimes M}\Big] \tag{62}$$

$$= \int \mathrm{d}U \operatorname{tr}\Big[(U\varphi U^\dagger)^{\otimes M} T((U\varphi U^\dagger)^{\otimes N})\Big] \tag{63}$$

$$\geq \int \mathrm{d}U \, F(T), \tag{64}$$

where the last inequality uses

$$\operatorname{tr}\Big[(U\varphi U^\dagger)^{\otimes M} T((U\varphi U^\dagger)^{\otimes N})\Big] \geq \inf_{|\psi\rangle} \operatorname{tr}\Big(\psi^{\otimes M} T(\psi^{\otimes N})\Big) = F(T).$$

Taking the infimum over $|\varphi\rangle \in \mathcal{H}$, we get $F(\overline{T}) \geq F(T)$. Now let $\tau_N := \frac{1}{d_N}\pi_N$ where π_N is the projector onto $\operatorname{Sym}^N(\mathcal{H})$. Using that $U^{\otimes N}\pi_N(U^\dagger)^{\otimes N} = \pi_N$ for all $U \in \mathrm{U}_d$, we get

$$U^{\otimes M}\overline{T}(\tau_N)(U^\dagger)^{\otimes M} = \overline{T}(U^{\otimes N}\tau_N(U^\dagger)^{\otimes N}) = \overline{T}(\tau_N)$$

for all $U \in \mathrm{U}_d$. Schur-Weyl duality then implies $\overline{T}(\tau_N) = \lambda\tau_M + (1-\lambda)\sigma$ where $\sigma \perp \operatorname{Sym}^M(\mathcal{H})$ and $\lambda \in [0,1]$.

We also have for every $|\varphi\rangle \in \mathcal{H}$ that $\pi_N - |\varphi\rangle\langle\psi|^{\otimes N} \geq 0$. Therefore

$$0 \leq \overline{T}(\pi_N - |\varphi\rangle\langle\varphi|^{\otimes N}) \tag{65}$$

$$= \overline{T}(\pi_N) - \overline{T}(\varphi^{\otimes N}) \tag{66}$$

$$= d_N\lambda\tau_M + d_N(1-\lambda)\sigma - \overline{T}(\varphi^{\otimes N}). \tag{67}$$

We can further compute

$$0 \leq \operatorname{tr}\left[\varphi^{\otimes M}\overline{T}(\pi_N - |\varphi\rangle\langle\varphi|^{\otimes N})\right] = d_N\lambda \underbrace{\operatorname{tr}\left(\varphi^{\otimes M}\tau_M\right)}_{(*)} + d_N(1-\lambda)\underbrace{\operatorname{tr}\left(\varphi^{\otimes M}\sigma\right)}_{(**)} - \operatorname{tr}\left[\varphi^{\otimes M}\overline{T}(\varphi^{\otimes N})\right].$$

The quantities $(*)$ and $(**)$ can be simplified:

$$(*) = \operatorname{tr}\left(\varphi^{\otimes M}\pi_M d_M^{-1}\right) = \frac{1}{d_M}\operatorname{tr}\left(\varphi^{\otimes M}\right) = \frac{1}{d_M},$$

$$(**) = \operatorname{tr}\left(\pi_M\varphi^{\otimes M}\pi_M\sigma\right) = \operatorname{tr}\left(\varphi^{\otimes M}\pi_M\sigma\pi_M\right) = 0.$$

This gives $\operatorname{tr}\left[\varphi^{\otimes M}\overline{T}(\varphi^{\otimes N})\right]$. Therefore we have the following chain of inequalities, which completes the proof.

$$F(T) \leq F(\overline{T}) \leq \operatorname{tr}\left[\varphi^{\otimes M}\overline{T}(\varphi^{\otimes N})\right] \leq \frac{d_N}{d_M}\lambda \leq \frac{d_N}{d_M}. \qquad \blacksquare$$

Can the bound in the lemma be achieved? The answer is yes; the following map achieves it. Define

$$T(X) = \frac{d_N}{d_M}\pi_M(X \otimes \mathbb{1}_d^{\otimes M-N})\pi_M.$$

The action of this map on $X \in \mathcal{L}(\mathcal{H}^{\otimes N})$ can be understood as consisting of the following three steps.
Step 1. Extend state trivially from $\mathcal{H}^{\otimes N}$ to $\mathcal{H}^{\otimes M}$.
Step 2. Project down to symmetric subspace $\mathrm{Sym}^M(\mathcal{H})$.
Step 3. Normalize to get a quantum state.
To compute the fidelity $F(T)$ of this map, we have for arbitrary $|\varphi\rangle \in \mathcal{H}$,

$$\begin{aligned}
\operatorname{tr}\left[\varphi^{\otimes M}T(\varphi^{\otimes N})\right] &= \frac{d_N}{d_M}\operatorname{tr}\left[\varphi^{\otimes M}\pi_M(\varphi^{\otimes N}\otimes\mathbb{1})\pi_M\right] &(68)\\
&= \frac{d_N}{d_M}\operatorname{tr}\left[\pi_M\varphi^{\otimes M}\pi_M(\varphi^{\otimes N}\otimes\mathbb{1})\right] &(69)\\
&= \frac{d_N}{d_M}\operatorname{tr}\left[\varphi^{\otimes M}(\varphi^{\otimes N}\otimes\mathbb{1})\right] &(70)\\
&= \frac{d_N}{d_M}. &(71)
\end{aligned}$$

Therefore $F(T) = \frac{d_N}{d_M} \geq 1 - \frac{kd}{N}$ for $M = N + K$.
These results are due to [Wer98].

8.3. Further results on approximate cloning

(1) The approximate cloning map

$$T(\rho) = \frac{d_N}{d_M}\pi_M(\rho\otimes\mathbb{1}^{\otimes M-N})\pi_M \qquad (*)$$

is the unique cloning map achieving $F(T) = \frac{d_N}{d_M}$.

(2) The fidelity $F(T) = \inf_{|\psi\rangle} F(\psi^{\otimes M}, T(\psi^{\otimes N}))^2$ measures the quality of the full output state, which includes correlations between different systems. We might only be interested in comparing *single copies*; can we find a better map in this case? Interestingly, the answer is no. The cloning map in $(*)$ is also optimal for the single-copy worst-case fidelity

$$F_S(T) = \inf_{|\psi\rangle} F(\psi, \operatorname{tr}_{[M]\setminus\{1\}})T(\psi^{\otimes N}).$$

See [KW01].

(3) There are *asymmetric cloning* machines for which the single-copy fidelities on different sites are not necessarily equal. It is hard to obtain optimality results in general.

(4) There are also *state-dependent approximate cloning* protocols that exploit some known structure in the state to be cloned.

(5) An important application of approximate cloning is in quantum cryptography, specifically quantum key distribution (QKD). Here, a set of eavesdropping attacks can be described and analyzed using the approximate cloning framework, which leads to security proofs for QKD.

For more information on quantum cloning, see [Sca+05].

9. Spectrum estimation

9.1. Problem setup

Density operators describe the state of a quantum system. Mathematically, ρ is a quantum state iff ρ is positive semidefinite and $\operatorname{tr}\rho = 1$. A quantum state ρ has a spectral decomposition $\rho = \sum_{i=1}^{d} \lambda_i |e_i\rangle\langle e_i|$ with eigenvalues $(\lambda_i)_{i=1}^{d}$ satisfying $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$ and eigenvectors $\{|e_i\rangle\}_{i=1}^{d}$ $\langle e_i|e_j\rangle = \delta_{ij}$. In this chapter we are interested in the task of estimating the (unknown) density operators ρ of a quantum system. We focus on estimating the spectrum $\{\lambda_i\}_{i=1}^{d}$ of ρ. We make two assumptions:

(1) We have access to an experiment that prepares the system (exactly) in the state ρ.

(2) We can run this experiment n times and perform joint measurements on all n copies at the same time. The idea is that we will estimate spectrum of ρ by measuring $\rho^{\otimes n}$.

The goal is to devise a strategy that gives exact result with probability approaching 1 as $n \to \infty$.

9.2. Symmetries of spectrum estimation

The state $\rho^{\otimes n}$ is permutation invariant:

$$Q_\pi \rho^{\otimes n} Q_\pi^\dagger = \rho^{\otimes n}$$

for all $\pi \in S_n$.

Hence, without loss of generality, the desired measurement also has permutation invariance, since for any $P \geq 0$ we have

$$\operatorname{tr}\left(P\rho^{\otimes n}\right) = \operatorname{tr}\left(PQ_\pi \rho^{\otimes n} Q_\pi^\dagger\right) = \operatorname{tr}\left(Q_\pi^\dagger P Q_\pi \rho^{\otimes n}\right),$$

and so $\operatorname{tr}\left(P\rho^{\otimes n}\right) = \operatorname{tr}\left(\overline{P}\rho^{\otimes n}\right)$ with

$$\overline{P} = \frac{1}{n!} \sum_{\pi \in S_n} Q_\pi P Q_\pi^\dagger.$$

We also know that ρ and $U\rho U^\dagger$ have the same eigenvalues for any unitary $U \in \mathrm{U}_d$. Therefore, we can impose $U^{\otimes n}$ invariance on measurement operators as well.

We have both S_n and U_d invariance, so Schur-Weyl duality gives the decomposition

$$(\mathbb{C}^d)^{\otimes n} = \bigoplus_{\lambda \vdash_d n} \underbrace{V_\lambda}_{S_n \text{ irrep}} \otimes \underbrace{W_\lambda}_{\mathrm{U}_d \text{ irrep}} .$$

For $\lambda \vdash_d n$ let P_λ be the projection onto $V_\lambda \otimes W_\lambda$, then $P_\lambda \geq 0$ and $\sum_{\lambda \vdash_d n} P_\lambda = \mathbb{1}_{\mathbb{C}^d}^{\otimes n}$ (that is, P_λ is a measurement). Furthermore $[P_\lambda, Q_\pi] = 0$ for all $\pi \in S_n$ and $[P_\lambda, U^{\otimes n}] = 0$ for all $U \in \mathrm{U}_d$. Invariance under both S_n and U_d is satisfied by P_λ; it is thus a good candidate for spectrum measurement.

Question. What does the outcome "$\lambda \vdash_d n$" mean?

Observation. Let $\lambda = (\lambda_1, ..., \lambda_d) \vdash_d n$, that is, $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$ and $\sum_{i=1}^{d} \lambda_i = n$. Therefore $\overline{\lambda} := \frac{\lambda}{n}$ is a valid spectrum of a quantum state! That is, $\overline{\lambda_i} \geq 0$ and $\sum_{i=1}^{d} \overline{\lambda_i} = 1$.

Idea of spectrum estimation. Let ρ have spectrum $r = (r_1, ..., r_d)$ (WLOG $v_1 \geq v_2 \geq \cdots \geq v_d$).

(1) Measure $\rho^{\otimes n}$ w.r.t $\{P_\lambda\}$.

(2) For outcome $\lambda \vdash_d n$, set $\hat{r} = \frac{\lambda}{n}$.

(3) $\operatorname{Prob}(\hat{r} \neq r) \to 0$ as $n \to \infty$.

The measurement in (1) is often called *weak Schur sampling*. The main result of this chapter is to show (3).

9.3. Weak Schur sampling

Our goal is to bound the probability of obtaining outcome "λ" (where $\lambda \vdash_d n$ is a Young diagram) in weak Schur sampling. That is, denoting by P_λ the projector onto $V_\lambda \otimes W_\lambda$ in the Schur-Weyl decomposition, we want to bound

$$\operatorname{tr}\left(P_\lambda \rho^{\otimes n}\right),$$

where ρ is the unknown quantum state whose spectrum we want to estimate.

Since $Q_\pi \rho^{\otimes n} Q_\pi^\dagger = \rho^{\otimes n}$, we can write

$$\rho^{\otimes n} = \bigoplus_{\lambda \vdash_d n} \mathbb{1}_{V_\lambda} \otimes \rho_\lambda$$

for some positive semidefinite operators $\rho_\lambda \in \operatorname{End}(W_\lambda)$.
Recall that $W_\lambda = e_T(\mathbb{C}^d)^{\otimes n}$, where T is the standard Young tableau of shape $\lambda \vdash_d n$. The first step is to characterize W_λ so that we understand the effect of P_λ on $\rho^{\otimes n}$.

Definition 33. Let $x, y \in \mathbb{R}^d$, and denote by $x^\downarrow$, $y^\downarrow$ the vectors of components of x, y sorted in non-increasing order (e.g. $x_1^\downarrow \geq \cdots \geq x_d^\downarrow$). Then y is said to *majorize* x, in symbols $x \prec y$ if

- $\sum_{i=1}^q x_i^\downarrow \leq \sum_{i=1}^q y_i^\downarrow$ for all $q = 1, ..., d-1$,
- $\sum_{i=1}^d x_i = \sum_{i=1}^d y_i$.

Now consider the spectral decomposition $\rho = \sum_{i=1}^d r_i |e_i\rangle\langle e_i|$, and form the tensor product basis $B = \{\bigotimes_{j=1}^n |e_{i_j}\rangle : i_j \in [d]\}$ of $(\mathbb{C}^d)^{\otimes n}$. For $|v\rangle \in B$ let $f = (f_1, ..., f_d)$ be the *frequency distribution* of $|v\rangle$: f_i is the number of times $|e_i\rangle$ appears in $|v\rangle$. Note that f is an (ordered) partition of n.

Lemma 9.1. Let $|v\rangle \in B$ with frequency distribution f, and let T be the standard Young tableau of shape $\lambda \vdash_d n$. Then $e_T|v\rangle = 0$ unless $f \prec \lambda$.

Proof. Observe that if T has a column with indices j and k such that $|e_{i_j}\rangle = |e_{i_k}\rangle$ in $|v\rangle$, then $e_T|v\rangle = 0$. This is because $e_T \propto v_T c_T$ antisymmetrizes over columns and $c_T = c_T(\mathbb{1} - (jk))$ (proof: exercise.).
Now WLOG assume $f_1 \geq f_2 \geq \cdots \geq f_d$. If $e_T|v\rangle \neq 0$, then $f_1 \leq \lambda_1$ (length of the first row of λ), because otherwise some column would have two indices j and k with $|e_{i_j}\rangle = |e_{i_k}\rangle$ in $|v\rangle$ (where $i_j = i_k$ has frequency f_1), in which case $e_T|v\rangle = 0$ (the basis elements $|e_{i_j}\rangle$ "spill over" into the second row).
Likewise, if $f_1 + f_2 > \lambda_1 + \lambda_2$, then the same thing happens in row 3 or further down, hence $f_1 + f_2 \geq \lambda_1 + \lambda_2$ if $e_T|v\rangle \neq 0$.
Continuing in this manner, we get

$$\sum_{i=1}^q f_i \leq \sum_{i=1}^q \lambda_i \text{ for all } q = 1, ..., d-1,$$

and

$$\sum_{i=1}^d f_i = n = \sum_{i=1}^d \lambda_i$$

if $e_T|v\rangle \neq 0$. ■

Proposition 9.2. Let ρ be a density operator with spectrum $r = (r_1, ..., r_d)$ where $r_1 \geq r_2 \geq \cdots \geq r_d$. Let $\lambda = (\lambda_1, ..., \lambda_d) \vdash_d n$ and $\bar{\lambda} = \frac{\lambda}{n}$. Then

$$\operatorname{tr}\left(P_\lambda \rho^{\otimes n}\right) \leq (n+1)^{\frac{d(d-1)}{2}} \exp\left(-nD(\bar{\lambda}||r)\right),$$

where $D(\bar{\lambda}||r)$ is the Kullback-Leibler divergence between $\bar{\lambda}$ and r, defined by

$$D(p||q) = \sum_i p_i \log \frac{p_i}{q_i}$$

for probability distributions p, q with $\operatorname{supp} p := \{i : p_i \neq 0\} \subset \operatorname{supp} q$. The KL divergence satisfies $D(p||q) \geq 0$ and $D(p||q) = 0$ if and only if $p = q$.

Proof. Recall that for $\lambda \vdash_d n$ we denote by $\mathrm{SYT}(\lambda)$ the set of standard Young tableau of shape λ. Then

$$P_\lambda = \sum_{T \in \mathrm{SYT}(\lambda)} e_T,$$

with the Young projector e_T associated to $T \in \mathrm{SYT}(\lambda)$. Note that

$$|\mathrm{SYT}(\lambda)| = \dim V_\lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h(i,j)} \leq \frac{n!}{\prod_{i=1}^d \lambda_i!}$$

where the last bound is left as an exercise.
Hence for $\lambda \vdash_d n$, we have

$$\operatorname{tr}\left(P_\lambda \rho^{\otimes n}\right) = \sum_{T \in \mathrm{SYT}(\lambda)} \operatorname{tr}\left(e_T \rho^{\otimes n}\right).$$

Fix some $T \in \mathrm{SYT}(\lambda)$, and recall that $\rho^{\otimes n}$ has eigenvectors $|v\rangle \in B$ (with B the tensor product basis of eigenvectors of ρ defined previously) with eigenvalues $\prod_i r_i^{f_i}$, where $f = (f_1, ..., f_d)$ is the frequency distribution of $|v\rangle$. We can thus write

$$\rho^{\otimes n} = \sum_{|v\rangle \in B} \prod_i r_i^{f_i} |v\rangle\langle v|.$$

Now, by the previous lemma,

$$\begin{aligned} \operatorname{tr}\left(e_T \rho^{\otimes n}\right) &= \sum_{|v\rangle \in B} \prod_i r_i^{f_i} \operatorname{tr}(e_T |v\rangle\langle v|) && (72) \\ &= \sum_{\substack{|v\rangle \in B \\ f \prec \lambda}} \prod_i r_i^{f_i} \operatorname{tr}(e_T |v\rangle\langle v|), && (73) \end{aligned}$$

To bound this expression further, we use the following simple fact from majorization theory (see exercises):
if $x \prec y$ and $u \in \mathbb{R}^d$ is arbitrary, $\langle x^\downarrow, u^\downarrow \rangle \leq \langle y^\downarrow, u^\downarrow \rangle$.
Choosing $x = f, y = \lambda$ and $u = (\log r_1, ..., \log r_d)$, we get

$$\langle f, u \rangle = \sum_{i=1}^d f_i \log r_i \leq \sum_{i=1}^d \lambda_i \log r_i = \langle \lambda, u \rangle.$$

Exponentiating this yields $\prod_{i=1}^d r_i^{f_i} \leq \prod_{i=1}^d r_i^{\lambda_i}$, and hence

$$\begin{aligned} \operatorname{tr}\left(e_T \rho^{\otimes n}\right) &= \sum_{\substack{|v\rangle \in B \\ f \prec \lambda}} \prod_i r_i^{f_i} \operatorname{tr}(e_T |v\rangle\langle v|) && (74) \\ &\leq \prod_i r_i^{\lambda_i} \operatorname{tr}\left[e_T \sum_{\substack{|v\rangle \in B \\ f \prec \lambda}} |v\rangle\langle v| \right] && (75) \\ &\leq \prod_i r_i^{\lambda_i} \operatorname{tr} e_T \qquad \text{since } \sum_{\substack{|v\rangle \in B \\ f \prec \lambda}} |v\rangle\langle v| \leq \mathbb{1} && (76) \\ &= \prod_i r_i^{\lambda_i} \dim W_\lambda && (77) \\ &\leq \prod_i r_i^{\lambda_i} (n+1)^{\frac{d(d-1)}{2}}, && (78) \end{aligned}$$

where we used the dimension bound $\dim W_\lambda \leq (n+1)^{\frac{d(d-1)}{2}}$ (see e.g. Christandl's PhD thesis.) Putting everything together we have

$$\begin{aligned} \operatorname{tr}\left(P_\lambda \rho^{\otimes n}\right) &= \sum_{T \in \mathrm{SYT}(\lambda)} \operatorname{tr}\left(e_T \rho^{\otimes n}\right) && (79)\\ &\leq (n+1)^{\frac{d(d-1)}{2}} \sum_{T \in \mathrm{SYT}(\lambda)} \prod_i r_i^{\lambda_i} && (80)\\ &= (n+1)^{\frac{d(d-1)}{2}} \frac{n!}{\prod_i \lambda_i!} \prod_i r_i^{\lambda_i}. && (81) \end{aligned}$$

The result now follows from a well-known bound on the multinomial coefficient $\binom{n}{\lambda} = \frac{n!}{\lambda_1! \cdots \lambda_d!}$, namely $\binom{n}{\lambda} \leq \prod_{i=1}^d \left(\frac{n}{\lambda_i}\right)^{\lambda_i}$ together with the observation that

$$\begin{aligned} -nD(\overline{\lambda}||r) &= -n \sum_i \frac{\lambda_i}{n} \log\left(\frac{\lambda_i}{nr_i}\right) && (82)\\ &= \sum_i -\lambda_i \log\left(\frac{\lambda_i}{nr_i}\right) && (83)\\ &= \sum_i \log\left(\frac{nr_i}{\lambda_i}\right)^{\lambda_i}, && (84) \end{aligned}$$

so that $\exp\left(-nD(\overline{\lambda}||r)\right) = \prod_i r_i^{\lambda_i} (\frac{n}{\lambda_i})^{\lambda_i}$. ■

9.4. Asymptotics of spectrum estimation

We have proved that for a quantum state ρ with spectrum $r = (r_1, ..., r_d)$, $r_i \geq r_{i+1}$ and $\lambda \vdash_d n$,

$$\operatorname{tr}\left(P_\lambda \rho^{\otimes n}\right) \leq (n+1)^{\frac{d(d-1)}{2}} \exp\left(-nD(\overline{\lambda}||r)\right),$$

where $\overline{\lambda} = \frac{\lambda}{n}$ and $D(\cdot||\cdot)$ is the so called relative entropy. We can extend this bound to a set S of possible spectra as follows. Set

$$P_S = \sum_{\substack{\lambda \vdash n \\ \overline{\lambda} \in S}} P_\lambda,$$

and note that

$$\operatorname{tr}\left(P_S \rho^{\otimes n}\right) \leq (n+1)^{\frac{d(d-1)}{2}} \exp\left(-n \min_{\substack{\lambda \vdash n \\ \overline{\lambda} \in S}} D(\overline{\lambda}||r)\right),$$

which follows from picking the λ with the slowest convergence, or equivalently the minimum $D(\overline{\lambda}||r)$, and using

$$|S| \leq |\{\lambda \vdash_d n\} \leq (n+1)^d.$$

(this heavily overestimates the number of Young diagrams with n boxes in d rows, but it is sufficient for our purposes.)
Finally we consider the $\varepsilon-$ball

$$B_\varepsilon = \{r' : \sum_i |r_i - r_i'| < \varepsilon\}$$

around the true spectrum r. Choosing $S = \overline{B_\varepsilon(r)}$, we obtain:

Proposition 9.3. Let ρ be a quantum state with (ordered) spectrum $r = (r_1, ..., r_d)$, and for given $\delta > 0$, let

$$P_\lambda = \sum_{\substack{\lambda \vdash n \\ \overline{\lambda} \in B_\delta(r)}} P_\lambda.$$

Then for any $\varepsilon > 0$ there exists n_0 such that for all $n \geq n_0$

$$\operatorname{tr}\left(P_\lambda \rho^{\otimes n}\right) \geq 1 - \varepsilon.$$

10. Quantum state tomography

10.1. Warm-up: pure state estimation

Quantum state tomography is the task of obtaining a classical description of an unknown quantum state ρ. We make the same assumptions as in spectrum estimation:

(1) We can prepare identical copies of the unknown state ρ.
(2) We can make joint measurements on all copies simultaneously.

The task then is to find a measurement on $\rho^{\otimes n}$ that yields asymptotically accurate estimate of ρ.
In this section we will consider the simpler task of pure state estimation: assuming a quantum system $\mathcal{H}$ is prepared in an unknown pure state $|\psi\rangle$, we want to estimate ψ by measuring $|\psi\rangle\langle\psi|^{\otimes n}$.
We know from previous chapters:

(1) $\psi^{\otimes n}$ is permutation-invariant, and moreover

(2) $\psi^{\otimes n}$ is supported on the symmetric subspace.

Our ansatz for the measurement will be

$$\{\binom{n+d-1}{n}|\varphi\rangle\langle\varphi|^{\otimes n}\}_{|\psi\rangle\in\mathcal{H}'}$$

where $|\varphi\rangle$ is distributed according to the Haar measure $\mathrm{d}\varphi$ on pure states.
This is a continuous POVM on $\operatorname{Sym}^n(\mathcal{H})$, it satisfies:

- $|\varphi\rangle\langle\varphi|^{\otimes n} \geq 0$ for all $|\varphi\rangle \in \mathcal{H}$.
- $\binom{n+d-1}{n}\int \mathrm{d}\varphi\, |\varphi\rangle\langle\varphi|^{\otimes n}$ is the projector onto $\operatorname{Sym}^n(\mathcal{H})$ and hence equal to the identity on that space.

Remark. Recall from Chapter 5 that projective measurements correspond to the spectral decomposition of Hermitian observables. A POVM can be "purified" to a projective measurement on a larger space (which is the system and its environment). How can we then make sense of a continuous POVM on a finite-dimensional Hilbert space?
This was answered by Chiribella et al. in [CDS07]. A continuous POVM can be expressed as a continuous random variable taking values in some set Ω with probability density p_Ω. For each $\omega \in \Omega$, there is a discrete finite POVM M_ω. The outcome of the continuous POVM is obtained as follows:

(1) sample $\omega \in \Omega$ according to p_Ω
(2) Measure system with M_ω.

Using the ansatz for the measurement given by $\{Q_\varphi\}$ with

$$Q_\varphi := \binom{n+d-1}{n}|\varphi\rangle\langle\varphi|^{\otimes n},$$

we propose the following protocol for pure state estimation:

(1) Measure $|\psi\rangle\langle\psi|^{\otimes n}$ with respect to $\{Q_\varphi\}$.
(2) Outcome $|\hat{\psi}\rangle$ is our estimator for $|\psi\rangle$.

Claim.

$$\mathbb{E}_{\hat{\psi}}(F(\hat{\psi},\psi)^2) = \mathbb{E}_{\hat{\psi}}\left(\left|\langle\hat{\psi}\rangle\,|\psi\rangle\right|^2\right) \geq 1 - \frac{d}{n},$$

and so

$$\mathbb{E}_{\hat{\psi}}(D(\hat{\psi},\psi)) \leq \sqrt{\frac{d}{n}} \to 0 \text{ as } n \to \infty.$$

The proof of this claim is left as an exercise. This was essentially calculated in Chapter 7.

10.2. Symmetries of state tomography

Suppose we have an unknown quantum state ρ with spectral decomposition $\rho = \sum_i |e_i\rangle\langle e_i|\lambda_i$. Our goal is to find a measurement on $\rho^{\otimes n}$ that yields asymptotically accurate estimates of ρ. In the previous section for pure state estimation, our ansatz for the measurement was a collection of POVM operators proportional to $|\varphi\rangle\langle\varphi|^{\otimes n} = \left(U|\phi_0\rangle\langle\phi_0|U^\dagger\right)^{\otimes n}$, where $|\varphi_0\rangle$ is a fixed pure state with spectrum $(1, 0, \ldots, 0)$ and U is a random unitary operator.

The idea for full tomography is to replace $|\phi_0\rangle\langle\phi_0|$ above by a diagonal state with an estimate of the spectrum and take Haar-random U as estimate for the eigenbasis. The guess for $\mathrm{Spec}(\rho)$ is obtained as in spectrum estimation: measure $\rho^{\otimes n}$ with respect to the Schur-Weyl decomposition $\{P_\lambda\}_{\lambda\vdash n}$ where P_λ projects onto the λ component in $(\mathbb{C}^d)^{\otimes n} = \oplus_{\lambda\vdash n} V_\lambda \otimes W_\lambda$.
It remains now to incorporate U in the measurement. To this end, denote by $\mathcal{D}(\mathcal{H})$ the set $\{\rho \in \mathcal{L}(\mathcal{H}) : \rho \geq 0, \operatorname{tr}\rho = 1\}$ of density operators on $\mathcal{H}$. We look for a continuous POVM $\{M_\sigma\}_{\sigma\in\mathcal{D}(\mathcal{H})}$ with the following symmetries:

(1) Permutation invariance: $Q_\pi M_\sigma Q_\pi^\dagger = M_\sigma$ for all $\pi \in S_n, \sigma \in \mathcal{D}(\mathcal{H})$ (since $\rho^{\otimes n}$ has this symmetry.)
(2) Unitary covariance: $M_{U\sigma U^\dagger} = U^{\otimes n} M_\sigma (U^\dagger)^{\otimes n}$ for all $U \in \mathrm{U}_d, \sigma \in \mathcal{D}(\mathcal{H})$ since we consider ρ and $U\rho U^\dagger$ to be equally likely, and hence

$$\operatorname{tr}(\rho^{\otimes n} M_\sigma) = \operatorname{tr}\Big(U^{\otimes n}\rho^{\otimes n}(U^\dagger)^{\otimes n} U^{\otimes n} M_\sigma (U^\dagger)^{\otimes n}\Big) \tag{85}$$

$$= \operatorname{tr}\left((U\rho U^\dagger)^{\otimes n} M_{U\sigma U^\dagger}\right). \tag{86}$$

We now come up with an ansatz satisfying the two symmetries above. Let $\lambda \vdash_d n$ be a Young diagram. Set $\bar{\lambda} = \frac{1}{n}\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and for $U \in \mathrm{U}_d$ set $\sigma(\lambda, U) := U\bar{\lambda}U^\dagger$. Define measurement operators for $\sigma = \sigma(\lambda, U)$ as

$$M_\sigma = M(\lambda, U) = c_\lambda P_\lambda (U\bar{\lambda}U^\dagger)^{\otimes n} P_\lambda. \tag{87}$$

On outcome (λ, U) we take the estimator $\sigma = U\bar{\lambda}U^\dagger$. It remains to determine the constant c_λ.
Observe that the operator $M_\lambda := \int \mathrm{d}U\, M(\lambda, U)$ satisfies:

- $Q_\pi M_\lambda Q_\pi^\dagger = M_\lambda$ for all $\pi \in S_n$.
- $U^{\otimes n} M_\lambda (U^\dagger)^{\otimes n} = M_\lambda$ for all $U \in \mathrm{U}_d$.
- $M_\lambda = P_\lambda M_\lambda P_\lambda = \in \operatorname{End}(V_\lambda \otimes W_\lambda)$. (recall that Schur Weyl duality gives the decomposition $(\mathbb{C}^d)^{\otimes n} = \oplus_{\lambda\vdash n} V_\lambda \otimes W_\lambda$) where V_λ has dimension d_λ and W_λ has dimension m_λ.

We would like $M_\lambda = P_\lambda$ to be true, since then $\mathbb{1} = \sum_\lambda P_\lambda = \sum_\lambda M_\lambda = \sum_\lambda \int \mathrm{d}U\, M(\lambda, U)$.
To compute c_λ, take the trace of both sides of the equation 87:

$$\operatorname{tr} M_\lambda = c_\lambda \int \mathrm{d}U \operatorname{tr}\left[P_\lambda\left((U\bar{\lambda}U^\dagger)^{\otimes n}\right)P_\lambda\right] \tag{88}$$

$$\tag{89}$$

We now use the fact that $P_\lambda (U\bar{\lambda}U^\dagger)^{\otimes n} P_\lambda = \mathbb{1}_{V_\lambda} \otimes w_\lambda(U\bar{\lambda}U^\dagger)$, where w_λ is the irrep of $\mathrm{GL}(\mathcal{H})$ on $\mathcal{H}^{\otimes n}$ labeled by λ. Thus,

$$\operatorname{tr}\left[P_\lambda (U\bar{\lambda}U^\dagger)^{\otimes n} P_\lambda\right] = \operatorname{tr}\left[\mathbb{1}_{V_\lambda} \otimes w_\lambda(U\bar{\lambda}U^\dagger)\right] \tag{90}$$

$$= d_\lambda \operatorname{tr}\Big(w_\lambda(U\bar{\lambda}U^\dagger)\Big) \tag{91}$$

$$= d_\lambda s_\lambda(U\bar{\lambda}U^\dagger) \tag{92}$$

where the *Schur polynomial* s_λ is the character of the irreducible representation (w_λ, W_λ) of $\mathrm{GL}(\mathcal{H})$ (or U_d) on $\mathcal{H}^{\otimes n}$ labeled by $\lambda \vdash_d n$.
Since characters are functions of eigenvalues (as traces), we have $s_\lambda(U\bar{\lambda}U^\dagger) = s_\lambda(\bar{\lambda})$ and hence

$$
\begin{aligned}
\operatorname{tr}(P_\lambda) &= \dim(V_\lambda \otimes W_\lambda) = d_\lambda m_\lambda && (93)\\
\operatorname{tr} M_\lambda &= c_\lambda \int \mathrm{d}U \operatorname{tr}\Big(P_\lambda (U\bar{\lambda}U^\dagger)^{\otimes n} P_\lambda\Big) && (94)\\
&= c_\lambda \int \mathrm{d}U\, d_\lambda s_\lambda(\bar{\lambda}) && (95)\\
&= c_\lambda d_\lambda s_\lambda(\bar{\lambda}). && (96)
\end{aligned}
$$

Finally, using $\operatorname{tr} P_\lambda = \operatorname{tr} M_\lambda$ gives $c_\lambda = \frac{m_\lambda}{s_\lambda \bar{\lambda}}$.

10.3. Error analysis of our tomography protocol

We will need the following bounds on Schur polynomials:

Lemma 10.1. Let $\lambda \vdash_d n$.

(1) For $\bar{\lambda} = \frac{1}{n}\operatorname{diag}(\lambda)$, we have $s_\lambda(\bar{\lambda}) \geq e^{-nH(\bar{\lambda})}$, where $H(X) = \sum_i -x_i \log x_i$ is the Shannon entropy of a probability distribution $x = (x_1, x_2, \ldots, x_d)$.

(2) Let $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ with $F(\rho,\sigma) = F$. Assume that $\operatorname{rank}(\rho) \leq r \leq d$. Then,

$$
s_\lambda(\rho\sigma) = \begin{cases} = 0 & \text{if} \quad \lambda_{r+1} > 0 \\ \leq m_\lambda e^{-2nH(\bar{\lambda})} F^{2n} & \text{otherwise.} \end{cases}
$$

Proof. See [Haa+16]. ■

We also record the bound

$$
d_\lambda = \dim V_\lambda \leq e^{nH(\bar{\lambda})},
$$

which we implicitly used in the spectrum estimation chapter.

With these bounds in place we can estimate the probabilities of our continuous tomography POVM $\{M_\lambda, U\}_{\lambda \vdash n, U \in \mathrm{U}_d}$:

$$
\begin{aligned}
\operatorname{tr}\left[M(\lambda,U)\rho^{\otimes n}\right] &= \frac{m_\lambda}{s_\lambda(\bar{\lambda})} \operatorname{tr}\left[P_\lambda (U\bar{\lambda}U^\dagger)^{\otimes n} P_\lambda \rho^{\otimes n}\right] && (97)\\
&= \frac{m_\lambda}{s_\lambda(\bar{\lambda})} \operatorname{tr}\left[\underbrace{P_\lambda (U\bar{\lambda}U^\dagger)^{\otimes n} P_\lambda}_{\mathbb{1}_{V_\lambda} \otimes w_\lambda(U\bar{\lambda}U^\dagger)} \underbrace{P_\lambda \rho^{\otimes n} P_\lambda}_{\mathbb{1}_{V_\lambda} \otimes w_\lambda(\rho)}\right] && (98)\\
&= \frac{m_\lambda}{s_\lambda(\bar{\lambda})} \operatorname{tr} \underbrace{\left[w_\lambda(U\bar{\lambda}U^\dagger) w_\lambda(\rho)\right]}_{w_\lambda(U\bar{\lambda}U^\dagger \rho)} && (99)\\
&= \frac{m_\lambda}{s_\lambda(\bar{\lambda})} s_\lambda(U\bar{\lambda}U^\dagger \rho) && (100)\\
&\leq m_\lambda e^{2nH(\bar{\lambda})} m_\lambda e^{-2nH(\bar{\lambda})} F^{2n} && (101)\\
&\leq m_\lambda^2 F^{2n} && (102)\\
&\leq (n+1)^{2dr} F^{2n}, && (103)
\end{aligned}
$$

where we used $m_\lambda \leq (n+1)^{dr}$ for λ with $\lambda_k = 0$ for $k > r+1$ and we set $F = F(\rho, U\bar{\lambda}U^\dagger)$.Thus, for all $\varepsilon > 0$, with $\hat{\rho} = U\bar{\lambda}U^\dagger$, we have $\Pr(F(\rho,\hat{\rho}) \leq 1-\varepsilon) \leq (n+1)^{2dr}(1-\varepsilon)^{2n}$.

For more details on this proof, see [Haa+16].

11. Universal Quantum Source Compression

11.1. Classical sources and entropy

Consider a classical random variable X that emits symbols $x \in \{1, \dots, d\}$ with probability p_x. As an example, a biased coin gives H with probability $p \in [0, 1]$ and T with probability $1 - p$.
We assume that we receive a sequence of symbols from $X : x^n = (x_1, \dots, x_n) \in \{1, \dots, d\}^n$. A common assumption is that the source is independent and identically distributed (i.i.d.) with probability distribution or *memoryless*, and so $\Pr(x^n) = \prod_{i=1}^{n} \Pr(x_i)$.
A central question in information theory is to ask how much we information we gain when we learn $x^n = (x_1, \dots, x_n)$. Two extreme examples are:

(1) For a *deterministic source* with $p_{\hat{x}} = 1$ for some fixed $\hat{x} \in [d]$ and $p_y = 0$ for all $y \neq \hat{x}$, we learn nothing new when we receive $x^n = (\hat{x}, \dots, \hat{x})$.

(2) For a *uniformly random source* with $p_x = \frac{1}{d}$ for all $x \in [d]$, all output sequences x^n are equally probable (with probability $\frac{1}{d^n}$), and so a specific observed sequence x^n conveys a lot of information.

One of Shannon's many contributions was to make these observations quantitative using the concept of *entropy*.

Definition 34. For a random variable $X \sim p_x$, the *surprisal* of an event $x \in [d]$ is defined as

$$I(X) := \log \frac{1}{p_x} = -\log p_x.$$

Intuitively, the less likely an event is, the more surprising it is, and the more information we gain. The expected surprisal of a random variable X is called the *entropy* of the source X.

Definition 35. For a random variable $X \sim p_x$, the *entropy* of the source X is defined as

$$H(X) := \sum_{x \in [d]} p_x I(X) = -\sum_{x \in [d]} p_x \log p_x.$$

Note that we use the convention that $0 \log 0 = 0$, since $x \log x \to 0$ as $x \to 0$. Hence, if $p_x = 0$ for some x, then x has infinite surprisal, but receives no weight in $H(X)$.
Properties of Shannon entropy.

(1) For any random variable X taking values in $[d]$, $0 \leq H(X) \leq \log d$. The bounds are achieved when X is deterministic or uniformly random, respectively, as in the two extreme examples above. That is, X is deterministic iff $H(X) = 0$ and X is uniform iff $H(X) = \log d$.

(2) Concavity: let $X_1 \sim p_x$ and $X_2 \sim q_x$ be random variables on the same alphabet, and for $\lambda \in [0, 1]$ define a random variable $Z = \lambda X_1 + (1 - \lambda) X_2$. Then $H(Z) \geq \lambda H(X_1) + (1 - \lambda) H(X_2)$.

11.2. Compressing a classical source

Our task is to compress the signals $x^n = (x_1, \dots, x_n)$ of an iid source $X \sim p_x$ without losing information in the limit $n \to \infty$. The previous section suggests that information content of a source X is quantified by Shannon entropy $H(X)$. Shannon proved in 1948 in [Sha48] that $H(X)$ is the optimal compression rate.

The idea of source compression is to use the fact that some output signals of the source occur more frequently (determined by iid probability distribution $p^{\times n}$)) than others, and hence there is redundancy in the information. There are two ways of carrying out compression: *variable-length* and *fixed-length*. In variable-length coding, more frequent signals are assigned shorter code words (for instance in Huffman coding), and in fixed-length coding, all signals are assigned the same length code words (decoding in this case is easier). We focus on fixed-length coding. How do we characterize the "frequent" signals of a source? We use the concept of *typicality*, which we now define.

Definition 36. Let $(X_i)_{i \in \mathbb{N}}$ be iid random variables each taking values in $[d]$ and with common probability mass function p_x, $x \in [d]$. For $x^n = (x_1, x_2, \dots, x_n) \in [d]^n$, let $p(x^n) = \prod_{i=1}^{n} p_{x_i}$. Fixing $\varepsilon > 0$, the *ε-typical set* $T_\varepsilon^{(n)}$ consists of those sequences $x^n \in [d]^n$ for which

$$2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}$$

where $X \sim p_x$.

This definition captures a notion of *typicality* in the following sense. Assume that each letter $x \in [d]$ appears roughly np_x times in a "typical" sequence x^n. Then

$$p(x^n) \cong \prod_{x\in[d]} p_x^{np_x} = \prod_{x\in[d]} 2^{np_x \log p_x} \tag{104}$$
$$= 2^{n\sum_{x\in[d]} p_x \log p_x} \tag{105}$$
$$= 2^{-nH(X)}. \tag{106}$$

Theorem 11.1. Fix $\varepsilon > 0$. For any $\delta > 0$, there is $n_0 \in \mathbb{N}$ such that the following statements hold for all $n \geq n_0$:

(1) $H(X) - \varepsilon \leq -\frac{1}{n}\log p(x^n) \leq H(X) + \varepsilon$ for all $x^n \in T_\varepsilon^{(n)}$.
(2) $\Pr(T_\varepsilon^{(n)}) \geq 1 - \delta$.
(3) $|T_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$.
(4) $|T_\varepsilon^{(n)}| \geq (1-\delta)2^{n(H(X)-\varepsilon)}$.

Proof. See Ch. 14 in [Wil16]. ■

We now prove that any rate $R > H(X)$ is achievable, which is one part of Shannon's compression theorem. Suppose we have an iid source $X \sim p_x$. Fix a rate $R > H(X)$ and choose $\varepsilon > 0$ such that $H(X) + \varepsilon < R$. For any $\delta > 0$, there is n_0 such that for $n \geq n_0$, there are at most $|T_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)} < 2^{nR}$ typical sequences. Now:

(1) Index elements in $T^{(n)_\varepsilon}$ in some way using no more that $\lceil nR \rceil$ bits.
(2) Encoding: for a received signal x^n, decide if $x^n \in T_\varepsilon^{(n)}$. If yes, encode x^n using the index of x^n in $T_\varepsilon^{(n)}$, and prefix with symbol 1. If no, encode x^n using a fixed-length code of length $\lceil nR \rceil$, and prefix with 0.
(3) Decoding: on receiving sequence $1\ldots$, output respective typical sequence. On receiving sequence $0\ldots$, declare an error. The latter only occurs with a probability at most δ.

In the limit $n \to \infty$ this defines a code with rate $r = \lim_{n\to\infty} \frac{1}{n}(nR+1) = R$ and error $e \to 0$.
Conversely, any code with rate $R < H(X)$ necessarily has error $e \not\to 0$ as $n \to \infty$. The proof uses typicality again. This proves that the Shannon entropy $H(X)$ is the optimal compression rate.

11.3. Strong typicality and universal compression

In the previous section we saw how source compression can be done based on typicality. The advantage of this approach is that it has an easy proof using the law of large numbers. The disadvantage, however, is that encoding and decoding depend on source statistics. In this section we will devise a code that only depends on the entropy of the source, which is the optimal source compression rate. We will use the concept of *strong typicality*, which we now define.
For a sequence $x^n = (x_1, \ldots, x_n) \in [d]^n$ and $x \in [d]$, let

$$N(x|x^n) = |\{i : x_i = x\}|\,.$$

Definition 37. The type t_{x^n} of a sequence x^n is a probability distribution on $[d]$ defined as $t_{x^n}(x) = \frac{1}{n}N(x|x^n)$.

For example, let $d = 3, n = 5, x^n = (0,1,0,2,2)$. Then x^n has type $t_{x^n} = (0.4, 0.2, 0.4)$.
Since $N(x|x^n)$ can only take $n+1$ possible values, there are at most $(n+1)^d$ possible types. This is only polynomial in the sequence length n. Let $T_P \subset [d]^n$ denote the set of sequences x^n of type $t_{x^n} = P$. Then

$$(n+1)^{-d}2^{nH(P)} \leq |T_P| \leq 2^{nH(P)}.$$

For a proof see [CK81].

Definition 38. Let $X \sim p_x$ be a source on $[d]$, and fix $\varepsilon > 0$. A sequence x^n is said to be *ε-strongly typical* if $|t_{x^n} - p_x| \leq \varepsilon$ for all $x \in [d]$ such that $p_x > 0$ and $N(x|x^n) = 0$ if $p_x = 0$. The set of all ε-strongly typical sequences is denoted by $T_{X,\varepsilon}^{(n)}$.

Proposition 11.2. Properties of strongly typical sequences:

(1) For all $\delta > 0$, we have $\Pr(T^{(n)}_{X,\varepsilon}) \geq 1 - \delta$.

(2) $|\frac{1}{n}\log|T^{(n)}_{X,\varepsilon}| - H(X)| \leq c\varepsilon$ for some $c > 0$ and sufficiently large n.

(3) For some constant $c > 0$,

$$2^{-n(H(X)+c\varepsilon)} \leq \Pr(x^n) \leq 2^{-n(H(X)-c\varepsilon)}.$$

Proof. See 14.7 in [Wil16]. ∎

The third property in the proposition above says that strong typicality implies typicality as defined in Section 11.2, which is often called weak typicality. The first two properties in the proposition above give rise to a source compression protocol that only depends on $H(X)$: for fixed $R > H(X)$, define

$$A^{(n)} := \cup_{P:H(P)<R} T_P,$$

the set of all sequences of type P such that $H(P) < R$. Then we have by [CK81]

$$|A^{(n)}| \leq (n+1)^d 2^{nR} \tag{107}$$

because $|T_P| \leq 2^{nH(P)}$ and the number of types is at most $(n+1)^d$. We also have

$$\Pr(x^n \notin A^{(n)}) \leq (n+1)^d \exp\left[-n \min_{Q:H(Q)\geq R} D(Q||P_X)\right]. \tag{108}$$

The protocol consists of only keeping sequences in $A^{(n)}$, for which by equation 107 we need at most

$$\frac{1}{n}\log\left[(n+1)^d 2^{nR}\right] = d\frac{\log(n+1)}{n} + R \xrightarrow{n\to\infty} R \text{ bits},$$

with the error decaying exponentially in n by equation 108.

11.4. Quantum source compression

A quantum source emits quantum state with certain probabilities. We restrict to pure state sources. Let $(p_x, |\psi_x\rangle)_{x\in[d]}$ be a quantum state ensemble, where $|\psi_x\rangle \in \mathcal{H}$ are pure states on a D-dimensional Hilbert space. The signal $|\psi_x\rangle$ is emitted with probability p_x. The iid assumption in this case is to assume the source emits sequences of states $|\psi_{x^n}\rangle = |\psi_{x_1}\rangle \otimes |\psi_{x_2}\rangle \otimes \cdots \otimes |\psi_{x_n}\rangle$ (which is analogous to the sequence $x^n = (x_1, \ldots, x_n) \in [d]^n$ in the classical case) with probability $\Pr(\psi_{x^n}) = \prod_{i=1}^n p_{x_i}$.
Let $\rho = \sum_{x\in[d]} p_x |\psi_x\rangle\langle\psi_x|$ be the *ensemble average density operator*. Then the average density operator after the source has emitted n signals is given by $\rho^{\otimes n}$. A source compression protocol consists of:

(1) an encoding or compression map

$$\varepsilon : \mathcal{L}(\mathcal{H}^{\otimes n}) \to \mathcal{L}(\tilde{\mathcal{H}}_n)$$

with $\dim \tilde{\mathcal{H}}_n < \dim \mathcal{H}^{\otimes n} = D^n$.

(2) a decoding operation $\mathcal{D} : \mathcal{L}(\tilde{\mathcal{H}}_n) \to \mathcal{L}(\mathcal{H}^{\otimes n})$.

Define the error as

$$\varepsilon_n = 1 - \underbrace{\sum_{x^n} p_{x^n} F(\psi_{x^n}, \mathcal{D} \circ \varepsilon(\psi_{x^n}))}_{\text{average fidelity}}.$$

If $\varepsilon_n \to 0$ for $n \to \infty$, we call

$$R = \lim_{n\to\infty} \frac{1}{n} \log\dim\tilde{\mathcal{H}}_n$$

an *achievable compression rate*, and the infimum R^* of all achievable rates is called the *optimal rate of compression*. In the classical case the optimal compression rate is given by the Shannon entropy of the source; in this case the equivalent notion is that of *von Neumann entropy* of the source.

Definition 39. The von Neumann entropy $S(\rho)$ of a density operator ρ with eigenvalues $\lambda = (\lambda_i)_{i=1,\ldots,D}$ is defined as

$$S(\rho) = -\sum_{i=1}^{D} \lambda_i \log \lambda_i.$$

If $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$ is a spectral decomposition, we can define the *matrix logarithm*

$$\log \rho = \sum_{i:\lambda_i > 0} \log \lambda_i |e_i\rangle\langle e_i|.$$

The von Neumann entropy of ρ is then given by $S(\rho) = \operatorname{tr} \rho \log \rho$.

Proposition 11.3. Properties of von Neumann entropy.

(1) $0 \leq S(\rho) \leq \log D$ where $D = \dim \mathcal{H}(\rho \in \mathcal{L}(\mathcal{H}))$. $S(\rho) = 0$ iff $\rho = |\psi\rangle\langle\psi|$ is pure. $S(\rho) = \log D$ if and only if $\rho = \frac{1}{D}\mathbb{1}_{\mathcal{H}}$ is completely mixed state.
(2) $S(\rho) = S(U\rho U^\dagger)$ for any unitary $U \in \mathrm{U}(\mathcal{H})$.
(3) $S(\lambda\rho_1 + (1-\lambda)\rho_2) \geq \lambda S(\rho_1) + (1-\lambda)S(\rho_2)$ for any $\lambda \in [0,1]$ and $\rho_1, \rho_2 \in \mathcal{L}(\mathcal{H})$.
(4) For any pure state $|\psi\rangle_{AB}$, we have $S(\psi_A) = S(\psi_B)$ using Schmidt decomposition.
(5) A pure state $|\psi\rangle_{AB}$ is entangled iff $S(\psi_A) > 0$.

How might we achieve quantum source compression at a rate equal tot the von Neumann entropy of the source? Schumacher in 1995 in [Sch95] showed that this is possible using a quantum version of typicality. Let $\rho = \sum_x p_x |x\rangle\langle x|$ be a spectral decomposition of a density operator $\rho \in \mathcal{H}$. Consider the state $\rho^{\otimes n}$ with spectral decomposition $\rho^{\otimes n} = \sum_{x^n} p_{x^n} |x^n\rangle\langle x^n|$, where $p_{x^n} := \prod_{i=1}^{n} p_{x_i}$ and $|x^n\rangle := \otimes_{i=1}^{n} |x_i\rangle$.

Definition 40. For $\varepsilon > 0$, the typical subspace $T_\varepsilon^{(n)}$ of a source $\rho = \sum_x p_x |x\rangle\langle x|$ is defined as

$$T_\varepsilon^{(n)} := \operatorname{span}\{|x^n\rangle : x^n \text{ is } \varepsilon\text{-typical }\} \subset \mathcal{H}^{\otimes n}.$$

The projector onto $T_\varepsilon^{(n)}$ is given by $\Pi_\varepsilon^{(n)} := \sum_{x^n \in T_\varepsilon^{(n)}} |x^n\rangle\langle x^n|$. Note that the symbol $T_\varepsilon^{(n)}$ is used to denote both the set of ε-typical sequences and the ε-typical subspace; it will be clear from the context which is meant.

Proposition 11.4. Properties of the typical subspace.

(1) For all $\delta > 0$, and n sufficiently large, $\operatorname{tr}\left(\Pi_\varepsilon^{(n)} \rho^{\otimes n}\right) \geq 1 - \varepsilon - \delta$.
(2) Let $S = S(\rho)$. Then for some constant $c > 0$,

$$\dim T_\varepsilon^{(n)} = \operatorname{tr} \Pi_\varepsilon^{(n)} \leq 2^{n(S + c\varepsilon)}.$$

(3) The operator $\tilde{\rho}_n := \Pi_\varepsilon^{(n)} \rho^{\otimes n} \Pi_\varepsilon^{(n)}$ is the "typical" part of $\rho^{\otimes n}$ and satisfies $\tilde{\rho}_n \cong 2^{-nS} \Pi_\varepsilon^{(n)}$. Furthermore, $\tilde{\rho}_n = \rho^{\otimes n}$ when n is large.

Schumacher's quantum source compression protocol.

(1) Perform the typical subspace measurement to project the source signals to the typical subspace.
(2) Using some enumeration of the typical sequences in $T_\varepsilon^{(n)}$, construct a map $U_f = \sum_{x^n \in T_\varepsilon^{(n)}} |f(x^n)\rangle_W \langle x^n|_{A^n}$, where $A^n = \mathcal{H}^{\otimes n}$ and W is the typical subspace. The subspace W has dimension at most $2^{n(S(\rho)+\varepsilon)}$. U_f is the inverse of an isometry, i.e., $U_f U_f^\dagger = \mathbb{1}_W$.
(3) Decoding: essentially apply U_f^{-1}.

This achieves compression at a rate

$$R = \lim_{n\to\infty} \frac{1}{n} \log \dim W = S(\rho),$$

with error $\varepsilon_n \to 0$ as $n \to \infty$.
One can also show that no asymptotically faithful compression protocol can achieve rates below the entropy $S(\rho)$, which proves that $S(\rho)$ is the optimal possible rate for quantum source compression.

11.5. Universal compression based on Schur-Weyl duality

Schumacher's compression protocol achieves optimal compression rate but is defined in terms of the spectral decomposition of the source state. In this section, we will see a compression protocol that depends only on $S(\rho)$. This protocol is based on symmetries and Schur-Weyl duality.
Quantum source compression has two symmetries:

(1) Permutation symmetry: $Q_\pi \rho^{\otimes n} Q_\pi^\dagger = \rho^{\otimes n}$ for all $\pi \in S_n$.
(2) Unitary symmetry: $S(\rho) = S(U\rho U^\dagger)$ for all $U \in \mathrm{U}(\mathcal{H})$ (in other words: entropy only depends on the spectrum of ρ).

Schur-Weyl duality gives the decomposition

$$\mathcal{H}^{\otimes n} \cong \bigotimes_{\lambda \vdash_D n} V_\lambda \otimes W_\lambda.$$

Let P_λ be the projector onto $V_\lambda \otimes W_\lambda$. For $\lambda \vdash_D n$ define $\overline{\lambda} = \frac{1}{n}\lambda$ (as in spectrum estimation). Now fix $R > S(\rho)$ and define

$$\Pi_R := \sum_{\lambda : H(\overline{\lambda}) \leq R} P_\lambda.$$

This is a quantum version of the universal classical source compression code of section 11.3. Using Π_R as the projector in a source compression protocol, we can show (see [Hay02]):

(1) With $\tilde{\mathcal{H}}_n := \Pi_R \mathcal{H}^{\otimes n} = \bigoplus_{\lambda : H(\overline{\lambda}) \leq R} V_\lambda \otimes W_\lambda$,

$$\dim \tilde{\mathcal{H}}_n = \operatorname{tr} \Pi_R \leq \mathrm{poly}(n) 2^{nR}.$$

The corresponding protocol thus has rate

$$\lim_{n \to \infty} \frac{1}{n} \log \dim \tilde{\mathcal{H}}_n \leq R.$$

(2) Exponential decay of decoding error:

$$\varepsilon_n \leq 2(n+D)^{4D} \exp\left(-n \min_{H(b) \geq R} D(b||\lambda) \right),$$

where λ are the eigenvalues of the source ρ and as before $D(b||\lambda)$ is the Kullback-Leibler distance between b and λ. Since $S(\rho) = H(\lambda) < R \leq H(b)$, we have $b \neq \lambda$ for all b in the above optimization, and hence $\min_{H(b) \geq R} D(b||\lambda) > 0$. Thus, $\varepsilon_n \to 0$ exponentially as $n \to \infty$. for any rate $R > S(\rho)$.

References

[Alc18] Judith Alcock-Zeilinger. *The Special Unitary Group, Birdtracks and Applications in QCD*. 2018.

[CDS07] Giulio Chiribella, Giacomo Mauro D'Ariano, and Dirk Schlingemann. "How Continuous Quantum Measurements in Finite Dimensions Are Actually Discrete". *Physical Review Letters* 98.19 (May 2007).

[Chr06] Matthias Christandl. "The structure of bipartite quantum states-insights from group theory and cryptography, PhD thesis, University of Cambridge" (2006).

[CK81] Imre Csiszar and János Korner. *Information theory: coding theorems for discrete memoryless systems, New York, New York: Academic Press*. 1981.

[CM06] Matthias Christandl and Graeme Mitchison. "The Spectra of Density Operators and the Kronecker Coefficients of the Symmetric Group". *Communications in Mathematical Physics* 261.3 (2006), pp. 789–797.

[Haa+16] Jeongwan Haah et al. "Sample-optimal tomography of quantum states". *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*. ACM, 2016, pp. 913–925.

[Har13] Aram W. Harrow. "The church of the symmetric subspace, arXiv" (2013). preprint.

[Hay02] Masahito Hayashi. "Exponents of quantum fixed-length pure state source coding". *Physical Review A* 66 (2002).

[Kna16] Anthony W. Knapp. *Representation Theory of Semisimple Groups: An Overview Based on Examples, Princeton, NJ: Princeton University Press*. 2016.

[KW01] M. Keyl and R. Werner. “Estimating the spectrum of a density operator”. *Physical Review A 64* (2001).

[Sca+05] Valerio Scarani et al. “Quantum cloning”. *Reviews of Modern Physics* 77.4 (Nov. 2005), pp. 1225–1256.

[Sch95] Benjamin Schumacher. “Quantum coding”. *Phys. Rev. A* 51 (4 1995), pp. 2738–2747.

[Ser77] Jean-Pierre Serre. *Linear Representations of Finite Groups, Graduate Texts in Mathematics, New York: Springer*. 1977.

[Sha48] Claude Shannon. “A Mathematical Theory of Communication”. *The Bell System Technical Journal* 27 (1948), pp. 379–423.

[Tel05] Constantin Teleman. *Representation Theory*. 2005.

[Wal18] Michael Walter. *Symmetry and Quantum Information, Lecture notes*. 2018.

[Wat18] John Watrous. *The Theory of Quantum Information, Cambridge University Press*. 2018.

[Wer98] Reinhard Werner. “Optimal Cloning of Pure States”. *Physical Review A* 58 (1998).

[Wil16] Mark M. Wilde. *Quantum information theory*. 2nd ed. Cambridge University Press, 2016.