Schur-Weyl duality, quantum data compression, tomography

Lecture 7 *Michael Walter, Stanford University*

These lecture notes are not proof-read and are offered for your convenience only. They include additional detail and references to supplementary reading material. I would be grateful if you email me about any mistakes and typos that you find.

Today, we will summarize the "Schur-Weyl toolbox" that we developed in lecture 6 to solve the spectrum estimation problem. We will then apply it to the task of compressing a quantum information source.

7.1 The Schur-Weyl toolbox

Let us recapitulate the machinary that we developed to solve the spectrum estimation problem. Just like any representation of SU(2), the Hilbert space of n qubits can be decomposed in the form

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_j \otimes \mathbb{C}^{m(n,j)}.$$

Last time, we discussed that the action of SU(2) could be extended first to SL(2) and then to arbitrary operators on $\mathbb{C}^2$: In eq. (6.3), we found that

$$X^{\otimes n} \cong \bigoplus_j T_X^{(n,j)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}},$$

where

$$T_X^{(n,j)} = (\det X)^{n/2}\, T^{(j)}_{X/\sqrt{\det X}}$$

is a polynomial in the matrix elements of X and hence makes sense for arbitrary X. You can verify this, e.g., by using the symmetric subspace model of the spin-j representation. In particular, this formula applies to unitary matrices U. It follows that the operators $T_U^{(n,j)}$ define a representation of the unitary group $U(2)$, which we will denote by $V_{n,j}$. Here, j tells us the spin of the representation when restricted to matrices in SU(2), and n reminds us of the way that multiples $\alpha \mathbb{1}_{\mathbb{C}^2}$ of the identity matrix act by α^n. Since every unitary can be written as αU with $\alpha \neq 0$ and $U \in \mathrm{SU}(2)$, this information specifies the representation completely. It is clear that the representations $V_{n,j}$ are irreducible, since they are even irreducible for the subgroup SU(2).

We can also consider $(\mathbb{C}^2)^{\otimes n}$ as a representation of the symmetric group S_n. Since $[R_\pi, U^{\otimes n}] = 0$, Schur's lemma (lemma 3.2) implies that

$$R_\pi \cong \bigoplus_j \mathbb{1}_{V_{n,j}} \otimes R_\pi^{(n,j)}$$

for some operators $R_\pi^{(n,j)}$ on $\mathbb{C}^{m(n,j)}$. So far, the Hilbert spaces $\mathbb{C}^{m(n,j)}$ were simply vectors spaces – but now we see that the operators $R_\pi^{(n,j)}$ turn them into representations of S_n. We will denote these representations by $W_{n,j}$. The representations $W_{n,j}$ are irreducible and pairwise inequivalent. You will verify this and the following statements in problem 3.5.

Thus, we have the following decomposition of the Hilbert space of n qubits:

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus V_{n,j} \otimes W_{n,j} \tag{7.1}$$

which holds as a representation of both $U(2)$ and S_n (equivalently, of the product group $U(2) \times S_n$). The spaces $\{V_{n,j}\}$ and $\{W_{n,j}\}$ are pairwise inequivalent, irreducible representations of $U(2)$ and of S_n, respectively. Equation (7.1) shows that they are "paired up" perfectly in the n-qubit Hilbert space. This result is known as *Schur-Weyl duality*, and it has a number of important consequences.

For example, any operator that commutes with all $U^{\otimes n}$ is necessarily a linear combination of the operators R_π. Dually, any operator that commutes with all R_π is necessarily a linear combination of operators of the form $X^{\otimes n}$ (even $U^{\otimes n}$). Mathematically, we say that the two representations span each other's *commutants*. Schur-Weyl duality also implies that the projectors

$$P_j \cong \bigoplus_{j'} \delta_{j,j'} \mathbb{1}$$

not only have both symmetries of the spectrum estimation problem (i.e., that they commute with both the $U^{\otimes n}$ and the R_π), but that they are in fact the most fine-grained projective measurement with this property.

Table 1 assembles all important facts and formulas about the representation theory of the n-qubit Hilbert space that we obtained past week (the "Schur-Weyl toolbox"). It contains one formula, eq. (7.5), which is proved just like eq. (6.8). We will use it to solve the quantum state tomography problem in section 7.3 below.

Remark 7.1. *So far, we have simply argued on abstract grounds that the Hilbert space of n qubits can be decomposed in the form (7.1). Here, the notation $\cong$ means that there exists a unitary intertwiner from the left-hand side to the right-hand side. But if we want to implement, e.g., spectrum estimation in practice, we need to know what this unitary operator looks like. In other words, we need to find a unitary operator that implements the transformation from the product basis*

$$|x_1, \dots, x_n\rangle = |x_1\rangle \otimes \dots \otimes |x_n\rangle$$

to the Schur-Weyl basis

$$|j, m, k\rangle$$

where $j \in \{\dots, \frac{n}{2} - 1, \frac{n}{2}\}$, $m \in \{-j, \dots, j\}$, $k \in \{1, \dots, m(n,j)\}$. Note that the right-hand side is not *a tensor product of three spaces, because the allowed values for m and k depend on j. However, we can embed it into a larger space where $|j,m,k\rangle = |j\rangle \otimes |m\rangle \otimes |k\rangle$ is a product basis vector. In lecture 10 we will learn how to implement this transformation – called the* quantum Schur transform *– by a quantum circuit.*

Beyond qubits

How does the preceding generalize beyond qubits? This is best explained by making a simple coordinate change and instead of by (n, j) parametrizing all representations by

$$\lambda = (\lambda_1, \lambda_2) = \left(\frac{n}{2} + j, \frac{n}{2} - j\right) \in \mathbb{Z}^2.$$

We can identify λ with a so-called *Young diagram* with two rows, where we place λ_1 boxes in the first and λ_2 boxes in the second row. E.g.,

$$\lambda = (7, 3) = \ydiagram{7,3}$$

We always demand that $\lambda_1 \geq \lambda_2$, corresponding to $j \geq 0$. Note that the total number of boxes is $\lambda_1 + \lambda_2 = n$, while $2j = \lambda_1 - \lambda_2$ is the difference of row lengths.

If we write $V_\lambda := V_{n,j}$ and $W_\lambda := W_{n,j}$, then the Schur-Weyl duality (7.1) becomes

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_\lambda V_\lambda \otimes W_\lambda, \tag{7.2}$$

where we sum over all Young diagrams with n boxes and at most two rows.

Remark 7.2. *In example 3.1, we already met the irreducible representations of S_3 and labeled them by Young diagrams. The representations $W_{\ydiagram{3}}$ and $W_{\ydiagram{2,1}}$ that occur in $(\mathbb{C}^2)^{\otimes 3}$ are precisely the ones that we already met in example 3.1. You will verify this in problem 4.1.*

On the other hand, because the antisymmetric subspace $\bigwedge^3 \mathbb{C}^2 = \{0\}$ is zero-dimensional, the sign representation $W_{\ydiagram{1,1,1}}$ does not appear at all.

The notation λ is quite suggestive. Indeed, let us define the *normalization* of a Young diagram λ by $\bar{\lambda} = \lambda/n = (\lambda_1/n, \lambda_2/n)$, where $n = \lambda_1 + \lambda_2$. This is a probability distribution, and

$$\bar{\lambda}_1 = \frac{1}{2} + \frac{j}{n} = \hat{p}, \quad \bar{\lambda}_2 = \frac{1}{2} - \frac{j}{n} = 1 - \hat{p}.$$

Thus, spectrum estimation can be rephrased as follows: When we measure $\{P_\lambda\}$ on $\rho^{\otimes n}$ and the outcome is λ, then $\bar{\lambda}$ is a good estimate for the spectrum of ρ.

The key point now is the following: eq. (7.2) generalizes quite directly from qubits to arbitrary d. This is because the irreducible representations of $U(d)$ are labeled by Young diagrams with (at most) d rows, while the irreps of S_n are labeled by Young diagrams with n boxes. See, e.g., Harrow (2005), Christandl (2006), Walter (2014) for further detail.

7.2 Quantum data compression

We will now discuss quantum data compression in more precise terms (Schumacher, 1995). We consider a *quantum information source* described by an ensemble $\{p_x, |\psi_x\rangle\}$ of qubit pure states. It emits sequences

$$|\psi(\vec{x})\rangle = |\psi_{x_1}\rangle \otimes \ldots \otimes |\psi_{x_n}\rangle \in (\mathbb{C}^2)^{\otimes n}$$

with probabilities

$$p(\vec{x}) = p_{x_1} \ldots p_{x_n}.$$

The task of *quantum data compression* is to design an compressor that encodes a sequence $|\psi(\vec{x})\rangle \in (\mathbb{C}^2)^{\otimes n}$ into some state of Rn qubits and a corresponding decompressor – R is called the *compression rate* at *block length n*. Unlike the state of a coin, we cannot in general hope to precisely recover the original state. Instead, the decompressor should produce a state $|\tilde{\psi}(\vec{x})\rangle$ that has high overlap with the original state (say, on average):

$$\sum_{\vec{x}} p(\vec{x})\, E\left[|\langle\psi(\vec{x})|\tilde{\psi}(\vec{x})\rangle|^2\right] \approx 1. \tag{7.6}$$

Schur-Weyl duality:

$$
\begin{aligned}
(\mathbb{C}^2)^{\otimes n} &\cong \bigoplus_{j=\dots,\frac{n}{2}-1,\frac{n}{2}} V_{n,j} \otimes W_{n,j}, \\
X^{\otimes n} &\cong \bigoplus_j T_X^{(n,j)} \otimes \mathbb{1}_{W_{n,j}}, \quad \text{where} \quad T_X^{(n,j)} := (\det X)^{n/2}\, T^{(j)}_{X/\sqrt{\det X}}, \\
R_\pi &\cong \bigoplus_j \mathbb{1}_{V_{n,j}} \otimes R_\pi^{(n,j)}.
\end{aligned}
$$

$V_{n,j}$ and $W_{n,j}$ are pairwise inequivalent, irreducible representations of $U(2)$ and S_n, respectively.

Dimensions:

$$
\begin{aligned}
\dim V_{n,j} &= 2j + 1 \leq n + 1, \\
\dim W_{n,j} &= m(n,j) \leq 2^{nh(\hat{p})}, \quad \text{where} \quad \hat{p} = \frac{1}{2} + \frac{j}{n}.
\end{aligned} \tag{7.3}
$$

Estimates:

$$
2^{-n\left[h(\hat{p})+\delta(\hat{p}\|p)\right]} \leq \operatorname{tr}\left[T_\rho^{(n,j)}\right] \leq (2j+1)2^{-n\left[h(\hat{p})+\delta(\hat{p}\|p)\right]} \quad \text{where } \rho \text{ has eigenvalues } \{p, 1-p\}, \tag{7.4}
$$

More generally, if $X \geq 0$ and $k > 0$:

$$
\operatorname{tr}\left[T_{X^k}^{(n,j)}\right] \leq (2j+1)2^{-nk\left[h(\hat{p})+\delta(\hat{p}\|x)\right]} (\operatorname{tr} X)^{kn}, \quad \text{where } \frac{X}{\operatorname{tr} X} \text{ has eigenvalues } \{x, 1-x\}. \tag{7.5}
$$

Spectrum estimation:

$$
\begin{aligned}
P_j &\cong \bigoplus_{j'} \delta_{j,j'} \mathbb{1}_{V_{n,j}} \otimes \mathbb{1}_{W_{n,j}}, \\
\rho^{\otimes n} &\cong \bigoplus_j T_\rho^{(n,j)} \otimes \mathbb{1}_{V_{n,j}} =: \bigoplus_j p_j\, \rho_{V_{n,j}} \otimes \tau_{W_{n,j}},
\end{aligned}
$$

and so

$$
\begin{aligned}
p_j = \operatorname{tr}\left[P_j \rho^{\otimes n}\right] &\leq (n+1)^2 2^{-n\delta(\hat{p}\|p)} \leq (n+1)^2 2^{-n\frac{2}{\ln 2}(\hat{p}-p)^2} \\
\operatorname{tr}\left[\widetilde{P}_n \rho^{\otimes n}\right] &\geq 1 - (n+1)^2 2^{-n\frac{2}{\ln 2}\varepsilon^2}
\end{aligned}
$$

where $\widetilde{P}_n := \sum_{j:|\hat{p}-p|<\varepsilon} P_j$ is the projector onto the "ε-spectrum typical subspace" of ρ.

Table 1: The Schur-Weyl toolbox for i.i.d. quantum information theory (in the case of qubits).

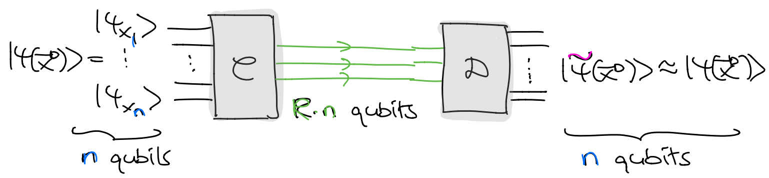

Figure 10: Illustration of the compression of a quantum information source.

The average value $E[\dots]$ refers to the fact that the decompressed state $|\tilde{\psi}(\vec{x})\rangle$ for a given $|\psi(\vec{x})\rangle$ is typically random. How should we go about solving this problem?

At the end of last lecture, we constructed, for every $p \in [\frac{1}{2}, 1]$ and $\varepsilon > 0$, projectors $\tilde{P}_n = \sum_{j:|p-\hat{p}|<\varepsilon} P_j$ onto a subspace $\tilde{\mathcal{H}}_n$ of $(\mathbb{C}^2)^{\otimes n}$ such that

$$\dim \tilde{\mathcal{H}}_n \le (n+1)^2 2^{nh(p)+\delta(\varepsilon)},$$

and

$$\operatorname{tr}\left[\tilde{P}_n \rho^{\otimes n}\right] \to 1 \tag{7.7}$$

for all density operators ρ with eigenvalues $\{p, 1-p\}$ (cf. table 1).

What is the density operator ρ that we should care about? Every ensemble gives rise to a density operator $\rho = \sum_x p_x |\psi_x\rangle \langle\psi_x|$, describing the *average state* emitted by the qubit source (we discussed this in lecture 3).

Remark. *The states $|\psi_x\rangle$ emitted by the source do* not *have to be orthogonal. Thus, the eigenvalues $\{p, 1-p\}$ of ρ used to construct $\tilde{P}_n$ are* not *in general the same as the probabilities p_x of the ensemble. E.g., in problem 1.3 you computed that $\frac{1}{2}(|1\rangle\langle 1| + |-\rangle\langle -|)$ has eigenvalues $\frac{1}{2} \pm \frac{1}{2\sqrt{2}}$.*

This suggests the following two-step *quantum data compression protocol* that is completely analogous to the way by which we compressed sequences of coin flips in section 5.1:

- Alice measures the observable $\tilde{P}_n$ (i.e., she performs the projective measurement $\{\tilde{P}_n, \mathbb{1} - \tilde{P}_n\}$).
- If the outcome is 1, then the post-measurement state

$$\frac{\tilde{P}_n |\psi(\vec{x})\rangle}{\|\tilde{P}_n |\psi(\vec{x})\rangle\|} \in \tilde{\mathcal{H}}_n$$

  lives in the subspace $\tilde{\mathcal{H}}$ only. Thus, Alice can send this state over to Bob by transmitting roughly $n(h(p) + \delta)$ qubits.

- If the outcome is 0, she simply sends over some fixed state. (Alternatively, she might signal failure – as in our coin flip protocol.)

Bob then uses the sent-over state in $\tilde{\mathcal{H}} \subseteq (\mathbb{C}^2)^{\otimes n}$ as the decompressed state. For large n, this protocol achieves a quantum compression rate of roughly $R = h(p) + \delta$.

Remark. *As discussed in class, in order to be able to "send over" the post-measurement state we first need to identify the subspace $\tilde{\mathcal{H}}$ with $N \approx n(h(p) + \delta)$ many qubits. For example, Alice* *could first apply a unitary U that maps the subspace $\tilde{\mathcal{H}}$ into the subspace of states of the form $|\phi\rangle_{A_1 \dots A_N} \otimes |0\rangle_{A_{N+1}} \otimes \dots \otimes |0\rangle_{A_n}$. Alice would then send over the first N of her qubits. Upon receiving those, Bob would apply $U^\dagger$ to obtain the decompressed sate. Mathematically, this is not very interesting, but physically this is quite important because we usually do not get to choose our physical qubits!*

Let us analyze the average fidelity (7.6) achieved by our compression protocol. If the input state is $|\psi(\vec{x})\rangle$ then according to the Born rule the measurement of the observable $\tilde{P}_n$ yields outcome 1 with probability

$$q(\vec{x}) := \langle \psi(\vec{x}) | \tilde{P}_n | \psi(\vec{x}) \rangle .$$

As already mentioned above, the post-measurement state in this case is

$$\frac{\tilde{P}_n |\psi(\vec{x})\rangle}{\|\tilde{P}_n |\psi(\vec{x})\rangle\|} = \frac{\tilde{P}_n |\psi(\vec{x})\rangle}{\sqrt{q(\vec{x})}}$$

and so this is the state $|\tilde{\psi}(\vec{x})\rangle$ that Bob obtains at his end. Thus, eq. (7.6) can be bounded as follows:

$$\sum_{\vec{x}} p(\vec{x})\, E\left[|\langle \psi(\vec{x}) | \tilde{\psi}(\vec{x}) \rangle|^2\right] \geq \sum_{\vec{x}} p(\vec{x})\, q(\vec{x})\, |\langle \psi(\vec{x})| \frac{\tilde{P}_n |\psi(\vec{x})\rangle}{\sqrt{q(\vec{x})}} \rangle|^2 = \sum_{\vec{x}} p(\vec{x})\, |\langle \psi(\vec{x}) | \tilde{P}_n | \psi(\vec{x}) \rangle|^2$$
$$= \sum_{\vec{x}} p(\vec{x})\, q(\vec{x})^2 \geq \left(\sum_{\vec{x}} p(\vec{x})\, q(\vec{x}) \right)^2$$

The first inequality is because we lower bound the overlap in the case that the outcome is 0; the second inequality is Jensen's inequality that we already used previously in class. But now note that

$$\sum_{\vec{x}} p(\vec{x})\, q(\vec{x}) = \sum_{\vec{x}} p(\vec{x}) \operatorname{tr}\left[|\psi(\vec{x})\rangle \langle \psi(\vec{x})| \, \tilde{P}_n \right] = \operatorname{tr}\left[\left(\sum_{\vec{x}} p(\vec{x}) \; |\psi(\vec{x})\rangle \langle \psi(\vec{x})| \right) \tilde{P}_n \right] = \operatorname{tr} \rho^{\otimes n} \tilde{P}_n \to 1$$

by eq. (7.7). Thus, our compression protocol will successfully compress a quantum information source with associated density operator ρ at rate $h(p) + \delta$. We can make $\delta > 0$ arbitrarily small by choosing $\varepsilon > 0$ smaller and smaller (note, however, that this requires the block length n to increase). This compression rate turns out to be optimal, as we will find in lecture 9.

This motivates us to define the *von Neumann entropy* of a density operator ρ as

$$S(\rho) = -\operatorname{tr} \rho \log \rho.$$

For qubits, $S(\rho) = h(p)$, as you can verify by expanding the trace in the eigenbasis of ρ. Thus, the von Neumann entropy that you might already know from your quantum physics research has a well-defined *operational interpretation*: It is the optimal compression rate of *any* quantum information source with associated density operator ρ. This is in complete analogy to one of the many roles played by the Shannon entropy in classical information theory. Next time, we will discuss a number of other meanings of the von Neumann entropy related to entanglement.

Remark. *This emphasizes a fundamental idea in information theory: We often seek to find characterizations of entropic quantities as optimal rates for information processing tasks. In the asymptotic limit of $n \to \infty$, the von Neumann entropy plays a rather universal role. However, at finite block lengths $n < \infty$, there is not just one entropy but a whole zoo of entropic quantities that information theorists are interested in, each targeted at different tasks (Faist, 2013).*

An interesting fact about our compression protocol is that the projectors $\tilde{P}_n$ depended only on the eigenvalues p and $1-p$, not on the eigenbasis of the density operator ρ. Thus the compression protocol designed above works for all qubit sources whose associated density operator has eigenvalues $\{p, 1-p\}$. On problem 3.3 you will show that by a very simple extension of this idea one obtains a truely *universal quantum compression protocol*: *It is targeted at a fixed compression rate S_0 and is able to compress an arbitrary qubit source whose density operator has entropy $S(\rho) < S_0$.* This universality is *not* automatic using the textbook approach to asymptotic quantum information theory, and it is one of the main advantages of the Schur-Weyl toolbox introduced in section 7.1.

7.3 Supplement: Quantum state tomography

Starting with our solution to the spectrum estimation problem, we can also solve the problem of estimating an unknown quantum state from many copies – a task that is also known as *quantum state tomography*. That is, given $\rho^{\otimes n}$, we would like to design a POVM measurement that yields an estimate $\hat{\rho} \approx \rho$ with high probability,

$$\rho^{\otimes n} \longrightarrow \hat{\rho} \approx \rho.$$

We follow the approach of Haah et al. (2015) (but see the original paper by Keyl (2006) and other exciting recent works by O'Donnell and Wright (2015, 2016)).

The POVM measurement

The general idea is that we would like to design a POVM measurement $\{Q_{j,U}\}$ with *two* outcomes j and U, such that the estimate is

$$\hat{\rho} = U \begin{pmatrix} \hat{p} & \\ & 1-\hat{p} \end{pmatrix} U^\dagger.$$

As before, j is a discrete parameter that we will use for the eigenvalue estimate $\hat{p} = \frac{1}{2} + \frac{j}{n}$, while U is a continuous parameter that rotates the diagonal matrix with eigenvalues $\{\hat{p}, 1-\hat{p}\}$ into the proper eigenbasis. In order for $\{Q_{j,U}\}$ to be a POVM, we need that $Q_{j,U} \geq 0$ as well as

$$\sum_j \int dU \, Q_{j,U} = \mathbb{1}, \tag{7.8}$$

where $\int dU$ denotes the *Haar measure* of the unitary group $U(2)$. This is the unique probability measure on $U(2)$ such that all expectation values are invariant under the substitution $U \mapsto VUW^\dagger$ for unitaries V, W. Moreover, we would like for the POVM $\{Q_{j,U}\{$ to be a refinement of $\{P_j\}$, so that the j have the same meaning as before. That is, if we forget about the outcome U then we would like to get the same statistics for j as if we performed the measurement $\{P_j\}$. Mathematically, this means that we would like to demand that

$$\int dU \, Q_{j,U} = P_j \tag{7.9}$$

which clearly implies eq. (7.8). What does such a POVM look like?

We will make the ansatz

$$Q_{j,U} \propto P_j \hat{\rho}^{\otimes n} P_j = P_j U^{\otimes n} \begin{pmatrix} \hat{p} & \\ & 1-\hat{p} \end{pmatrix}^{\otimes n} U^{\dagger,\otimes n} P_j \cong T_{\hat{p}}^{(n,j)} \otimes \mathbb{1}_{W_{n,j}}.$$

To see that this is natural, we observe that, for $j = \frac{n}{2}$, P_j is the projector Π_n onto the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^2)$. Moreover, $\hat{p} = 1$, hence

$$\hat{\rho} = U |0\rangle \langle 0| U^\dagger =: |\hat{\psi}\rangle \langle \hat{\psi}| ,$$

and so

$$Q_{n/2,U} \propto \Pi_n \hat{\rho}^{\otimes n} \Pi_n = |\hat{\psi}\rangle^{\otimes n} \langle \hat{\psi}|^{\otimes n}$$

is exactly the uniform POVM (2.8) that we used for pure state estimation in lecture 2. Thus, our POVM measurement $Q_{j,U}$ is a true generalization of what we did for pure states – that's already an encouraging sign. Moreover, note that $Q_{j,U}$ has permutation symmetry (i.e., $[R_\pi, Q_{j,U}] = 0$) and it is *covariant* with respect to the unitary group in the following sense: For all $V \in U(2)$,

$$\mathrm{tr}\left[\rho Q_{j,U}\right] = \mathrm{tr}\left[V \rho V^\dagger Q_{j,VU}\right],$$

where we note that estimate corresponding to the POVM element $Q_{j,VU}$ is $V\hat{\rho}V^\dagger$. We could summarize this as $\rho \mapsto V\rho V^\dagger \rightsquigarrow \hat{\rho} \mapsto V\hat{\rho}V^\dagger$.

We will now show that eq. (7.9) holds true by a suitable choice of normalization constant. For this, we first note that

$$\int dU\, Q_{j,U} \cong \underbrace{\int dU\, T^{(n,j)}_{\hat{\rho}}}_{\propto \mathbb{1}_{V_{n,j}}} \otimes \mathbb{1}_{W_{n,j}} \propto P_j$$

as a consequence of Schur's lemma. Indeed, the indicated operator is a self-intertwiner on the irreducible representation $V_{n,j}$, because

$$\begin{aligned} T^{(n,j)}_V \left(\int dU\, T^{(n,j)}_{\hat{\rho}} \right) T^{(n,j)}_{V^\dagger} &= T^{(n,j)}_V \left(\int dU\, T^{(n,j)}_U T^{(n,j)}_{\begin{pmatrix} \hat{p} & \\ & 1-\hat{p} \end{pmatrix}} T^{(n,j)}_{U^\dagger} \right) T^{(n,j)}_{V^\dagger} \\ &= \int dU\, T^{(n,j)}_{VU} T^{(n,j)}_{\begin{pmatrix} \hat{p} & \\ & 1-\hat{p} \end{pmatrix}} T^{(n,j)}_{(VU)^\dagger} = \int dU\, T^{(n,j)}_U T^{(n,j)}_{\begin{pmatrix} \hat{p} & \\ & 1-\hat{p} \end{pmatrix}} T^{(n,j)}_{U^\dagger} = \int dU\, T^{(n,j)}_{\hat{\rho}}; \end{aligned}$$

in the second to last step we used that the integral is invariant under the substitution $U \mapsto VU$. It is now easy to figure out the correct normalization constant – we merely need to compare traces. On the one hand, in view of the definition of $Q_{j,U}$, its trace that does not depend on U, and so

$$\mathrm{tr}\left[\int dU\, Q_{j,U} \right] = \mathrm{tr}\, Q_{j,U} = \mathrm{tr}\left[T^{(n,j)}_{\hat{\rho}} \right] (\dim W_{n,j})$$

for any U that we like. On the other hand,

$$\mathrm{tr}\, P_j = (\dim V_{n,j})(\dim W_{n,j}) = (2j+1)(\dim W_{n,j}).$$

Thus, the appropriately normalized POVM elements are

$$Q_{j,U} = \frac{2j+1}{\mathrm{tr}\left[T^{(n,j)}_{\hat{\rho}} \right]} P_j \hat{\rho}^{\otimes n} P_j.$$

The fidelity between two quantum states

In section 4.3 we discussed the trace distance $T(\rho,\sigma)$ as a distance measure between quantum states (whether pure or mixed). Another very useful measure was the overlap, $|\langle\phi|\psi\rangle|$, which we only defined for pure states. The overlap also generalizes nicely to mixed states, but the expression is more complicated: It is the following quantity, known as the *fidelity*:

$$F(\rho,\sigma) = \operatorname{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} = \operatorname{tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}$$

As in problem 1.4, $\sqrt{M}$ denotes the square root of a positive semidefinite operator M, defined by taking the square root of all eigenvalues. The fidelity might seem like a strange definition – but actually it is precisely the maximal overlap that can be obtained between any two purifications. That is,

$$F(\rho,\sigma) = \max_{|\phi\rangle_{AB},|\psi\rangle_{AB}} |\langle\phi_{AB}|\psi_{AB}\rangle|$$

where we optimize over all pure states $|\phi\rangle_{AB}, |\psi\rangle_{AB}$ such that $\operatorname{tr}_B[|\phi\rangle\langle\phi|_{AB}] = \rho$, $\operatorname{tr}_B[|\psi\rangle\langle\psi|_{AB}] = \sigma$. In particular, if $\rho = |\phi\rangle\langle\phi|$ and $\sigma = |\psi\rangle\langle\psi|$ are themselves pure then the fidelity agrees with the overlap. (You can also check this explicitly from the definition, since in that case $\sqrt{\rho} = \rho$ and $\sqrt{\sigma} = \sigma$.) In general, the trace distance and fidelity are related by the *Fuchs-van de Graaf inequalities*:

$$1 - F(\rho,\sigma) \le T(\rho,\sigma) \le \sqrt{1 - F^2(\rho,\sigma)} \tag{7.10}$$

Analysis of the measurement

Similarly as when analyzing the spectrum estimation measurement, our goal is to show that $\operatorname{tr}[Q_{j,U}\rho^{\otimes n}]$ is exponentially small unless $\rho \approx \hat{\rho}$. Thus, we want to bound For this, we will use the full strength of the Schur-Weyl toolbox. We start with

$$\begin{aligned}\operatorname{tr}\left[Q_{j,U}\rho^{\otimes n}\right] &= \frac{2j+1}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}\right]} \operatorname{tr}\left[P_j\hat{\rho}^{\otimes n}P_j\rho^{\otimes n}\right] = \frac{2j+1}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}\right]} \operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}T_{\rho}^{(n,j)} \otimes \mathbb{1}_{W_{n,j}}\right] \\ &= \frac{(2j+1)m(n,j)}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}\right]} \operatorname{tr}\left[T_{\sqrt{\rho}\hat{\rho}\sqrt{\rho}}^{(n,j)}\right] = \frac{(2j+1)m(n,j)}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,j)}\right]} \operatorname{tr}\left[T_{\sqrt{\sqrt{\rho}\hat{\rho}\sqrt{\rho}}^2}^{(n,j)}\right].\end{aligned}$$

In the second to last step, we have used that $T_X^{(n,j)}T_Y^{(n,j)} = T_{XY}^{(n,j)}$ for arbitrary operators, as well as cyclicity of the trace. We now use the upper bound (7.3), the lower bound in (7.4) (observing that $\widehat{\rho}$ has eigenvalues $\{\hat{p}, 1-\hat{p}\}$), and the upper bound (7.5) (with $k = 2$). The result is that

$$\operatorname{tr}\left[Q_{j,U}\rho^{\otimes n}\right] \le \frac{(2j+1)2^{nh(\hat{p})}}{2^{-nh(\hat{p})}}(2j+1)2^{-2n(h(\hat{p})+\delta(\hat{p}\|x))}\left(\operatorname{tr}\sqrt{\sqrt{\rho}\hat{\rho}\sqrt{\rho}}\right)^{2n} \le (n+1)^2 F(\rho,\hat{\rho})^{2n},$$

where in the second step we used $\delta(\hat{p}\|x) \ge 0$ as well as $2j \le n$. This is the desired upper bound. Indeed, it implies that

$$\begin{aligned}&\Pr(T(\hat{\rho},\rho) \ge \varepsilon) \le \Pr(F(\hat{\rho},\rho)^2 \le 1-\varepsilon^2) \le \sum_j \int dU (n+1)^2(1-\varepsilon^2)^n \\ &\le (n+1)^3 2^{n\log(1-\varepsilon^2)} \le (n+1)^3 2^{-n\varepsilon^2}\end{aligned}$$

(The first inequality is a consequence of the upper bound in eq. (7.10). The last holds whenever $\varepsilon \le \frac{1}{2}$ and is only for illustration.)

Bibliography

Aram W Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. PhD thesis, 2005. arXiv:quant-ph/0512255.

Matthias Christandl. *The structure of bipartite quantum states-insights from group theory and cryptography*. PhD thesis, 2006. arXiv:quant-ph/0604183.

Michael Walter. *Multipartite quantum states and their marginals*. PhD thesis, 2014. arXiv:1410.6820.

Benjamin Schumacher. Quantum coding, *Physical Review A*, 51(4):2738, 1995.

Philippe Faist. Welcome to the entropy zoo. 2013. URL `http://www.statslab.cam.ac.uk/biid2013/slides/EntropyZoo.pdf`. Beyond IID.

Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. 2015.

Michael Keyl. Quantum state estimation and large deviations, *Reviews in Mathematical Physics*, 18 (01):19–60, 2006. arXiv:quant-ph/0412053.

Ryan O'Donnell and John Wright. Efficient quantum tomography. 2015. arXiv:1508.01907.

Ryan O'Donnell and John Wright. Efficient quantum tomography ii. 2016. arXiv:1612.00034.