PHYSICS 491: Symmetry and Quantum Information April 18, 2017

Shannon theory, data compression, spectrum estimation

Lecture 5 *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.

5.1 A first glance at information theory: data compression

Imagine that Alice has acquired a biased coin, with heads coming up with $p = 75\%$ probability. She is excited about her purchase and wants to let Bob know about the result of her coin flips. If the flips the coin once, how many bits does she need to communicate the result to Bob? Clearly, she needs at least one bit. Otherwise, since both outcomes are possible, she would make an error 25% of the time.

Now suppose that Alice flips her coin not only once, but a large number of times – say n times. She would still like to communicate the results of her coin flips to Bob. Clearly, Alice could send over one bit immediately after each coin flip. Can she do better by waiting and looking at the whole sequence of coin flips? If we assume that her coin flips are *independent* then we would expect that heads will come up $j \approx pn$ times for large enough n. This suggests the following compression scheme:

- If the number of coin flips j is not within $(p \pm \varepsilon)n$, Alice gives up and signals failure.
- Otherwise, she sends j over to Bob, as well as the index i of her particular sequence of coin flips in a list $\mathcal{L}_j$ that contains all possible coin flips with j heads and $n - j$ tails.

If our two protagonists have agreed beforehand on the lists $\mathcal{L}_j$ (you might call them a *codebook*), then Bob will have no trouble decoding the sequence of coin flips – he merely looks up the i-th entry in the list $\mathcal{L}_j$. Note that, for any fixed $\varepsilon > 0$, the probability of failure in the first step is arbitrarily small – this is a consequence of the strong law of large numbers.

Remark. *If failure is not an option, Alice may instead send the uncompressed sequence of coin flips instead of giving up. This leads to a similar analysis and will be left as an exercise.*

What is the compression rate of this protocol? To send j, we need roughly $(\log n)/n$ bits per coin flip, which is negligible for large n.[1] How many sequences are there with j heads and $n - j$ tails? This is given by the binomial coefficient $\binom{n}{j}$. Thus, to communicate the index $i \in \{1, \dots, \binom{n}{j}\}$, Alice needs to send roughly $\frac{1}{n} \log \binom{n}{j}$ bits per coin flip. To estimate this rate, we note that for any $x \in [0, 1]$,

$$x^j (1-x)^{n-j} \binom{n}{j} \leq (x + (1-x))^n = 1$$

[1] Here and throughout the rest of these lecture notes, log denotes the logarithm to the base two.

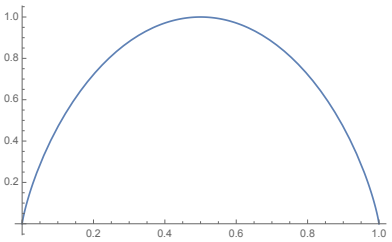

Figure 7: The binary entropy function $h(p)$ defined in eq. (5.2).

and hence, choosing $x = \frac{j}{n}$,

$$\frac{1}{n}\log\binom{n}{j} \leq -\frac{j}{n}\log\frac{j}{n} - (1-\frac{j}{n})\log(1-\frac{j}{n}). \tag{5.1}$$

Since $\frac{j}{n} \approx p$, the right-hand side is approximately equal to the *binary (Shannon) entropy*

$$h(p) := -p\log p - (1-p)\log(1-p). \tag{5.2}$$

See fig. 7 for a plot of the binary entropy function.

In total, the protocol sketched above will achieve a compression rate of roughly $h(p) \leq 1$ bits per coin flip. E.g., $h(75\%) = 0.81$ – so Alice achieve savings of roughly of 19%. We can get arbitrarily close to $h(p)$ by decreasing ε, at the expense of n having to become larger and larger for the probability of failure to vanish. It is not hard to see the compression rate $h(p)$ is optimal. This is Shannon's famous *noiseless coding theorem* – it is called "noiseless" since we assume that the communication line from Alice to Bob is perfect.

The coin flip example illustrates the traditional core principles of information theory, or *Shannon theory*: We are interested in finding *optimal asymptotic rates* for information processing tasks such as compression (the task that you have just solved), information transmission over noisy channels, etc. *Quantum information theory* has very analogous goals – except that now we are dealing with *quantum information* rather than classical information. At a fundamental level, this means that we are interested in the asymptotic behavior of a large number of independent copies of a quantum state ρ, i.e., in $\rho^{\otimes n}$ for large n (the so-called *i.i.d.* limit).

Example 5.1 (Warning)**.** *If $\rho = |\psi\rangle\langle\psi|$ is a pure state then $\rho^{\otimes n} = |\psi\rangle^{\otimes n}\langle\psi|^{\otimes n}$ is an operator on the symmetric subspace. We explored this quite extensively in lectures 2 to 4. However, as soon as ρ is a mixed state, $\rho^{\otimes n}$ is no longer supported purely on the symmetric subspace. A simple example is the maximally mixed state $\tau = \mathbb{1}/d$. Clearly, $\tau^{\otimes n} = \mathbb{1}/d^n$ is supported on all of $(\mathbb{C}^d)^{\otimes n}$. Thus we need to develop new techniques.*

Remark 5.2. *In recent years, there has been an increased interest in understanding optimal information processing rates in non-asymptotic scenarios. This is largely beyond the scope of these lectures, although we might have a brief glance at these ideas in the last week of class.*

5.2 Spectrum estimation

Today, we will start developing the appropriate machinary for working with independent copies of a quantum state, $\rho^{\otimes n}$. A popular approach that you will find in many textbooks is to work in the

eigenbasis of ρ in order to turn the quantum problem into a classical problem (e.g., Nielsen and Chuang, 2002, Wilde, 2013). *In this class we will pursue a different, and arguably more "invariant" route.* What this means exactly will become clear over the coming lectures, but the practical advantage of exploiting all available symmetries will be that we are naturally led to *universal protocols* that work not only for a single state ρ but for whole classes of states (e.g., all states ρ with the same eigenvalues).

When we discussed the symmetric subspace, our motivation was to solve an estimation problem, namely, the estimation an unknown pure state $|\psi\rangle$ given n copies $|\psi\rangle^{\otimes n}$. Today, we will again be interested in an estimation problem: We would like to estimate the eigenvalues of an unknown density operator ρ, given n copies $\rho^{\otimes n}$. That is, if $p_1 \geq \cdots \geq p_d$ denote the eigenvalues of ρ then we would like to define a measurement $\{Q_{\hat{p}}\}$ such that, when we measure on $\rho^{\otimes n}$, we obtain an outcome such that $\hat{p} \approx p$. This task is known as the *spectrum estimation* problem (Keyl and Werner, 2001). It is an easier problem than estimating the full density operator ρ, and it allows us to focus on the key difference between pure and mixed states – their eigenvalue spectrum. We will spend the rest of today's lecture and part of lecture 6 solving the spectrum estimation problem.

The tools that we will develop in the course of solving this problem will be prove useful for working with asymptotic quantum information more generally. In lecture 7, we will use them to compress quantum information and we will also sketch how one can estimate the entire unknown quantum state ρ from $\rho^{\otimes n}$, thereby solving the task of quantum states estimation of mixed state, also known as *quantum state tomography.*

Symmetries of the spectrum estimation problem

If ρ is a quantum state on $\mathbb{C}^d$ then the state $\rho^{\otimes n}$ is a quantum state on $(\mathbb{C}^d)^{\otimes n}$. As discussed in section 3.1, this space is a representation for two groups: (i) the permutation group S_n, with representation operators R_π, and (ii) the unitary group $U(d)$, with representation operators $T_U = U^{\otimes n}$. The operator $\rho^{\otimes n}$ is *permutation-invariant* as defined last time, i.e., it commutes with permutations, $[R_\pi, \rho^{\otimes n}] = 0$ for all $\pi \in S_n$. We may explicitly verify this on a product basis:

$$\begin{aligned} R_\pi \rho^{\otimes n} |x_1, \ldots, x_n\rangle &= R_\pi \left(\rho |x_1\rangle \otimes \ldots \otimes \rho |x_n\rangle\right) = \rho |x_{\pi^{-1}}\rangle \otimes \ldots \otimes \rho |x_{\pi^{-1}}\rangle \\ &= \rho^{\otimes n}\left(|x_{\pi^{-1}}\rangle \otimes \ldots \otimes |x_{\pi^{-1}}\rangle\right) = \rho^{\otimes n} R_\pi |x_1, \ldots, x_n\rangle . \end{aligned}$$

On the other hand, $\rho^{\otimes n}$ does *not* commute with the action of the unitary group: Instead,

$$U^{\otimes n} \rho^{\otimes n} U^{\dagger, \otimes n} = (U \rho U^\dagger)^{\otimes n}$$

which amounts to replacing $\rho \mapsto U \rho U^\dagger$. *This operation changes the eigenbasis, but leaves the eigenvalues the same.* In other words, while the permutation symmetry is a symmetry of the state, the unitary symmetry is a symmetry of the problem that we are trying to solve! This suggests that both symmetries should play an important role, and it prompts us to investigate the representation $(\mathbb{C}^d)^{\otimes n}$ more closely.

Example 5.3 (Warmup)**.** *Suppose we are just given two copies of the unknown quantum state, i.e., $\rho^{\otimes 2}$. This is a density operator on*

$$(\mathbb{C}^d)^{\otimes 2} = \mathrm{Sym}^2(\mathbb{C}^d) \oplus \bigwedge\nolimits^2(\mathbb{C}^d).$$

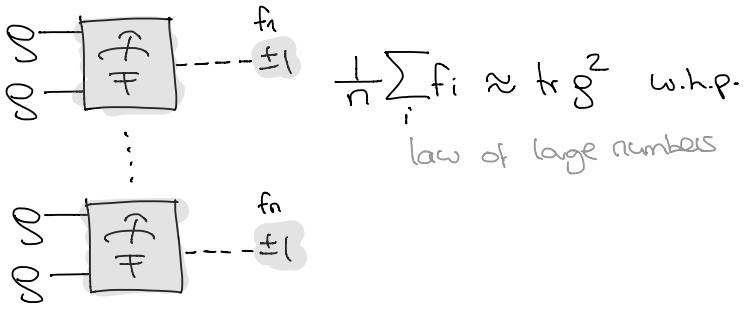

Figure 8: By measuring the swap operator on independent copies of $\rho^{\otimes 2}$, we can estimate the purity $\operatorname{tr} \rho^2$ of the quantum state.

Both the symmetric and the antisymmetric subspace are irreducible representations (you show this in problem 2.3 for the symmetric subspace; the antisymmetric subspace can be treated completely analogously). The permutation group S_2 has just two elements, the identity permutation and the nontrivial permutation $\pi = 1 \leftrightarrow 2$. The corresponding operator is known as the swap operator

$$F = R_\pi = \sum_{a,b} |a,b\rangle \langle b,a| .$$

It commutes both with the representation of $U(d)$ as well as the one of S_2 (any operator commutes with itself and with the identity matrix). Thus, F is an observable of exactly the kind that we are looking for. Its eigenvalues are $+1$ on the symmetric subspace and -1 on the antisymmetric subspace. In problem 2.1, you show the following "swap trick":

$$\langle F \rangle = \operatorname{tr} \rho^{\otimes 2} F = \operatorname{tr} \rho^2.$$

The quantity $\operatorname{tr} \rho^2$ is called the purity *of ρ, since it is equal to 1 only if the state ρ is a pure state. (It is closely related to* Rényi-2 entropy $S_2(\rho) = -\log \operatorname{tr} \rho^2$ *that you study in problem 2.1.) The important point though is that if ρ has eigenvalues $r_1 \geq \cdots \geq r_d$ then*

$$\operatorname{tr} \rho^2 = \sum_k r_k^2,$$

and hence already this simple measurement allows us to learn something about the eigenvalues of ρ.

Just to be perfectly clear: When measuring the observable F on $\rho^{\otimes 2}$, the measurement outcome is either ± 1. Only when repeated many times on independent copies of $\rho^{\otimes 2}$ will these signs average to $\operatorname{tr} \rho^2$ (fig. 8).

For qubits, $d = 2$, example 5.3 provides a complete solution (since $p_1 + p_2 = 1$, there is only a single unknown, which can be determined from $\operatorname{tr} \rho^2 = p_1^2 + p_2^2$). In the following, we will discuss a different solution which fully exploits the symmetries of the problem and generalizes readily to any d. The protocol is due to Keyl and Werner (2001) and we will follow the proof strategy of Christandl and Mitchison (2006). It will prove to be an important building block for several quantum information applications that we will discuss in the remainder of this course.

Towards a solution of the spectrum estimation problem

We start by decomposing the Hilbert space of n qubits into irreducible representations of $\mathrm{SU}(2)$. The answer can be written ni the form:

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j V_j \otimes \mathbb{C}^{m(n,j)}, \tag{5.3}$$

where V_j denotes the irreducible representation of $\mathrm{SU}(2)$ with spin j and $m(n,j)$ are the multiplicities that we need to determine. That is, for any $U \in \mathrm{SU}(2)$ we have that

$$U^{\otimes n} \cong \bigoplus_j T_U^{(j)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}} = \left[\begin{array}{c|c|c} T_U^{(0)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,0)}} & & \\ \hline & T_U^{(1/2)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,1/2)}} & \\ \hline & & \ddots \end{array}\right]. \tag{5.4}$$

Here we write $T_U^{(j)}$ for the representation operators of the spin-j representation.

Recall that we are looking for a measurement that commutes with both the action of $\mathrm{SU}(2)$ and S_n. The projection operator P_j onto a direct summand in eq. (5.3) seems like a plausible candidate. It measures the total spin – generalizing example 5.3. By design, P_j commutes with the action of the unitary group. Indeed, in view of eq. (5.4) it clearly commutes with $U \in \mathrm{SU}(2)$, and any element of $U(2)$ can be written in the form $e^{i\phi}U$ where $U \in \mathrm{SU}(2)$.

Does P_j also commute with the action of S_n? Yes, this follows from $[R_\pi, U^{\otimes n}] = 0$ and Schur's lemma, as you will verify in problem 3.5. We have found the desired candidate measurement!

In the remainder of today's lecture, we will start analyzing the projective measurement $\{P_j\}$. That is, we would like to bound the probabilities

$$\Pr(\text{outcome j}) = \mathrm{tr}\left[\rho^{\otimes n} P_j\right]. \tag{5.5}$$

Note that these probabilities remain unchanged if we substitute $\rho \mapsto U\rho U^\dagger$, as P_j commutes with $U^{\otimes n}$. Since we can always diagonalize ρ by a unitary there is thus no harm in assuming that ρ is already a diagonal matrix

$$\rho = \begin{pmatrix} p & \\ & 1-p \end{pmatrix} \tag{5.6}$$

with $p \geq 1-p$, i.e., $p \in [\frac{1}{2}, 1]$. Our goal will be to show that (5.5) is exponentially small in n most of the time – except when we can obtain a good estimate of the spectrum from j (we will later see that $\hat{p} := \frac{1}{2} + \frac{j}{n} \approx p$ provides such an estimate).

How would we go about analyzing eq. (5.5)? The idea is that $\rho^{\otimes n}$ looks just like the representation operators $U^{\otimes n}$ – except that ρ is almost never a unitary matrix! To go beyond unitaries, we need to talk about some more representation theory.

Representation theory of $\mathrm{SU}(2)$ and $\mathrm{SL}(2)$

As we have already used several times in this course, the irreducible representations of $\mathrm{SU}(2)$ are labeled by their spin $j \in \{0, \frac{1}{2}, 1, \frac{3}{2}, \dots\}$. We denote the spin-$j$ irrep by V_j and its representation operators by $T_U^{(j)}$. The representation V_j is of dimension $2j+1$.

Remark 5.4. *In your quantum mechanics class, you have probably analyzed the representation theory of* $\mathrm{SU}(2)$ *by considering its "generators": For any traceless Hermitian matrix* H, $U = \exp(iH)$ *is in* $\mathrm{SU}(2)$. *Given a representation* $\tilde{\mathcal{H}}$ *of* $\mathrm{SU}(2)$ *with representation operators* $\tilde{T}_U$, *we can define*

$$\widetilde{H} = \frac{1}{i}\frac{d}{dt}\Big|_{=0}\tilde{T}_{\exp(itH)}.$$

Sometimes this is called the representation of the Lie algebra of $\mathrm{SU}(2)$ *(though technically speaking the Lie algebra of* $\mathrm{SU}(2)$ *consists of the* anti*hermitian traceless matrices). Note that the assignment* $H \mapsto \widetilde{H}$ *is linear. Since the real vector space of traceless Hermitian matrices is spanned by the Pauli operators X,Y,Z (the "generators"), we can fully understand the representation* $\tilde{\mathcal{H}}$ *by considering the operators* $\widetilde{X}, \widetilde{Y}, \widetilde{Z}$.

In your quantum mechanics class, you likely followed this approach to analyze the irreducible representations of $\mathrm{SU}(2)$. *For example, you might remember that* V_j *has a basis* $|j,m\rangle$, *where* $m = -j, \ldots, j-1, j$, *such that*

$$\widetilde{Z}\,|j,m\rangle = 2m\,|j,m\rangle\,.$$

Moreover,

$$\widetilde{Q} = (\widetilde{X})^2 + (\widetilde{Y})^2 + (\widetilde{Z})^2 = 4j(j+\frac{1}{2})\mathbb{1}_{V_j}.$$

The operator $\widetilde{Q}$ *is called the* quadratic Casimir operator *of* $\mathrm{SU}(2)$, *and we used the fact that it acts by a scalar on each irreducible representation of* $\mathrm{SU}(2)$ *in lecture 1 to find a qubit .*

In the previous lectures, we used to great effect that the symmetric subspace is irreducible – and you will show this in problem 2.3 by following precisely the strategy outlined in the preceding remark. This means that $\mathrm{Sym}^n(\mathbb{C}^2)$ ought to be one of the spin-j irreps. It is very easy to see that $j = \frac{n}{2}$, and we record this important fact:

$$V_j \cong \mathrm{Sym}^{2j}(\mathbb{C}^2). \tag{5.7}$$

It gives us a very simple way of realizing the spin-j representation concretely, as will be prove useful in just a momenet.

An important fact that was perhaps never explicitly spelled out in your quantum mechanics class is the following: Any unitary representation of $\mathrm{SU}(2)$ can be extended to a (holomorphic, non-unitary) representation of the group $\mathrm{SL}(2)$ in a unique way. For example, our representation $T_U = U^{\otimes n}$ of $\mathrm{SU}(2)$ on $(\mathbb{C}^d)^{\otimes n}$ can be extended to $T_g = g^{\otimes n}$ for $g \in \mathrm{SL}(2)$. We can also restrict this action to the symmetric subspace. Since we can define the spin-j representation using the symmetric subspace (eq. (5.7)), we can likewise define $T_g^{(j)}$ for any $g \in \mathrm{SL}(2)$. Thus, for any $g \in \mathrm{SL}(2)$, eq. (5.3) reads

$$g^{\otimes n} \cong \bigoplus_j T_g^{(j)} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}}. \tag{5.8}$$

Remark. *A general way of defining the extension from* $\mathrm{SU}(2)$ *to* $\mathrm{SL}(2)$ *is as follows: In remark 5.4, we defined* $\widetilde{H}$ *for Hermitian matrices we can safely extend it by linearity to arbitrary complex traceless matrices* M. *But then* $\exp(M)$ *is an arbitrary matrix in* $\mathrm{SL}(2)$ *and this allows us to extend an arbitrary unitary representation of* $\mathrm{SU}(2)$ *to* $\mathrm{SL}(2)$*: For* $g = \exp(M)$, *define* $R_g := \exp(\widetilde{M})$. *It is not hard to see that a subspace is invariant for* $\mathrm{SU}(2)$ *iff it is invariant for the operators* $\widetilde{H}$ *iff it is invariant for the operators* $\widetilde{M}$ *iff it is invariant for* $\mathrm{SL}(2)$. *This can be used to argue that the finite-dimensional representation theory of* $\mathrm{SU}(2)$ *and of* $\mathrm{SL}(2)$ *is completely identical.*

Bounding the probability distribution

Why is this important? We are interested in understanding the operator $\rho^{\otimes n}$ on $(\mathbb{C}^2)^{\otimes n}$. Suppose that our density matrix ρ has no zero eigenvalues. Then it is is invertible and

$$\tilde{\rho} := \rho/\sqrt{\det \rho}$$

is an element in the group $\mathrm{SL}(2)$, and we can interpret $\tilde{\rho}^{\otimes n}$ as the corresponding representation operator on $(\mathbb{C}^2)^{\otimes n}$! By eq. (5.8), it follows that

$$\rho^{\otimes n} = (\det \rho)^{n/2}\, \tilde{\rho}^{\otimes n} \cong (\det \rho)^{n/2} \bigoplus_j T^{(j)}_{\tilde{\rho}} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}} = \bigoplus_j \underbrace{(\det \rho)^{n/2}\, T^{(j)}_{\tilde{\rho}}}_{=:T^{(n,j)}_{\rho}} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}} \tag{5.9}$$

By continuity, this equation can be extended to all $\rho \geq 0$.

Remark. *Since any operator X can be infinitesimally perturbed to become invertible, we can use the same strategy to analyze $X^{\otimes n}$ for arbitrary operators X on $\mathbb{C}^2$.*

As a consequence of eq. (5.9), our desired probability (5.5) reads

$$\operatorname{tr}\left[P_j \rho^{\otimes n}\right] = \operatorname{tr}\left[T^{(n,j)}_{\rho} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}}\right] = (\det \rho)^{n/2} \operatorname{tr}\left[T^{(j)}_{\tilde{\rho}} \otimes \mathbb{1}_{\mathbb{C}^{m(n,j)}}\right] = m(n,j)(\det \rho)^{n/2} \operatorname{tr}\left[T^{(j)}_{\tilde{\rho}}\right].$$

How can we compute the right-hand side trace? By eq. (5.7) we can simply compute the trace of $\tilde{\rho}^{\otimes 2j}$ on the symmetric subspace:

$$\operatorname{tr}\left[T^{(j)}_{\tilde{\rho}}\right] = \sum_{k=0}^{2j} \langle\!\langle k \| \tilde{\rho}^{\otimes 2j} \| k \rangle\!\rangle = (\det \rho)^{-j} \sum_{k=0}^{2j} \langle\!\langle k \| \rho^{\otimes 2j} \| k \rangle\!\rangle = (\det \rho)^{-j} \sum_{k=0}^{2j} p^k (1-p)^{2j-k} \leq (\det \rho)^{-j} (2j+1) p^{2j}.$$

Here, we compute the trace in the occupation number basis

$$\| k \rangle\!\rangle \propto |0\rangle^{\otimes k} |1\rangle^{\otimes (2j-k)} + \text{permutations}$$

of the symmetric subspace (see eq. (2.5) and problem 2.3). In the third step, we used that ρ is diagonal, and in the last step we bounded each summand by p^{2j} using that $p \geq 1-p$ (see eq. (5.6)). Thus:

$$\begin{aligned} &\operatorname{tr}\left[T^{(n,j)}_{\rho}\right] = (\det \rho)^{n/2} \operatorname{tr}\left[T^{(j)}_{\tilde{\rho}}\right] \leq (2j+1)(\det \rho)^{n/2-j} p^{2j} = (2j+1) p^{\frac{n}{2}+j} (1-p)^{\frac{n}{2}-j} \\ &= (2j+1) 2^{n\left[\left(\frac{1}{2}+\frac{j}{n}\right)\log p + \left(\frac{1}{2}-\frac{j}{n}\right)\log(1-p)\right]} = (2j+1) 2^{n\left[\hat{p}\log p + (1-\hat{p})\log(1-p)\right]}. \end{aligned} \tag{5.10}$$

where we have defined $\hat{p} := \frac{1}{2} + \frac{j}{n}$. If we plug this back into the preceding equation then we obtain

$$\operatorname{tr}\left[P_j \rho^{\otimes n}\right] \leq (2j+1) m(n,j) 2^{n\left[\hat{p}\log p + (1-\hat{p})\log(1-p)\right]}.$$

This already looks quite suggestively as if the eigenvalue p has something to do with $\hat{p}$!

However, we still need to determine the multiplicities $m(n,j)$. We will do this next time – it will allow us to solve the spectrum estimation problem completely. We will then put the tools developed into a more general context and use them to tackle a number of important applications.

Bibliography

Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

Mark M Wilde. *Quantum information theory*. Cambridge University Press, 2013.

Michael Keyl and Reinhard F Werner. Estimating the spectrum of a density operator, *Physical Review A*, 64(5):052311, 2001. arXiv:quant-ph/0102027.

Matthias Christandl and Graeme Mitchison. The spectra of quantum states and the Kronecker coefficients of the symmetric group, *Communications in Mathematical Physics*, 261(3):789–797, 2006. arXiv:quant-ph/0409016.