- *The* ***double commutant theorem****: in situations where one has an isotypic decomposition like in Lemma 78, this states that one action being the* commutant *of the other in the sense of the above bullet point allows one to further deduce that the Schur functors in the isotypic decomposition are irreps.*

The first fact has a remarkably elementary proof, see https://math.univ-lyon1.fr/~aubrun/recherche/schur-weyl.pdf*. The second fact can be found in any standard representation theory text, e.g.* [EGH+09, Theorem 5.18.1].

At this juncture, what we really need is a user-friendly description of the *characters* of the representations $\mathbb{S}_\lambda(V_\lambda)$. As we will see, these are given by the **Schur polynomials** s_λ. This will follow from the following lemmas:

Lemma 81. *Let $|\vec{i}\rangle = |i_1\rangle \otimes \cdots \otimes |i_N\rangle$ be a tensor product of basis states, i.e. $i_1, \ldots, i_N \in [d]$. For every $j \in [d]$, suppose it appears ν_j times among $i_1, \ldots, i_N$. Then for any $\lambda \vdash [N]$ and any diagonal matrix $D = \mathrm{diag}(x_1, \ldots, x_d)$, we have*

$$D^{\otimes N} c_\lambda |\vec{i}\rangle = x_1^{\nu_1} \cdots x_d^{\nu_d} c_\lambda |\vec{i}\rangle \ .$$

PROOF. For any $\pi \in \mathcal{S}_N$, note that $D^{\otimes N} P(\pi) |\vec{i}\rangle = x_1^{\nu_1} \cdots x_d^{\nu_d}$. As c_λ is a linear combination of $P(\pi)$'s, the claim immediately follows. □

Definition 82. *Given $\lambda, \lambda' \vdash [N]$, we say that λ* ***majorizes*** *λ', denoted $\lambda' \prec \lambda$, if $\sum_{i \le j} \lambda'_i \le \sum_{i \le j} \lambda_i$ for all j.*

Lemma 83. *Let ν_j denote the number of times j appears among $i_1, \ldots, i_N$. Let $\nu^{\mathsf{sorted}} \vdash [N]$ denote the partition given by sorting the entries of ν in decreasing order.*

Then $c_\lambda |\vec{i}\rangle = 0$ if $\nu^{\mathsf{sorted}} \not\prec \lambda$. As a special case, this implies that $c_\lambda |\vec{i}\rangle = 0$ if λ has more than d rows.

PROOF. We can associate to $|\vec{i}\rangle$ a Young tableau T with shape λ by filling in the entries of $\vec{i}$ from left to right and top to bottom. If any column of T has a repeated entry, then $c_\lambda |\vec{i}\rangle = 0$.

The condition that $\nu^{\mathsf{sorted}} \not\prec \lambda$ implies that there is some j for which $\sum_{i \le j} \nu_i^{\mathsf{sorted}} > \sum_{i \le j} \lambda_i$, which implies that some column of T has a repeated entry by pigeonhole principle. □

Corollary 84. *Let $q_\lambda : \mathrm{GL}_d \to \mathrm{GL}(\mathbb{S}_\lambda V_\lambda)$ denote the GL_d-representation associated to the Schur functor $\mathbb{S}_\lambda V_\lambda$. For any $M \in \mathrm{GL}_d$ with eigenvalues $x_1, \ldots, x_d$, we have*

$$\mathrm{tr}(q_\lambda(M)) = \sum_{\text{SSYT } T} x^T \triangleq s_\lambda(\vec{x}) \,,$$

where the sum ranges over all semi-standard Young tableaux with shape λ over alphabet $[d]$, and x^T denotes the monomial $x_1^{\nu_1} \cdots x_d^{\nu_d}$ where ν_j is the number of occurrences of entry j in T. The polynomials s_λ are called the Schur polynomials.

PROOF. A fact we will need but will not prove is that $c_\lambda |\vec{i}\rangle$ for any $|\vec{i}\rangle$ which does not correspond to a semi-standard Young tableau can be expressed as a linear combination of $c_\lambda |\vec{j}\rangle$ for $|\vec{j}\rangle$'s which do correspond to semi-standard Young tableau. This is a nontrivial fact, a consequence of the so-called *Garnir relations*, whose proof is out of the scope of these notes. The upshot of this fact is that the collection of $\{c_\lambda |\vec{i}\rangle\}$ for all $|\vec{i}\rangle$ which do correspond to semi-standard Young tableaux spans $\mathbb{S}_\lambda V_\lambda$, in fact, forms a basis.

Next, note that any character χ of a representation μ of GL_d only depends on the eigenvalues of the input: if $M \in \mathrm{GL}_d$ has diagonalization $M = U^{-1}DU$, then $\chi(M) = \chi(U^{-1}DU) = \mathrm{tr}(\mu(U)^{-1}\mu(D)\mu(U)) = \mathrm{tr}(\mu(D)) = \chi(D)$.

So we may assume that M in the corollary is diagonal, in which case Lemmas 81 and 83 imply that

$$\mathrm{tr}(q_\lambda(M)) = \sum_{\nu:\nu^{\mathsf{sorted}} \preceq \lambda} K_{\lambda\nu} x_1^{\nu_1} \cdots x_d^{\nu_d}\,,$$

where ν ranges over all *ordered* partitions of $[N]$ and $K_{\lambda\nu}$ denotes the number of semi-standard Young tableaux with shape λ over alphabet $[d]$ such that each $j \in [d]$ occurs ν_j times.[2] The result then follows from the fact that $K_{\lambda\nu} = 0$ if $\nu^{\mathsf{sorted}} \not\preceq \lambda$, which we do not prove here. □

2.4. Schur Polynomial Facts

We will need two simple facts about the Schur polynomials.

Lemma 85. *For any partition $\lambda \vdash [N]$ with at most d rows,*

$$\dim(\mathbb{S}_\lambda V_\lambda) = s_\lambda(1^d) \leq N^{O(d^2)}\,.$$

Proof. The first equality follows from the fact that $s_\lambda(1^d) = \mathrm{tr}(q_\lambda(\mathrm{Id}))$ by Corollary 84, together with the fact that the character of a representation evaluated at the identity is the dimension of the representation.

The inequality is typically proved by invoking the identity

$$s_\lambda(1^d) = \prod_{1\leq i<j\leq d} \frac{\lambda_i - \lambda_j + j - i}{j-i}$$

together with the fact that $0 \leq \lambda_i - \lambda_j \leq N$, but the proof for this identity is nontrivial. Instead, we can prove the claimed bound by the following simple combinatorial argument. Recall that $s_\lambda(1^d)$ counts the number of SSYT of shape λ with entries from $[d]$. Given any such SSYT T, we can consider the sequence of SSYT's $(T^{\leq j})$ given by removing all blocks with an entry larger than j. Note that the shapes $\lambda^{\leq j}$ of these SSYT's can be used to uniquely recover the original T, because the difference between $T^{\leq j}$ and $T^{\leq j-1}$ specifies the locations of the j entries withint T. The number of such sequences of shapes $(\lambda^{\leq j})$ can be naively bounded by $N^{O(d^2)}$. □

Lemma 86. *Given two d-dimensional density matrices σ, ρ, for any partition $\lambda \vdash [N]$ with at most d rows,*

$$\mathrm{tr}(q_\lambda(\rho\sigma)) \leq \dim(\mathbb{S}_\lambda V_\lambda) \cdot e^{-2NH(\overline{\lambda})} F(\rho,\sigma)^{2N}\,.$$

Here $\overline{\lambda}$ denotes the distribution over $[d]$ with probability mass function given by $(\lambda_i/N)_{i\in[d]}$, and $H(\cdot)$ denotes Shannon entropy.

Proof. Let X be a psd matrix with eigenvalues $x_1 \geq \cdots \geq x_d$ sorted in non-increasing order, and define $\overline{x}_i = x_i/\mathrm{tr}(X)$.

By Fact 87 and the fact that there are $\dim(\mathbb{S}_\lambda V_\lambda)$ monomials in the definition of s_λ, we can bound

$$\mathrm{tr}(q_\lambda(X^2)) = s_\lambda(x_1^2, \ldots, x_d^2) \leq \dim(\mathbb{S}_\lambda V_\lambda) \cdot x_1^{2\lambda_1} \cdots x_d^{2\lambda_d}\,.$$

[2] Note that if ν and ν' are equal up to permutation of entries, $K_{\lambda\nu} = K_{\lambda\nu'}$, though this is not *a priori* clear.

Note that

$$\begin{aligned} x_1^{2\lambda_1} \cdots x_d^{2\lambda_d} &= \overline{x}_1^{2\lambda_1} \cdots \overline{x}_d^{2\lambda_d} \cdot \operatorname{tr}(X)^{2N} \\ &= \exp(2N \mathbb{E}_{i \sim \overline{\lambda}} \log \overline{x}_i) \cdot \operatorname{tr}(X)^{2N} \\ &\leq \exp(-2NH(\overline{\lambda})) \cdot \operatorname{tr}(X)^{2N} , \end{aligned}$$

where in the last step we used Gibbs' inequality.

To conclude the proof, take $X = \sqrt{\sigma^{1/2}\rho\sigma^{1/2}}$ so that $\operatorname{tr}(X) = F(\rho, \sigma)$. As the Schur polynomials are characters, $\operatorname{tr}(q_\lambda(\rho\sigma)) = \operatorname{tr}(q_\lambda(\sigma^{1/2}\rho\sigma^{1/2})) = \operatorname{tr}(q_\lambda(X^2))$, concluding the proof. $\square$

The above argument uses the following elementary fact:

Fact 87. *Given any partitions $\nu \preceq \lambda$ with at most d rows, and any nonnegative reals $x_1 \geq \cdots \geq x_d$,*

$$x_1^{\nu_1} \cdots x_d^{\nu_d} \leq x_1^{\lambda_1} \cdots x_d^{\lambda_d} .$$

Finally, we need the following simple lower bound on the Schur polynomials:

Lemma 88. *For any $\lambda \vdash [N]$ $s_\lambda(\overline{\lambda}) \geq e^{-NH(\overline{\lambda})}$.*

Proof. Every Schur polynomial is a sum of monomials and thus lower bounded by any single monomial. Consider the SSYT T with shape λ that has λ_j j's in the j-th row. Then

$$s_\lambda(\overline{\lambda}) = \overline{\lambda}^T = \overline{\lambda}_1^{N\overline{\lambda}_1} \cdots \overline{\lambda}_d^{N\overline{\lambda}_d} = e^{-NH(\overline{\lambda})}$$

as claimed. $\square$

3. Weak Schur Sampling

As discussed at the beginning of the previous section, the upshot of the isotypic decomposition in Lemma 78 is that there is a certain unitary U_{schur} over $\mathcal{H}$ that block-diagonalizes all states of the form $\rho^{\otimes N}$. More generally, for any $\pi \in \mathcal{S}_N$ and $M \in \mathrm{GL}_d$, we have

$$U_{\mathsf{schur}} P(\pi) Q(\rho) U_{\mathsf{schur}}^\dagger = \bigoplus_{\lambda \vdash [N]} p_\lambda(\pi) \otimes q_\lambda(M) ,$$

where p_λ is the $\mathcal{S}_N$-representation of π acting on the Specht module $V_\lambda \cong \mathbb{C}[\mathcal{S}_N] \cdot c_\lambda$, and q_λ is the GL_d-representation whose character is given by the Schur polynomial s_λ. In particular, when $\pi = \mathrm{Id}$ and $M = \rho$, then we obtain the block decomposition

$$U_{\mathsf{schur}} \rho^{\otimes N} U_{\mathsf{schur}}^\dagger = \bigotimes_{\lambda \vdash [N]} \mathrm{Id}_{V_\lambda} \otimes q_\lambda(\rho) .$$

The reason this block decomposition is helpful is that we have a good understanding of the blocks $q_\lambda(\rho)$ and how they behave when the copies of ρ get rotated.

The first step of our learning algorithm is thus to rotate the input $\rho^{\otimes N}$ into the basis prescribed by U_{schur} and perform a projective measurement onto one of these blocks, a procedure called *weak Schur sampling*.

Definition 89 (Weak Schur Sampling)**.** *Let ρ be a density matrix with eigenvalues $\overline{\lambda}_1^* \geq \cdots \geq \overline{\lambda}_d^* \geq 0$.*

Let Π_λ denote the projector to the isotypic component $V_\lambda \otimes \mathbb{S}_\lambda V_\lambda$. Weak Schur sampling *is the following procedure:*

(1) Perform a projective measurement $\{\Pi_\lambda\}_\lambda$ *on* $\rho^{\otimes N}$ *to obtain a state in* $V_\lambda \otimes \mathbb{S}_\lambda V_\lambda$ *with probability*

$$\mathrm{tr}(\mathrm{Id}_{V_\lambda} \otimes q_\lambda(\rho)) = \dim(V_\lambda) \cdot s_\lambda(\overline{\lambda}^*)\,.$$

The distribution over partitions $\lambda \vdash [N]$ *with this probability mass function is called the* ***Schur-Weyl distribution****, denoted* $\mathrm{SW}^N(\overline{\lambda}^*)$.

(2) Trace out the V_λ *register, resulting in the state with unnormalized density matrix* $q_\lambda(\rho)$*. Note that the trace of this is* $s_\lambda(\overline{\lambda}^*)$ *by Corollary 84, so the normalized density matrix is*

$$\tilde{\rho} \triangleq q_\lambda(\rho)/s_\lambda(\overline{\lambda}^*)\,.$$

Intuitively, the partition λ obtained by weak Schur sampling gives us a rough estimate $\overline{\lambda} = (\lambda_1/N, \ldots, \lambda_N/N)$ of the spectrum $\overline{\lambda}^*$ of ρ. This should be thought of as the quantum analogue of the classical algorithm for learning discrete distributions: given N samples from a distribution over $[d]$ which places mass $\overline{\lambda}^*_i$ on element i, the optimal estimator for $\overline{\lambda}^*$ is to output the **empirical histogram** $\overline{\lambda} = (\lambda_1/N, \ldots, \lambda_N/N)$, where λ_i is the number of samples in the dataset that are equal to i.

When learning quantum states however, there is a crucial missing piece even after we have estimated the spectrum of ρ: estimating the *eigenvectors* of ρ. This is where we will leverage the Schur polynomial estimates from the previous section.

4. Pretty Good Measurement

With the spectrum estimate $\overline{\lambda}$ and the post-measurement state $\tilde{\rho} = q_\lambda(\rho)/s_\lambda(\overline{\lambda}^*)$ in hand, a natural approach for learning the eigenbasis for ρ would be to sample a random unitary U and "measure $\tilde{\rho}$" with the operator $U\mathrm{diag}(\overline{\lambda})U^\dagger$ in each copy in our dataset. Of course, there is a type mismatch: our dataset is no longer an element of $\mathcal{H}$, so measuring $\tilde{\rho}$ with $Q(U\mathrm{diag}(\overline{\lambda})U^\dagger)$ doesn't quite make sense. But because q_λ is an irrep inside the representation $Q(\cdot)$, we know how $Q(U\mathrm{diag}(\overline{\lambda})U^\dagger)$ acts on $\tilde{\rho}$, namely via $q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)$. So by "measuring $\tilde{\rho}$ with $U\mathrm{diag}(\overline{\lambda})U^\dagger$," we really mean measuring with the operator $q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)$. Up to a normalizing constant, this is now entirely well-defined and provides the desired rotationally *equi*variant POVM needed for the second stage of our algorithm.

Lemma 90. *The POVM with elements*

$$\frac{\dim(\mathbb{S}_\lambda V_\lambda)}{s_\lambda(\overline{\lambda})} \cdot q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)\,\mathrm{d}U \tag{24}$$

is a valid POVM.

Proof. Note that this ensemble is invariant under conjugation by any unitary: for any $W \in U_d$ we have $q_\lambda(W) \cdot q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)q_\lambda(W^\dagger) = q_\lambda(WU\mathrm{diag}(\overline{\lambda})U^\dagger W^\dagger)$, and the Haar measure over U_d is invariant under left-multiplication by W by definition. So by irreducibility of q_λ, to show that the POVM elements integrate to $\mathrm{Id}_{\mathbb{S}_\lambda V_\lambda}$, it suffices to verify that their trace integrates to $\dim(\mathbb{S}_\lambda V_\lambda)$. This follows because

$$\mathrm{tr}(q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)) = \mathrm{tr}(q_\lambda(\mathrm{diag}(\overline{\lambda}))) = s_\lambda(\overline{\lambda})\,. \qquad \square$$

The pseudocode for our final learning algorithm is as follows:

Algorithm 1: OptimalTomography(ρ)

Input: N copies of unknown d-dimensional state ρ
Output: Estimate $\hat{\rho}$
Perform weak Schur sampling on $\rho^{\otimes N}$ (Definition 89) to obtain $\lambda \vdash [N]$ and post-measurement state $\tilde{\rho}$.
Measure $\tilde{\rho}$ with the POVM in Eq. (24) to obtain U.
return $\hat{\rho} = U \mathrm{diag}(\overline{\lambda}) U^\dagger$

With all of the machinery from Section 2, the proof that this works ends up being remarkably simple. The following shows that the further $\hat{\rho} = U\mathrm{diag}(\overline{\lambda})U^\dagger$ is from ρ, the less likely it is to output $\hat{\rho}$.

Lemma 91. *The infinitesimal probability of obtaining λ in the first step of* OptimalTomography *and U in the second step is at most $N^{O(d^2)} \cdot F(U\mathrm{diag}(\overline{\lambda})U^\dagger, \rho)^{2N}\, \mathrm{d}U$.*

Proof. Let us first compute the infinitesimal probability of observing U in the second stage of the algorithm, conditioned on measuring the post-measurement state $\tilde{\rho}$:

$$\frac{\dim(\mathbb{S}_\lambda V_\lambda)}{s_\lambda(\overline{\lambda}) s_\lambda(\overline{\lambda}^*)} \cdot \mathrm{tr}(q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger) \cdot q_\lambda(\rho))\, \mathrm{d}U = \frac{\dim(\mathbb{S}_\lambda V_\lambda)}{s_\lambda(\overline{\lambda}) s_\lambda(\overline{\lambda}^*)} \cdot \mathrm{tr}(q_\lambda(\rho U \mathrm{diag}(\overline{\lambda})U^\dagger))\, \mathrm{d}U\,,$$

where we used the fact that q_λ is a GL_d-representation. Recall that the probability of getting λ and $\tilde{\rho}$ from weak Schur sampling is $\dim(V_\lambda) \cdot s_\lambda(\overline{\lambda}^*)$, so the infinitesimal probability that the algorithm outputs a particular $\hat{\rho} = U\mathrm{diag}(\overline{\lambda})U^\dagger$ is

$$\dim(\mathbb{S}_\lambda V_\lambda) \dim(V_\lambda) \cdot \frac{\mathrm{tr}(q_\lambda(\rho U\mathrm{diag}(\overline{\lambda})U^\dagger))}{s_\lambda(\overline{\lambda})}\, \mathrm{d}U\,.$$

By Lemmas 77, 85, 86, and 88, this is at most

$$N^{O(d^2)} F(U\mathrm{diag}(\overline{\lambda})U^\dagger, \rho)^{2N}\, \mathrm{d}U$$

as claimed. Note the fortuitous cancellation of the entropy terms. □

We are now ready to prove the main result.

Theorem 92. *For any $\epsilon > 0$, there is $N = \tilde{O}((d^2 + \log 1/\delta)/\epsilon)$ such that given N copies of ρ,* OptimalTomography*(ρ) outputs an estimate $\hat{\rho}$ satisfying $F(\rho, \hat{\rho}) \geq 1 - \epsilon$ with probability at least $1 - \delta$.*

Proof. There are $\leq N^{O(d)}$ partitions $\lambda \vdash [N]$ with at most d rows, and $\int \mathrm{d}U = 1$, so the total probability contributed by (λ, U) for which $F(U\mathrm{diag}(\overline{\lambda})U^\dagger, \rho) < 1 - \epsilon$ is at most $N^{O(d^2)}(1-\epsilon)^{2N}$. Provided $N = \Omega((d^2 \log N + \log 1/\delta)/\epsilon)$, this is upper bounded by δ as desired. □

The bound is still off by a $\log(d/\epsilon)$ factor, and it is still open whether this can be tightened to match the best known lower bound of $\Omega((d^2 + \log 1/\delta)/\epsilon)$.