Symmetry and Quantum Information

Lecture Notes

Michael Walter, Ruhr University Bochum

Summary

These are notes for a course given at Stanford University and the University of Amsterdam for advanced undergraduates and graduate students. The course gives an introduction to quantum information theory. Somewhat unconventionally, it uses *symmetries* as a guiding principle to study fundamental features of quantum information and solve quantum information processing tasks. For more thorough introductions to the subject, see the textbooks by Nielsen and Chuang [NC10], Wilde [Wil17], or Watrous [Wat18], or the lecture notes by Preskill [Pre22].

Acknowledgements

I am grateful to Sepehr Nezami and Freek Witteveen for their careful proofreading of these notes. Thanks also to Jeroen Dekker, Felix Leditzky, Christian Majenz, Maris Ozols, Anthony Polloreno, Grant Salton, and Zhaoyou Wang for comments on earlier versions. Finally, I would like to thank Matthias Christandl, Aram Harrow, and Patrick Hayden for many inspiring discussions which shaped my understanding of quantum information and the subject of these notes.

Last updated: Aug 20, 2024. Please send corrections to michael.walter@rub.de.

Contents

Chapter 1

Introduction to quantum mechanics, uncertainty principle

By now, quantum information science is an established field, with theoreticians and experimentalists seeking to leverage the laws of quantum mechanics to process information and compute in fundamentally new and interesting ways. But quantum information theory also offers a fresh perspective on fundamental physics, by providing us with a versatile language and a useful toolbox to clarify abstract notions such as information and computing and how they are realized in the physical world.

This course on *Symmetry and Quantum Information* will give an introduction to this way of thinking and provide you with a concrete toolbox for your future endeavors in quantum information and computing. We will discuss a number of fundamental information theoretic problems, such as the storage, measurement, compression, and transmission of quantum information. Our guiding principle will be to identify the symmetries that are hidden behind these problems (an approach that many of you may well be familiar from your previous courses in mathematics and physics), and we will learn how to leverage those symmetries using the machinery of group representation theory to solve the problems at hand.

1.1 Axioms of quantum mechanics

Today, we start with an introduction to the axioms (laws, postulates) of quantum mechanics. We will careful go through each axiom and discuss a number of consequences and challenges that will motivate much of what we will study in this course. While the following list will be roughly what you remember from a previous course on quantum mechanics, you should think of it as a *first attempt*. As we go along this term, we will extend our repertoire and rephrase these rules in a way that (while equivalent) is more useful from the perspective of quantum information theory.

Axiom A (Systems)**.** *To every quantum mechanical system, we associate a* Hilbert space $\mathcal{H}$*. For a joint system composed of two subsystems A and B, with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the Hilbert space is the tensor product $\mathcal{H}_{AB} := \mathcal{H}_A \otimes \mathcal{H}_B$.*

Throughout this course we will restrict to finite-dimensional Hilbert spaces. Recall that a finite-dimensional Hilbert space is nothing but a vector space together with an inner product, which we denote by $\langle \cdot | \cdot \rangle$. By convention, the inner product is linear in the *second* argument.

The simplest quantum mechanical system is the *qubit* (short for *quantum bit*), described by the two-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^2$. Thus a system composed of n qubits corresponds to $(\mathbb{C}^2)^{\otimes n} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \ldots \otimes \mathbb{C}^2$. Note that the dimension of the latter space is 2^n, which is

exponential in the number of qubits. This explains some of the difficulty in simulating quantum mechanics on an ordinary "classical" (i.e., non-quantum) computer.

Axiom B (Pure States)**.** *Unit vectors* $|\psi\rangle \in \mathcal{H}$ *describe the* state *of quantum mechanical systems.*

Here we use Dirac's "bra-ket" notation, with "kets" $|\psi\rangle$ denoting vectors in $\mathcal{H}$ and "bras" $\langle\psi|$ denoting the corresponding dual vector in $\mathcal{H}^*$, i.e., $\langle\psi| := \langle\psi|\cdot\rangle$. Thus, "bra" and "ket" together give the inner product: $\langle\phi|\psi\rangle = \langle\phi|\,|\psi\rangle$. A unit vector is a vector $|\psi\rangle$ whose norm (or norm squared) is equal to one, i.e., $\langle\psi|\psi\rangle = 1$. We will denote by $X^\dagger$ the adjoint of an operator X between two Hilbert spaces. We can think of $|\psi\rangle \in \mathcal{H}$ as an operator $\mathbb{C} \to \mathcal{H}$, so that $\langle\psi| = |\psi\rangle^\dagger$.

Note that the notation $|\psi\rangle\langle\psi|$ is precisely the projection onto the one-dimensional subspace spanned by $|\psi\rangle$. When we say *projection* or *"projector"*, we always mean an orthogonal projection, that is, a linear operator P that satisfies $P^2 = P^\dagger = P$. In coordinates (i.e., for $\mathcal{H} = \mathbb{C}^d$):

$$|\psi\rangle = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_d \end{pmatrix}, \quad \langle\psi| = \begin{pmatrix} \overline{\psi_1} & \cdots & \overline{\psi_d} \end{pmatrix}, \quad \langle\phi|\psi\rangle = \sum_{i=1}^{d} \overline{\phi_i}\psi_i, \quad |\psi\rangle\langle\phi| = \begin{pmatrix} \psi_1\overline{\phi_1} & \cdots & \psi_1\overline{\phi_d} \\ \vdots & & \vdots \\ \psi_d\overline{\phi_1} & \cdots & \psi_d\overline{\phi_d} \end{pmatrix},$$

and the adjoint is given by the conjugate transpose: $X^\dagger = (\overline{X})^T = \overline{X^T}$.

When we speak of a *basis* of a Hilbert space then we always mean an *orthonormal* basis. The standard basis or *computational basis* of $\mathbb{C}^d$ is denoted by $|0\rangle, |1\rangle, \ldots, |d-1\rangle$. In particular, for a qubit, the computational basis is given by

$$|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

We can think of these as the two states of a classical bit that has been embedded into a qubit: $\{0,1\} \ni x \mapsto |x\rangle \in \mathbb{C}^2$. This makes sense because these two states are orthogonal, $\langle 0|1\rangle = 0$, and as we shall see below this means that they can be perfectly distinguished. For n qubits, we write

$$|i_1 \ldots i_n\rangle := |i_1, \ldots, i_n\rangle := |i_1\rangle \otimes \ldots \otimes |i_n\rangle$$

for the computational basis of $(\mathbb{C}^2)^{\otimes n}$.

The fact that for any two states $|\phi\rangle$ and $|\psi\rangle$ we have an entire continuum of "superposition" states $\alpha\,|\phi\rangle + \beta\,|\psi\rangle$ of states is sometimes called the *superposition principle*.

Some states of a composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ can be written as a tensor product, $|\Psi\rangle = |\alpha\rangle \otimes |\beta\rangle$. Such states are called *product states*; otherwise they are called *entangled*. An example of an entangled state of two qubits is the following,

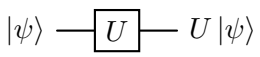
$$|\Phi^+\rangle := \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right) \in \mathbb{C}^2 \otimes \mathbb{C}^2, \tag{1.1}$$

which is known as a *maximally entangled state*, an *EPR pair*, or simply as an *ebit*. In Exercise 1.1 you can show that this state is indeed entangled.

In Chapter 2, we will see some first indications that entanglement can be a powerful resource for quantum information processing. In Chapter 3, we will see that it can lead to strong correlations that go beyond what can be produced by a classical local theory.

Axiom C (Unitary dynamics)**.** *Given a* unitary *operator* U *on* $\mathcal{H}$*, the transformation* $|\psi\rangle \mapsto U\,|\psi\rangle$ *is physical. In other words, in principle one can engineer an evolution of a quantum system for some finite time such that, when we start in an arbitrary initial state* $|\psi\rangle$*, the final state is* $U\,|\psi\rangle$*.*

Recall that a *unitary* operator U is one such that $UU^\dagger = U^\dagger U = I$, i.e., the adjoint is the inverse (we denote identity operators by I). Unitary matrices are precisely those linear maps that map unit vectors to unit vector. We denote the set of unitary operators by $\mathrm{U}(\mathcal{H})$. We will use pictures such as the following to indicate an evolution by some unitary U:



The relationship to the Schrödinger equation is that, in order to implement a given unitary, one can evolve the quantum system for some time with a suitable Hamiltonian.

The next axiom explains how to extract classical information from a quantum system. Before stating it, we recall the spectral theorem for Hermitian operators. This theorem asserts that any Hermitian operator O can be diagonalized, with real eigenvalues and an orthonormal eigenbasis. We can write this as $O = \sum_{x\in\Omega} x\, P_x$, where $\Omega \subseteq \mathbb{R}$ is the (finite) set of eigenvalues of O and where P_x denotes the orthogonal projection onto the eigenspace corresponding to eigenvalue $x \in \Omega$. Eigenspaces for distinct eigenvalues are orthogonal: we have $P_x P_y = \delta_{x,y} P_x$. If an eigenspace is one-dimensional and spanned by some unit vector $|e_x\rangle$, we can write $P_x = |e_x\rangle\langle e_x|$.

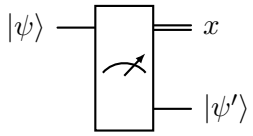
Axiom D (Observables)**.** *Any Hermitian operator O on $\mathcal{H}$ corresponds to an observable quantity or measurement. Let $O = \sum_{x\in\Omega} x\, P_x$ be the* spectral decomposition*. Then the Born rule asserts that the probability of outcome x in state $|\psi\rangle$ is given by*

$$\mathbf{Pr}_\psi(\textit{outcome } x) = \langle\psi|P_x|\psi\rangle = \|P_x|\psi\rangle\|^2. \tag{1.2}$$

(We will often omit the subscript ψ if the state is clear.) Moreover, if the outcome is x then the quantum state of the system changes ("collapses") into the post-measurement state

$$|\psi'\rangle = \frac{P_x|\psi\rangle}{\|P_x|\psi\rangle\|} = \frac{P_x|\psi\rangle}{\sqrt{\langle\psi|P_x|\psi\rangle}}. \tag{1.3}$$

Measurements will be indicated as follows:



The top right wire carries the measurement outcome, while the bottom one the post-measurement state (we may leave out one or the other). We follow the convention that single lines correspond to quantum systems, while double lines refer to classical information.

As a consequence of the Born rule in Eq. (1.2), the *expectation value* of the outcome of a quantum measurement can be succinctly expressed in terms of the observable O:

$$\mathbf{E}_\psi[\text{outcome}] = \sum_{x\in\Omega} x\,\langle\psi|P_x|\psi\rangle = \langle\psi|O|\psi\rangle\,,$$

Axiom D states that, in general, measurement outcomes are probabilistic and lead to a "collapse" of the quantum state. This is a very fundamental statement with numerous consequences. For example, it implies that quantum information *cannot be "cloned"*, that is, copied (in contrast to, say, the value of an ordinary bit in the memory of your computer). In fact, we will find that when we want to process quantum information, we have to do so in a way that avoids learning anything about the state of the quantum information itself – for learning is equivalent to measuring an aspect of the state, and measurements in general lead to a "collapse" of the

quantum state in the sense of Eq. (1.3). We will later see how to make this precise. This is a major challenge and closely related to the "fragility" of quantum information – but it also gives rise to a powerful way of constructing quantum communication protocols known as the *decoupling principle*, which we will discuss in Chapter 17.

Given an observable O, are there states $|\psi\rangle$ that are *not* "collapsed" by the measurement of an observable? In other words, are the states for which the post-measurement state is equal to the state before the measurement? It is not hard to see that this happens precisely when $|\psi\rangle$ is an eigenvector of O (by Eq. (1.3)), i.e., when $P_x |\psi\rangle = \delta_{x,x_0} |\psi\rangle$, where x_0 is the corresponding eigenvalue. Equivalently, this means that $\langle\psi|P_x|\psi\rangle = \delta_{x,x_0}$, i.e., the states that do not collapse are precisely those states for which the measurement outcome is deterministic ("certain").

A closely related question is when a pair of quantum states $\{|\alpha\rangle, |\beta\rangle\}$ can be *perfectly distinguished* by some observable. That is, when does there exists an observable O such that when we measure on $|\alpha\rangle$ we always obtain outcome $+1$, say, while if we measure on $|\beta\rangle$ we always obtain outcome -1, as in the following figure?

The answer is that this is possible precisely when the two states are orthogonal, i.e., $\langle\alpha|\beta\rangle = 0$. Indeed, in this case we can measure the observable $O = |\alpha\rangle\langle\alpha| - |\beta\rangle\langle\beta|$, which has $|\alpha\rangle$ as an eigenvector with eigenvalue $+1$ and $|\beta\rangle$ as an eigenvector with eigenvalue -1. In Exercise 1.2, you can show the converse: two states can be perfectly distinguished *only* when they are orthogonal.

We conclude our discussion of the axioms of quantum mechanics with one last observation. A careful look at Axioms A-D reveals that any two states $|\psi\rangle$ and $e^{\imath\theta}|\psi\rangle$ are completely indistinguishable. This means that there is some redundancy when we characterize states by unit vectors – we should really identify any two unit vectors that can be obtained from each other by multiplication with a complex number of absolute value one (a "global phase"). Mathematically, this means that we should work with the projective space $\mathbb{P}(\mathcal{H})$. One convenient way to achieve this is to consider $|\psi\rangle\langle\psi|$, which as mentioned is the orthogonal projection onto the one-dimensional subspace spanned by a unit vector $|\psi\rangle$. Note that the subspace and hence also the operator $|\psi\rangle\langle\psi|$ are insensitive to multiplying the state by an overall phase $e^{\imath\theta}$. Conversely, we can recover $|\psi\rangle$ up to phase by choosing any unit vector in the range of the operator $|\psi\rangle\langle\psi|$, so this achieves precisely what we wanted. A useful notation is to write $\psi := |\psi\rangle\langle\psi|$. Note that we can rephrase all our axioms in terms of ψ. For example, the unitary dynamics $|\psi\rangle \mapsto U|\psi\rangle$ now becomes $\psi \mapsto U\psi U^\dagger = U|\psi\rangle\langle\psi|U^\dagger$, Born's rule reads $\mathbf{Pr}_\psi(\text{outcome } x) = \operatorname{tr}[\psi P_x] = \operatorname{tr}[|\psi\rangle\langle\psi| P_x] = \langle\psi|P_x|\psi\rangle$, and the corresponding post-measurement state is $\psi' = P_x\psi P_x / \operatorname{tr}[P_x\psi]$.

What kind of object is ψ? It is positive semidefinite (which we write as $\psi \geq 0$) and its trace is $\operatorname{tr}[\psi] = \operatorname{tr}[|\psi\rangle\langle\psi|] = \langle\psi|\psi\rangle = 1$. In Chapter 7 we will see that these two properties define the notion of a *density operator*, which is a useful and physical generalization of the notion of a quantum state as introduced in Axiom B.

1.2 Measuring a qubit

For an ordinary bit, there is essentially only a single interesting measurement: Is the bit in state 0 or is it in state 1? For a quantum bit, however, Axiom D provides us with (infinitely) many

inequivalent measurements that we can perform. For example, consider the three *Pauli matrices*

$$\begin{aligned} X &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = |+\rangle\langle +| - |-\rangle\langle -|, \\ Y &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = |L\rangle\langle L| - |R\rangle\langle R|, \\ Z &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = |0\rangle\langle 0| - |1\rangle\langle 1|, \end{aligned} \tag{1.4}$$

which are all Hermitian and have eigenvalues ± 1 (so they are also unitary!). On the right-hand side, we indicated their spectral decomposition in terms of the eigenvectors

$$\begin{aligned} |+\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \qquad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle); \\ |L\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle), \qquad |R\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i\,|1\rangle); \end{aligned}$$

as well as the computational basis vectors $|0\rangle$ and $|1\rangle$, which we have already met. The basis vectors $|+\rangle$ and $|-\rangle$ make up the so-called *Hadamard basis*.

We briefly discuss some useful mathematical properties. First, the three Pauli matrices together with the identity matrix form a basis of the real vector space of the Hermitian 2×2 matrices. This means that any Hermitian operator on $\mathbb{C}^2$ can be written as $O = \alpha I + \beta X + \gamma Y + \delta Z$ for some suitable $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. In fact, they form an orthogonal basis with respect to the inner product $(O, O') := \mathrm{tr}[O^\dagger O'] = \mathrm{tr}[OO']$. The latter can be easily seen from the relations

$$X^2 = Y^2 = Z^2 = I$$

and

$$XY = iZ, \quad YZ = iX, \quad ZX = iY, \tag{1.5}$$

together with the fact that the Pauli matrices are traceless. Second, the Pauli matrices do *not* commute. This follows from Eq. (1.5), which implies that $[X, Y] := XY - YX = 2iZ$ and similarly $[Y, Z] = 2iX$ and $[Z, X] = 2iY$. In fact, the Pauli matrices *anti-commute*, meaning that $\{X, Y\} := XY + YX = 0$ and similarly $\{Y, Z\} = \{Z, X\} = 0$. In Exercise 1.3 you can show that this implies that if we measure two Pauli matrices then the order in which we do so is important!

1.3 An uncertainty relation

Above we discussed that when we measure an observable then the states for which outcome is deterministic are precisely the observable's eigenvectors. But no pair of Pauli operators has a joint eigenvector, as is clear from the spectral decompositions in Eq. (1.4). Accordingly, for *every* state $|\psi\rangle$ and any pair of Pauli operators, say X and Z, there is necessarily some uncertainty in either the measurement outcome for X or in the measurement outcome for Z (or both).

To make this statement quantitative, we need a way to quantify the uncertainty in a measurement outcome. To this end, we can use

$$\langle\psi|X|\psi\rangle^2 = (p_X(1) - p_X(-1))^2 = (2p_X(1) - 1)^2 = (1 - 2p_X(-1))^2, \tag{1.6}$$

where $p_X(x)$ denotes the probability of outcome x when measuring the observable X in state $|\psi\rangle$. Clearly,

$$0 \leq \langle\psi|X|\psi\rangle^2 \leq 1.$$

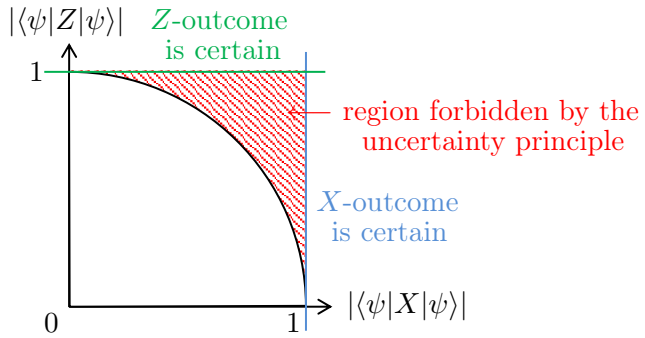


Figure 1.1: An illustration of the region excluded by the uncertainty relation in Theorem 1.1.

When are these values saturated? The upper bound is saturated precisely when either $p_X(1) = 1$ or $p_X(-1) = 1$, that is, when the measurement outcome is *certain*. On the other hand, the lower bound is saturated when $p_X(1) = p_X(-1) = 1/2$, which means that the measurement outcome is *completely uncertain* (i.e., uniformly random). Thus, $\langle\psi|X|\psi\rangle^2$ is a meaningful way to quantify the certainty of the measurement outcome when the quantum bit is in state $|\psi\rangle$ and we measure the observable X. We can similarly quantify the certainty of an Z measurement outcome by $\langle\psi|Z|\psi\rangle^2$.

By adding the upper bound for X and for Z, we obtain that

$$\langle\psi|X|\psi\rangle^2 + \langle\psi|Z|\psi\rangle^2 \leq 2.$$

But note that this upper bound can never be saturated – for otherwise $|\psi\rangle$ would be a state such that both outcomes are certain, and we have argued that no such state exists. This means that we must in fact have $\langle\psi|X|\psi\rangle^2 + \langle\psi|Z|\psi\rangle^2 < 2$. We will now show a significant strengthening. Namely, we will show that the sum of the two "certainties" cannot even exceed one (see also Fig. 1.1). Such a result is called an *uncertainty relation*.

Theorem 1.1 (Uncertainty relation for Pauli matrices)**.** *For every state $|\psi\rangle \in \mathbb{C}^2$, it holds that*

$$\langle\psi|X|\psi\rangle^2 + \langle\psi|Z|\psi\rangle^2 \leq 1 < 2, \tag{1.7}$$

and similarly for the other two pairs of Pauli matrices.

Proof. We prove the result by a brute-force calculation (in a couple of weeks we will be able to give a more beautiful argument). Let $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \in \mathbb{C}^2$ be an arbitrary unit vector. Then:

$$\begin{aligned}\langle\psi|X|\psi\rangle^2 + \langle\psi|Z|\psi\rangle^2 &= \big(\bar{\alpha}\beta + \bar{\beta}\alpha\big)^2 + \big(\bar{\alpha}\alpha - \bar{\beta}\beta\big)^2 \\ &= \bar{\alpha}^2\beta^2 + \alpha^2\bar{\beta}^2 + |\alpha|^4 + |\beta|^4 \\ &= (\alpha^2 + \beta^2)(\bar{\alpha}^2 + \bar{\beta}^2) = \big|\alpha^2 + \beta^2\big|^2,\end{aligned}$$

and now the claim follows because $|\psi\rangle$ is a unit vector and hence $|\alpha^2 + \beta^2| \leq |\alpha|^2 + |\beta|^2 = 1$. □

We close with one final remark on the interpretation of the above result. Similarly as in Eq. (1.6), we can write

$$|\langle\psi|X|\psi\rangle| = 2\max\{p_X(1), p_X(-1)\} - 1 = 2p_{\text{guess},X} - 1,$$

where the quantity

$$p_{\text{guess},X} := \max\{p_X(1), p_X(-1)\}$$

is often called a *guessing probability* because it can be interpreted as the maximal probability of correctly guessing the outcome of an X-measurement on the state $|\psi\rangle$ in advance of performing it (just go for the outcome that has the larger probability – there is no better way). Now observe that, Eq. (1.7) implies that $|\langle\psi|X|\psi\rangle| + |\langle\psi|Z|\psi\rangle| \leq \sqrt{2}$ using the Cauchy-Schwarz inequality. We can write this as follows in terms of guessing probabilities:

$$p_{\text{guess},X} + p_{\text{guess},Z} \leq 1 + \frac{1}{\sqrt{2}} < 2$$

This has a very transparent interpretation: it simply bounds the sum of the probabilities of guessing the two measurement outcomes correctly.

Uncertainty relations of the above form are powerful since they make nontrivial predictions for every quantum state (the upper bound is nontrivial and in fact independent of $|\psi\rangle$). More sophisticated uncertainty relations play an important role in quantum cryptography.

Exercises

1.1 **Entanglement criterion:** Let $|\Psi\rangle = \sum_{i,j} M_{ij} |i\rangle \otimes |j\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ be an arbitrary quantum state, expanded in the computational basis. Let M denote the $d \times d$-matrix with entries M_{ij}.

(a) Show that $|\Psi\rangle = |\alpha\rangle \otimes |\beta\rangle$ for some $|\alpha\rangle$, $|\beta\rangle \in \mathbb{C}^d$ if and only if the rank of M is one.
(b) Conclude that the ebit state $|\Phi^+\rangle$ defined in Eq. (1.1) is entangled.

1.2 **Perfectly distinguishing quantum states:** Show that if two quantum states $|\alpha\rangle$ and $|\beta\rangle$ can be perfectly distinguished by an observable, then the states must be orthogonal: $\langle\alpha|\beta\rangle = 0$.

1.3 **Order of meausurements:** In this problem, you will see how the order of measurements can matter in quantum mechanics. Let $|\psi\rangle$ be an arbitrary state of a qubit.

(a) Imagine that we first measure the Pauli matrix X, with outcome x, and then the Pauli matrix Z, with outcome z. Derive a formula for the joint probability, denoted $p(x \to z)$, of the two measurement outcomes.
(b) Derive a similar formula for the joint probability $p(x \leftarrow z)$ corresponding to first measuring Z and then X.
(c) Find a state $|\psi\rangle$ such that $p(x \to z) \neq p(x \leftarrow z)$.

In fact, this is a general feature of noncommutativity:

(d) Let O and O' be two arbitrary observables. Show that the order of measurement is irrelevant *for every state* precisely when $[O, O'] = 0$.

Chapter 2

Entanglement as a resource, generalized measurements

Last time we discussed the axioms of quantum mechanics and in particular the measurement of observables. In particular, for any observable O with spectral decomposition $O = \sum_{x\in\Omega} x\, P_x$, the probability of obtaining an outcome $x \in \Omega$ is given by Born's rule, $\langle\psi|P_x|\psi\rangle = \|P_x\,|\psi\rangle\|^2$, and the post-measurement state is given $P_x\,|\psi\rangle\,/\|P_x\,|\psi\rangle\|$. Note that the preceding only makes use of the collection of projections $\{P_x\}_{x\in\Omega}$ rather than the observable O itself. As discussed, we have that $P_x^2 = P_x^\dagger = P_x$, $\sum_x P_x = I$, and $P_x P_y = \delta_{x,y} P_x$ for all x, $y \in \Omega$. We will refer to any collection of projections $\{P_x\}_{x\in\Omega}$ which these properties as a *projective measurement*. To summarize:

$$O = \sum_{x\in\Omega} x\, P_x \quad \leftrightarrow \quad \{P_x\}_{x\in\Omega}.$$

While any projective measurement can be implemented by the measurement of an observable, it is often useful to directly work with the projections. For example, we can then allow the set of possible outcomes to be an arbitrary finite set Ω, not necessarily a subset of $\mathbb{R}$ (as one would get for the eigenvalues of a Hermitian operator). This is just a simple relabeling, but often convenient.

Before we launch into the main subject of today's lecture, let us discuss one last axiom that we did not spell out explicitly in the last lecture:

Axiom E (Operations on subsystems)**.** *Consider a quantum system composed of two subsystems, with joint Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. If we want to perform a unitary U_A on the subsystem modeled by $\mathcal{H}_A$, then the appropriate unitary on the joint system is $U_A \otimes I_B$. Similarly, if O_A is an observable on $\mathcal{H}_A$ then the appropriate observable on the joint system is $O_A \otimes I_B$. Equivalently, if $\{P_{A,x}\}_{x\in\Omega}$ is a projective measurement on $\mathcal{H}_A$ then the corresponding projective measurement on $\mathcal{H}_{AB}$ is $\{P_{A,x} \otimes I_B\}_{x\in\Omega}$.*

Note that the set of possible measurement outcomes remains the same (O_A and $O_A \otimes I_B$ have the same eigenvalues, albeit with different multiplicities), which is of course what we expect. Above and throughout these notes we follow the common convention of labeling subsystems by A ("Alice"), B ("Bob"), etc., and using subscripts to indicate the subsystems that mathematical objects such as states or operators are associated with.

Let us consider an example. Take the ebit state from Eq. (1.1),

$$|\Phi_{AB}^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right) = \frac{1}{\sqrt{2}} \sum_{i\in\{0,1\}} |i\rangle_A \otimes |i\rangle_B\,.$$

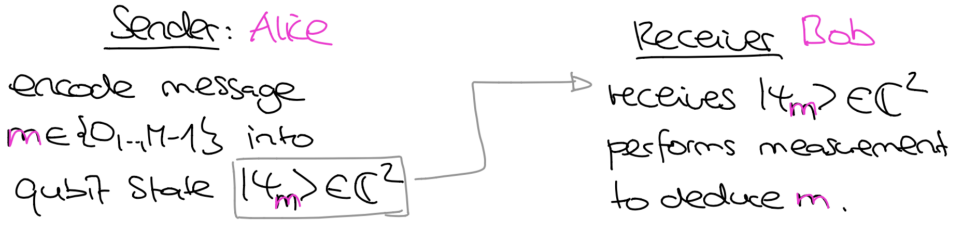
Take an arbitrary basis $\{|e_x\rangle\}_{x\in\{0,1\}}$ of $\mathbb{C}^2$ and denote by $P_{A,x} := |e_x\rangle\langle e_x|_A$ the corresponding projective measurement. If we apply this measurement on the first subsystem of the ebit, the probability of outcomes according to the Born rule is given by

$$\begin{aligned}\mathbf{Pr}_{\Phi^+}(\text{outcome } x) &= \langle\Phi^+_{AB}|P_{A,x}\otimes I_B|\Phi^+_{AB}\rangle \\ &= \frac{1}{2}\sum_{i,j}(\langle i|_A\otimes\langle i|_B)(P_{A,x}\otimes I_B)(|j\rangle_A\otimes|j\rangle_B) \\ &= \frac{1}{2}\sum_i \langle i_A|P_{A,x}|j_A\rangle \underbrace{\langle i_B|I_B|j_B\rangle}_{=\delta_{i,j}} \\ &= \frac{1}{2}\sum_i \langle i_A|P_{A,x}|i_A\rangle = \frac{1}{2}\operatorname{tr}[P_{A,x}] = \frac{1}{2},\end{aligned} \tag{2.1}$$

since the trace of a projection is equal to its rank. Thus, for any basis measurement we obtain either outcome with 50% probability. This is quite interesting – even though the joint system is in a well-defined state, measurement outcomes on the subsystem are completely uninformative. Nevertheless, ebits can be a useful resource in communication scenarios, as we will discuss next. This will also help us clarify the distinctions between bits and qubits.

2.1 Encoding bits into qubits: superdense coding

Consider a scenario where a sender – commonly called *Alice* – would like to send one out of M possible classical messages to a receiver – commonly called *Bob* – by sending a single qubit. Here is a sketch of a possible *communication protocol*:



What is the maximal number M of possible classical messages such that Bob can always perfectly decode? This requires that the quantum states $|\psi_m\rangle$ are all orthogonal, since only orthogonal quantum states can be distinguished perfectly (i.e., with zero probability of error), as we discussed last time. Thus, $M \leq 2$, since no three states in a two-dimensional Hilbert space can be orthogonal. Moreover, $M = 2$ can clearly be achieved – simply encode the two messages using any orthonormal basis, such as the computational basis: $m \mapsto |\psi_m\rangle := |m\rangle$ for $m \in \{0,1\}$. In summary, we found that by sending over a single qubit we can perfectly communicate a single bit, but no more.

In fact, as a consequence of the so-called *Holevo bound*, it es even impossible to communicate at an *asymptotic* rate higher than the trivial rate of one classical bit per qubit sent.

Superdense coding

We will now see that we can overcome this "no go" result by using entanglement. For this, consider the following set of vectors in $\mathbb{C}^2 \otimes \mathbb{C}^2$:

$$\begin{aligned}
|\beta_0\rangle &:= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = (I \otimes I)\,|\Phi^+\rangle\,, \\
|\beta_1\rangle &:= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) = (Z \otimes I)\,|\Phi^+\rangle\,, \\
|\beta_2\rangle &:= \frac{1}{\sqrt{2}}(|10\rangle + |01\rangle) = (X \otimes I)\,|\Phi^+\rangle\,, \\
|\beta_3\rangle &:= \frac{1}{\sqrt{2}}(|10\rangle - |01\rangle) = (XZ \otimes I)\,|\Phi^+\rangle\,.
\end{aligned} \tag{2.2}$$

Note that $\{|\beta_m\rangle\}_{m\in\{0,1,2,3\}}$ is an orthonormal basis, so the four states can be perfectly distinguished by a two-qubit measurement (namely the one with projections $P_{AB,m} = |\beta_m\rangle\langle\beta_m|$). This is called the *Bell basis* of two qubits. Moreover, as indicated on the right, each of the four states can be produced from the ebit by applying one out of the four unitaries I, Z, X, XZ on Alice's side. That is, we can write

$$|\beta_m\rangle_{AB} = (U_{A,m} \otimes I_B)\,|\Phi^+_{AB}\rangle\,,$$

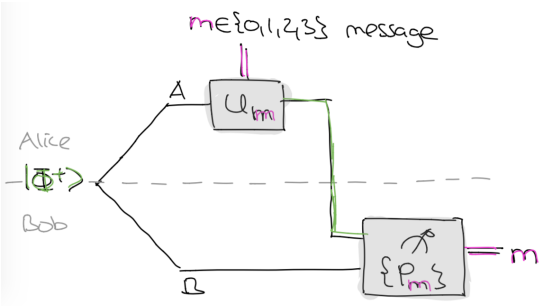
where $U_{A,0} = I$, $U_{A,1} = Z$, and so forth.

The preceding considerations suggest a communication protocol known as *superdense coding*. It allows Alice to communicate *two bits* (i.e., one out of four messages) by sending a single qubit to Bob, provided Alice and Bob already share an ebit [BW92]:

Protocol 2.1 (Superdense coding)**.** The goal is for Alice to communicate a two-bit message $m \in \{0, 1, 2, 3\}$ by sending one qubit to Bob.

1. Assume Alice and Bob share an ebit $|\Phi^+_{AB}\rangle$.
2. To send $m \in \{0, 1, 2, 3\}$, Alice applies the unitary $U_{A,m}$ to her qubit and sends the qubit over to Bob.
3. Upon receiving Alice's qubit, Bob applies the projective measurement $\{P_{AB,m}\}_{m\in\{0,1,2,3\}} = \{|\beta_m\rangle\langle\beta_m|_{AB}\}_{m\in\{0,1,2,3\}}$ on *both* qubits in his possession. The outcome is Alice's message m.

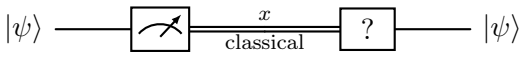
The correctness of the protocol follows directly by the preceding discussion. Here is an illustration of the protocol:



Of course, in order to establish the ebit between Alice and Bob, some form of prior (quantum) communication must have occurred. The point however is the following: the ebit state used in the protocol is completely independent (and can therefore be shared well in advance) of the message m that is later sent in superdense coding. Thus, *shared entanglement* is a *resource* that, once established, can be used for information processing tasks (such as the one we just saw: communicating classical bits at twice the rate than what is possible without shared entanglement).

2.2 Encoding qubits into bits, teleportation

Now we consider the "dual" problem to the above. Suppose Alice has a qubit in some arbitrary unknown state $|\psi\rangle$ in her possession (i.e., a *quantum message*!) that she would like to communicate to Bob. Can she do so if she is only able to send over a classical bit or bitstring x?

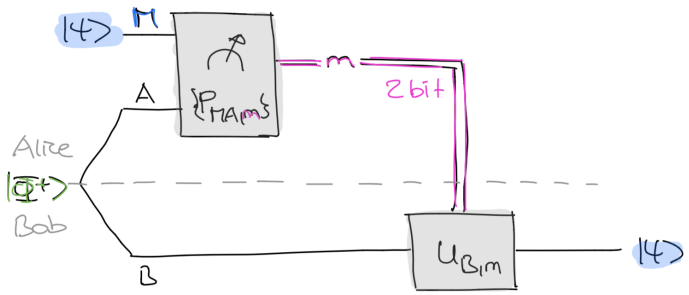


This is clearly impossible, provided that they want to achieve this task perfectly. An easy way to see this is that there are only finitely possible values for x, but infinitely many quantum states – so there must be two distinct quantum states corresponding to the same bitstring x, which is a contradiction.

A sharper argument is the following: Suppose that the protocol works for arbitrary qubit states, so in particular for $|0\rangle$ and $|+\rangle$. Then the protocol must send over different bitstrings x for these two states. Because any two bitstrings are perfectly distinguishable, this means that we have found a way to perfectly distinguish two quantum states that are not orthogonal. As discussed in the last lecture, this is impossible.[1] In summary, we found that is *not* possible to (perfectly) communicate an unknown qubit by sending any number of classical bits.

Teleportation

We will now see that this task becomes possible (and in fact two bits suffice) in the presence of shared entanglement. The protocol is called *(quantum) teleportation* and it looks as follows [BBC+93]:



The protocol uses the same elements as above, but in a different order and on different subsystems. Let us first write down the protocol more precisely and then see why it works:

Protocol 2.2 (Superdense coding)**.** The goal is for Alice to communicate a qubit M in some arbitrary unknown state by sending two bits to Bob.

1. Assume Alice and Bob share an ebit $|\Phi^+_{AB}\rangle$.
2. Next, Alice measures $\{P_{MA,m}\}_{m\in\{0,1,2,3\}} = \{|\beta_m\rangle\langle\beta_m|_{MA}\}_{m\in\{0,1,2,3\}}$ on *both* qubits in her possession, and sends the outcome m over to Bob.
3. Upon receiving m, Bob applies the unitary $U_{B,m}$. Now the qubit B is in the same state that M was in before the start of the protocol.

To verify that the protocol works as advertised we compute the joint state of all qubits after Alice's measurement if the message qubit M is initially in some arbitrary state $|\psi\rangle_M$ and the

[1] In Exercise 1.2 you showed that two qubit states can be distinguished by an *observable* measurement only if they are orthogonal. The same is true for arbitrary procedures. See Section 2.4 and Exercise 2.4.

measurement outcome is $m \in \{0,1,2,3\}$. Using the rules for measuring subsystems (Axioms D and E), we should calculate:

$$\begin{aligned}(P_{MA,m} \otimes I_B)\big(|\psi_M\rangle \otimes |\Phi^+_{AB}\rangle\big) &= (|\beta_{MA,m}\rangle \otimes I_B)(\langle\beta_{MA,m}| \otimes I_B)\big(|\psi_M\rangle \otimes |\Phi^+_{AB}\rangle\big) \\ &= (|\beta_{MA,m}\rangle \otimes I_B)\big(\langle\Phi^+_{MA}| \otimes I_B\big)\Big(U^\dagger_{M,m} \otimes I_{AB}\Big)\big(|\psi_M\rangle \otimes |\Phi^+_{AB}\rangle\big) \\ &= (|\beta_{MA,m}\rangle \otimes I_B)\underbrace{\big(\langle\Phi^+_{MA}| \otimes I_B\big)\big(I_M \otimes |\Phi^+_{AB}\rangle\big)} U^\dagger_{M,m}\,|\psi_M\rangle\,.\end{aligned}$$

We now compute the underbraced expression:

$$\begin{aligned}\big(\langle\Phi^+_{MA}| \otimes I_B\big)\big(I_M \otimes |\Phi^+_{AB}\rangle\big) &= \frac{1}{2}\sum_{i,j}(\langle i_M| \otimes \langle i_A| \otimes I_B)(I_M \otimes |j_A\rangle \otimes |j_B\rangle) \\ &= \frac{1}{2}\sum_{i,j}\delta_{i,j}\,|j_B\rangle\langle i_M| = \frac{1}{2}I_{M\to B},\end{aligned}$$

it is simply one half times $I_{M\to B}$, the identity map from qubit M to qubit B. Thus,

$$(P_{MA,m} \otimes I_B)\big(|\psi_M\rangle \otimes |\Phi^+_{AB}\rangle\big) = \frac{1}{2}\,|\beta_{MA,m}\rangle \otimes U^\dagger_{B,m}\,|\psi_B\rangle\,. \tag{2.3}$$

Thus we see that the three qubits are in state $|\beta_m\rangle_{MA} \otimes U^\dagger_{B,m}\,|\psi\rangle_B$ right after the measurement. In the last step of the protocol, Bob applies the unitary $U_{B,m}$ on his qubit to obtain the desired state in his subsystem. Thus the teleportation protocols indeed works as advertised.

Equation (2.3) also shows that the four possible measurement outcomes $m \in \{0,1,2,3\}$ occur with probability $1/4$ each – irrespective of the state $|\psi\rangle_M$ of the qubit that Alice wants to teleport to Bob. This means that Alice's measurement does not reveal any information about the teleported state. In Chapter 17 we will discuss the decoupling principle, which implies that this is both necessary and sufficient for a teleportation protocol to succeed.

What happens if the message qubit is entangled with another subsystem? In other words, what does the teleportation protocol do when the initial state is

$$|\psi\rangle_{MR} \otimes |\Phi^+\rangle_{AB}\,,$$

where R is an additional "reference" system? In Exercise 2.1 you can show that the result is $|\psi\rangle_{BR} \otimes |\beta_m\rangle_{AB}$. Thus, Bob's qubit is now entangled with R in the same way that previously Alice's qubit was entangled with E, which is exactly the desired behavior. This is also known as *entanglement swapping*, because it can be used to establish entanglement between subsystems that have not initially been entangled!

2.3 Resources for information processing tasks

We have now seen two examples where the maximally entangled state served as a *resource* that enables information processing tasks. Similarly, the capability of sending a classical or a quantum bit can be though of as resources. We can use some symbolic notation for this and denote the former by $[c \to c]$ and the latter by $[q \to q]$. More generally, we write formal linear combinations such as ebit $+\, 2[c \to c]$ for combinations of these resources and use inequalities $\geq$ to compare them. For example

$$[q \to q] \geq [c \to c]$$

means that if we are able to send over a qubit then we can also use this to send a bit. That is, an inequality such as the above means that the left-hand side resources are sufficient to implement the right-hand side ones (allowing arbitrary local quantum operations on Alice and Bob's side).

We can use this notation to conveniently summarize the discussion of the preceding sections. Thus,

$$\begin{aligned} [q \to q] &\not\geq 2[c \to c], \\ \text{ebit} + [q \to q] &\geq 2[c \to c], \end{aligned}$$

where the second inequality summarizes superdense coding. Likewise, the fact that no number $n \in \mathbb{N}$ of classical bits enables the capability of communicating qubits, while teleportation shows that two bits suffice in the presence of a shared ebit can be summarized as

$$\begin{aligned} n[c \to c] &\not\geq [q \to q], \\ \text{ebit} + 2[c \to c] &\geq [q \to q]. \end{aligned}$$

In other words, classical and quantum communication become equivalent when shared entanglement is free (up to a factor two):

$$2[c \to c] \equiv [q \to q] \pmod{\text{ebit}}$$

Furthermore, it is clear that

$$[q \to q] \geq \text{ebit},$$

because in order to establish a shared ebit, Alice can prepare it locally and send over one qubit to Bob, and it is intuitive plausible (and we will prove later) that for any $n \in \mathbb{N}$,

$$n\,\text{ebit} \not\geq [q \to q], \qquad n\,\text{ebit} \not\geq [c \to c],$$

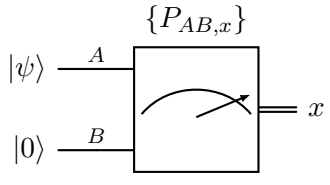
meaning that shared entanglement alone *cannot* be used to communicate.

2.4 POVM measurements

We conclude today's lecture by returning to the subject of measurements. So far, we always used observables or projective measurements

$$O = \sum_{x \in \Omega} x P_x \quad \leftrightarrow \quad \{P_x\}_{x \in \Omega}.$$

Are these the most general "measurement" procedures we can think of? Certainly not! For example, even if our goal is to extract information about a quantum system A, we could introduce an auxiliary system B that we initialize in some fixed state, say $|0\rangle_B$, and then apply an arbitrary projective measurement $\{P_{AB,x}\}_{x \in \Omega}$ on the joint system, as illustrated below:



What is the probability of an outcome $x \in \Omega$? According to the Born rule, Eq. (1.2), it is

$$\begin{aligned}\mathbf{Pr}_{|\psi\rangle}(\text{outcome } x) &= (\langle\psi_A| \otimes \langle 0_B|) P_{AB,x} (|\psi_A\rangle \otimes |0_B\rangle) \\ &= \langle\psi_A| \left(\underbrace{(I_A \otimes \langle 0_B|) P_{AB,x} (I_A \otimes |0_B\rangle)}_{=:Q_x} \right) |\psi_A\rangle\end{aligned}$$

where we have introduce new operators $\{Q_x\}_{x\in\Omega}$ on $\mathcal{H}_A$. These operators have the following properties:

(a) $Q_x \geq 0$ for all $x \in \Omega$ (meaning they are positive semidefinite), and
(b) $\sum_{x\in\Omega} Q_x = I_A$.

We will call any collection of operators $\{Q_x\}_{x\in\Omega}$ satisfying properties (a) and (b) a *POVM* with outcomes in Ω. POVM is short for *positive-operator valued measure*. The operators Q_x are called *POVM elements*. As we saw above, the probability of outcomes when performing a POVM measurement takes the familiar form of the *Born rule*:

$$\mathbf{Pr}(\text{outcome } x) = \langle\psi|Q_x|\psi\rangle\,. \tag{2.4}$$

A POVM measurement that has two possible outcomes is called a *binary POVM measurement*, and it has the form $\{Q, I - Q\}$, hence is specified by a single POVM element $0 \leq Q \leq I$. Note that the Q_x need not be pairwise orthogonal, nor will they in general be projections! In particular, it will no longer be true that we can rewrite $\langle\psi|Q_x|\psi\rangle$ as $\|Q_x\,|\psi\rangle\|^2$.

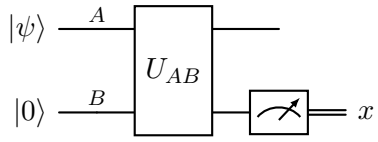
Example 2.3. The four operators $\{\frac{1}{2}\,|0\rangle\langle 0|\,, \frac{1}{2}\,|1\rangle\langle 1|\,, \frac{1}{2}\,|+\rangle\langle +|\,, \frac{1}{2}\,|-\rangle\langle -|\}$ make up a POVM with four possible outcomes. Measuring it is the same as performing either a measurement in the standard basis $|0\rangle\,, |1\rangle$ or in the Hadamard basis$|+\rangle\,, |-\rangle$, with 50% probability each.

Example 2.4. The operators $\{\frac{2}{3}\,|0\rangle\langle 0|\,, \frac{2}{3}\,|\alpha^+\rangle\langle\alpha^+|\,, \frac{2}{3}\,|\alpha^-\rangle\langle\alpha^-|\}$, where $|\alpha^\pm\rangle = \frac{1}{2}\,|0\rangle \pm \frac{\sqrt{3}}{2}\,|1\rangle$, make up a POVM with three outcomes. Indeed, it is easily verified that

$$\frac{2}{3}\,|0\rangle\,\langle 0| + \frac{2}{3}\,|\alpha^+\rangle\,\langle\alpha^+| + \frac{2}{3}\,|\alpha^-\rangle\,\langle\alpha^-| = I.$$

Unlike the previous example, this POVM cannot be decomposed in an interesting way.

POVM measurements are the most general "memoryless" measurements (as we defined them, with finitely many outcomes) provided by quantum mechanics. Importantly, any POVM can be implemented by a projective measurement on a larger system. However, note that a POVM only prescribes the probabilities of outcomes, but *not* the post-measurement state. In general, there are many different ways of implementing a POVM by a projective measurement on a larger system. For example, this can always be achieved as follows, as you can show in Exercise 2.3:



(2.5)

where B is a quantum system such that the Hilbert space $\mathcal{H}_B$ has one basis vector $|x_B\rangle$ for each possible measurement outcome $x \in \Omega$, U_{AB} is a unitary, and the measurement is simple the corresponding basis measurement. We can think of this intuitively as coupling our quantum system to a measurement apparatus, applying a unitary, and reading off the result by measuring

the apparatus. Note that the last two steps can be combined into a single projective measurement by taking $P_{AB,x} = U_{AB}^\dagger (I_A \otimes |x\rangle\langle x|_B) U_{AB}$.

We have just expanded our quantum information toolbox. While POVMs can always be implemented by projective measurements on a larger system, they can sometimes outperform projective measurements on the same system. In Exercise 2.2 you will see an example of this. On the other hand, is it still true that only orthogonal states can be perfectly distinguished, as you will show in Exercise 2.2.

Exercises

2.1 **Entanglement swapping:** Here you will discuss teleportation in a setting where the message qubit can be entangled with another system.

(a) Let $|\psi\rangle_{MR}$ be an arbitrary quantum state and consider the state $|\psi\rangle_{MR} \otimes |\Phi^+\rangle_{AB}$. Suppose that the M and A subsystems are in Alice' laboratory and the B subsystem is in Bob's laboratory, so that they can apply the teleportation protocol as in class. (Neither Alice nor Bob have access to the R subsystem.) Show that after completion of the teleportation protocol, the state of the B and R subsystems is $|\psi\rangle_{BR}$.

(b) Now assume that we have three nodes: Alice, Bob, and Charlie. Alice and Bob start out by sharing an ebit, and Bob and Charlie also start out by sharing an ebit. In other words, the initial state is $|\Phi^+\rangle_{AB_1} \otimes |\Phi^+\rangle_{B_2C}$. Explain how the three parties can establish an ebit between Alice and Charlie without sending any quantum information.

(c) Sketch how to extend the scheme in (b) to a linear chain of N nodes, assuming that initially only neighboring nodes share ebits.

2.2 **POVMs can outperform projective measurements [NC10, §2.2.6)]:** Imagine a qubit source that emits the two states $|0\rangle$ and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ with equal probability 1/2. As these are not orthogonal, they cannot be distinguished perfectly (Exercises 1.2 and 2.4).

Your task is to design a measurement that distinguishes the two states as well as you can in the following scenario. Your measurement is allowed to report one of *three* possible outcomes: that the true state is $|0\rangle$, that the true state is $|+\rangle$, or that you are not sure (the measurement has been inconclusive). However, it is *not allowed to ever give a wrong answer*! We define the success probability of such a measurement as the probability that you identify the true state.

(a) Show that for any projective measurements the success probability is at most 1/4.

(b) Find a POVM measurement that achieves a success probability strictly larger than 1/4.

2.3 **Implementing POVM measurements:** In this exercise, you will show that every POVM measurement can be realized by a projective measurement on a larger system. Thus, let $\{Q_x\}_{x\in\Omega}$ be an arbitrary POVM measurement on some Hilbert space $\mathcal{H}_A$.

(a) Let $\mathcal{H}_B$ be a Hilbert space with one basis vector $|x\rangle_B$ for each $x \in \Omega$, and fix some arbitrary $x_0 \in \Omega$. Show that the linear map

$$|\psi\rangle_A \otimes |x_0\rangle_B \mapsto \sum_x \sqrt{Q_x}\, |\psi\rangle_A \otimes |x\rangle_B \tag{2.6}$$

is an isometry (an isometry is a linear map V that preserves inner products; equivalently, $V^\dagger V = I$). Here, $\sqrt{Q_x}$ is the *square root* of the positive semidefinite operator Q_x defined by taking the square root of each eigenvalue while keeping the same eigenspaces.

Any isometry from a subspace into a larger Hilbert space can be extended to a unitary operator on the larger space. Thus there is a unitary U_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$ that extends the isometry (2.6).

(b) Use U_{AB} to design a projective measurement $\{P_{AB,x}\}$ on the joint system such that

$$Q_x = (I_A \otimes \langle x_0|_B)\, P_{AB,x}\, (I_A \otimes |x_0\rangle_B)$$

for all outcomes $x \in \Omega$.

(c) Conclude that any POVM measurement can be implemented as in (2.5).

2.4 **Distinguishing quantum states:** The *trace distance* between two (pure) quantum states $|\phi\rangle$ and $|\psi\rangle$ can be defined as follows:

$$T(\phi,\psi) = \max_{0\le Q\le I} (\langle\phi|Q|\phi\rangle - \langle\psi|Q|\psi\rangle) \tag{2.7}$$

Here, $0 \le Q \le I$ means that both Q and $I - Q$ are positive semidefinite operators.

(a) Imagine you are handed either $|\phi\rangle$ or $|\psi\rangle$ with probability 1/2 each. Show that the optimal probability of correctly identifying the state by a POVM measurement is given by

$$\frac{1}{2} + \frac{1}{2}T(\phi,\psi).$$

Without using this formula: Why can this probability never be smaller than 1/2?

(b) Conclude that only orthogonal states (i.e., $\langle\phi|\psi\rangle = 0$) can be distinguished perfectly.

(c) Show that the trace distance is a metric. That is, $T(\phi,\psi) = 0$ if and only if $|\phi\rangle = e^{\imath\theta}|\psi\rangle$, $T(\phi,\psi) = T(\psi,\phi)$, and the triangle inequality $T(\phi,\psi) \le T(\phi,\chi) + T(\chi,\psi)$ holds.

You will now derive an explicit formula for the trace distance. For this, consider the Hermitian operator $\Delta := |\phi\rangle\langle\phi| - |\psi\rangle\langle\psi|$ and its spectral decomposition $\Delta = \sum_i \lambda_i |e_i\rangle\langle e_i|$.

(d) Show that the operator $Q = \sum_{\lambda_i>0} |e_i\rangle\langle e_i|$ achieves the maximum in (2.7). Deduce from this the following formulas for the trace distance:

$$T(\phi,\psi) = \sum_{\lambda_i>0} \lambda_i = \frac{1}{2}\sum_i |\lambda_i|.$$

(e) Conclude that the optimal probability of distinguishing the two states in (a) remains unchanged if we restrict to projective measurements.

In class, we will also use the *fidelity*, which for pure states is simply the overlap $|\langle\phi|\psi\rangle|$:

(f) Show that trace distance and fidelity are related by the following formula:

$$T(\phi,\psi) = \sqrt{1 - |\langle\phi|\psi\rangle|^2}.$$

Hint: Argue that it suffices to verify this formula in the qubit case and with one state equal to $|0\rangle$*. Then use the formula from part (d).*

Thus, states with fidelity close to one are almost indistinguishable by any measurement. Note that the trace distance is a distance measure (it is ≈ 0 if the states are close), while the fidelity is a similarity measure (it is ≈ 1 if the states are close).

2.5 **Ambiguity of POVMs:** Find two implementations of the same POVM, as in (2.5), that produce different post-measurement states.

Chapter 3

Quantum correlations, non-local games, rigidity

In the past two lectures, we discussed some of the nonclassical features of quantum mechanics. In particular, we explored superpositions (such as $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$), entanglement ($|\Psi\rangle_{AB} \neq |\psi\rangle_A \otimes |\phi\rangle_B$), and non-commuting observables ($[X, Y] \neq 0$), and how these features impose both challenges (e.g., non-orthogonal states cannot be distinguished perfectly) and opportunities (e.g., entanglement gives rise to superdense coding and teleportation).

Today, we will discuss another way of quantifying the distinction between classical and quantum mechanics, namely through the *correlations* predicted by these theories. A modern perspective of studying and comparing correlations is through the notions of a *nonlocal game*. This is closely related to Bell inequalities, which you may remember from your quantum mechanics class – but we will discuss some interesting new aspects that you may not have seen before.

3.1 The GHZ game

In a *nonlocal game*, we imagine that a number of collaborating *players* play against a *referee*. The referee hands them *questions* and the players reply with appropriate *answers* that win them the game. The players' goal is to maximize their chance of winning. Before the game starts, may agree upon a joint strategy – but then they are separated from each other and cannot communicate while the game is being played (this can be ensured by the laws of special relativity). The point then is the following: *Since the players are constrained by the laws of physics, we can design games where players utilizing a quantum strategy may have an advantage.* This way of reasoning about quantum correlations is eminently operational and quantitative, as we will see. There are many well-known such games in the literature.

Here will discuss the *Greenberger-Horne-Zeilinger (GHZ) game* [Mer90, GHSZ90], but we note that another well-known game is the two-player *Clauser–Horne–Shimony–Holt (CHSH) game*. The GHZ game involves three players – Alice, Bob, and Charlie. Each receives as question a bit $x, y, z \in \{0, 1\}$ and their answers are likewise bits $a, b, c \in \{0, 1\}$. They win the game if the sum of their answers modulo two is as in the following table:

x	y	z	$a \oplus b \oplus c$
0	0	0	0
1	1	0	1
1	0	1	1
0	1	1	1

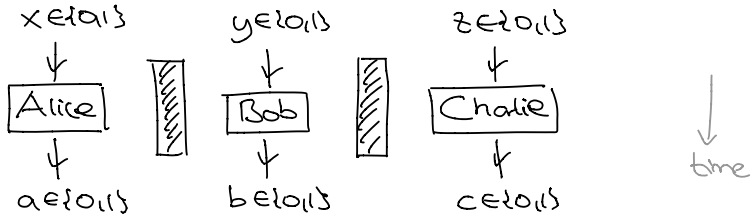


Figure 3.1: Setup of the three-player GHZ game. The winning condition is that $a \oplus b \oplus c = x \vee y \vee z$.

Note that the referee only asks four out of eight possible question triples xyz. The winning condition can be succinctly stated as follows: $a \oplus b \oplus c = x \vee y \vee z$. We write $\oplus$ for addition modulo 2 and $\vee$ for the logical OR. Figure 3.1 summarizes the setup.

3.2 Classical strategies

It is not hard to see that the GHZ game cannot be won if the players' strategies are described by a "local" and "realistic" theory. Here, "local" means that each player's answer only depends on their immediate surroundings, and "realistic" means that the strategy assigns a definite answer to any possible question (that is, in advance of the question being asked). Thus in a local and realistic theory we assume that

$$a = a(x), \quad b = b(y), \quad c = c(z).$$

When we say that the players may jointly agree on a strategy before the game is being played, we mean that they may select "question-answer functions" a, b, c in a correlated way. For example, when the players meet before the game is being played, they could toss a coin, resulting in some random $\lambda \in \{0, 1\}$, and agree on the strategy $a(x) = x \oplus \lambda$, $b(y) = y \oplus \lambda$, $c(z) = z \oplus \lambda$. Mathematically, this means that the functions a, b, c can be correlated random variables. This does not at all influence the argument that follows. Equivalently, we could say that λ is a "hidden variable", with some probability distribution p_λ, and consider $a = a(x, \lambda)$ as a deterministic function of both the input and the hidden variable. You can explore this in Exercise 3.3. If the players' strategy can be described by classical mechanics then the above would provide an adequate model. Thus, strategies of this form are usually referred to as *local hidden variable strategies* or simply as *classical strategies*.

Suppose now for sake of finding a contradiction that Alice, Bob, and Charlie can play the GHZ game perfectly using such a classical strategy. Then,

$$\begin{aligned} 1 &= 0 \oplus 1 \oplus 1 \oplus 1 \\ &= (a(0) \oplus b(0) \oplus c(0)) \oplus (a(1) \oplus b(1) \oplus c(0)) \oplus (a(1) \oplus b(0) \oplus c(1)) \oplus (a(0) \oplus b(1) \oplus c(1)) \\ &= 0. \end{aligned}$$

The first equality is plainly true, the second holds since we assumed that the strategy is perfect, and the last equality holds because $a(x) \oplus a(x) \equiv 0$ etc., whatever the value of $a(x)$. This is a contradiction! We conclude that there is no perfect classical winning strategy for the GHZ game.

This means that if the referee selects each possible question triple xyz with equal probability $1/4$, then the game can be won with probability at most

$$p_{\text{win,cl}} \leq 3/4, \tag{3.1}$$

since the players must get at least one of the four possible answers wrong (note that which one they get wrong might well be a random variable). This winning probability can be achieved by, e.g., the trivial strategy $a(x) = b(y) = c(z) \equiv 1$. Equation (3.1) can be called a *Bell inequality*. If you have seen this term before: do you see the connection?

3.3 Quantum strategies

In a *quantum strategy*, we imagine that the three players are described by quantum mechanics. Thus they start out by sharing an arbitrary joint state $|\psi\rangle_{ABC} \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, where $\mathcal{H}_A$ is the Hilbert space describing a quantum system in Alice's possession, etc., and upon receiving their questions $x, y, z \in \{0, 1\}$ they will each measure corresponding observables A_x, B_y, C_z on their respective Hilbert spaces. While it might not be immediately obvious, any classical strategy is also a quantum strategy. You can show this in Exercise 3.3.

It will be convenient to take the eigenvalues (i.e., measurement outcomes) of the observables to be in $\{\pm 1\}$ rather than in $\{0, 1\}$. That is, if Alice measure A_x and obtains outcome $(-1)^a$, she sends back answer a to the referee, and similarly for the other players. The condition that A_x has eigenvalues ± 1 can be succinctly stated as $A_x^2 = I$. In this case, the eigenvalues of the observable $A_x \otimes B_y \otimes C_z$ are $(-1)^{a+b+c} = (-1)^{a \oplus b \oplus c}$, and they correspond precisely to the sum modulo two of the answers. In particular, a *perfect* quantum strategy for the GHZ game in which the players win with 100% probability would consist of a quantum state $|\psi_{ABC}\rangle$ and observables $\{A_x\}$, $\{B_y\}$, $\{C_z\}$ such that $A_x^2 = B_y^2 = C_z^2 = I$ for all x, y, z and

$$\begin{aligned}
(A_0 \otimes B_0 \otimes C_0)\,|\psi_{ABC}\rangle &= +\,|\psi_{ABC}\rangle\,, \\
(A_1 \otimes B_1 \otimes C_0)\,|\psi_{ABC}\rangle &= -\,|\psi_{ABC}\rangle\,, \\
(A_1 \otimes B_0 \otimes C_1)\,|\psi_{ABC}\rangle &= -\,|\psi_{ABC}\rangle\,, \\
(A_0 \otimes B_1 \otimes C_1)\,|\psi_{ABC}\rangle &= -\,|\psi_{ABC}\rangle
\end{aligned} \tag{3.2}$$

(recall from Chapter 1 that an observable always give the same outcome precisely when the state is an eigenvector, with eigenvalue equal to that outcome). In Exercise 3.2 you can verify that, more generally, the probability of winning the GHZ game (for a uniformly random choice of questions xyz) can be written as:

$$\begin{aligned}
p_{\text{win},q} = \frac{1}{2} + \frac{1}{8} \langle\psi_{ABC}|\, A_0 \otimes B_0 \otimes C_0 - A_1 \otimes B_1 \otimes C_0 \\
- A_1 \otimes B_0 \otimes C_1 - A_0 \otimes B_1 \otimes C_1\, |\psi_{ABC}\rangle\,.
\end{aligned} \tag{3.3}$$

Remarkably, there is indeed a quantum strategy for the GHZ game that allows the players to win the game with probability $p_{\text{win,q}} = 1$. We assume that the players share the three-qubit state

$$|\Gamma_{ABC}\rangle = \frac{1}{2}\left(|000\rangle - |110\rangle - |101\rangle - |011\rangle\right) \in \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2, \tag{3.4}$$

where the first qubit is in Alice's possession, the second in Bob's, and the third in Charlie's (see also Exercise 3.1). Upon receiving $x = 0$, Alice measures the Pauli observable $A_0 = Z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ on her qubit, while upon receiving $x = 1$ she measures the Pauli observable $A_1 = X = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Bob and Charlie perform exactly the same strategy on their qubits. To see that this quantum strategy wins the GHZ game every single time, we only need to verify (3.2) for this strategy. Indeed:

$$(Z \otimes Z \otimes Z)\,|\Gamma_{ABC}\rangle = |\Gamma_{ABC}\rangle\,,$$

$$(X \otimes X \otimes Z) |\Gamma_{ABC}\rangle = \frac{1}{2} (|110\rangle - |000\rangle - (-1) |011\rangle - (-1) |101\rangle) = - |\Gamma_{ABC}\rangle ,$$

and similarly $(X \otimes Z \otimes X) |\Gamma_{ABC}\rangle = (Z \otimes X \otimes X) |\Gamma_{ABC}\rangle = - |\Gamma_{ABC}\rangle$.

This shows that in a precise quantitative sense, quantum mechanics enables stronger non-local correlations than what is possible using any local realistic classical theory.

A glance a device-independent quantum cryptography

When the three players perform the optimal strategy described above then not only do their answers satisfy the winning condition but their answers a, b, c are in fact *uniformly random*, subject only to the constraint that $a \oplus b \oplus c$ must sum to the desired value $x \vee y \vee z$. In particular, $a, b \in \{0, 1\}$ are two independent random bits. You can easily verify this by inspection. For example, for $x = y = z = 0$, Alice, Bob, and Charlie each measure their local Z observable. The eigenvectors are $|abc\rangle$ and so it is clear from Eq. (3.4) that we obtain $abc \in \{000, 110, 101, 011\}$ with equal probability $1/4$.

The randomness obtained in this way is also *private*. We will only discuss this in a very heuristic sense and you should be sceptical of the details, but I would still like to give you an impression. Suppose that apart from Alice, Bob, Charlie, there is also an evil eavesdropper (Evan) who would like to learn about the random bits generated in this way. Their joint state can be described by a pure state $|\psi_{ABCE}\rangle$. If Alice, Bob, and Charlie indeed share the state $|\Gamma_{ABC}\rangle$ then it must be the case that $|\psi_{ABCE}\rangle = |\Gamma_{ABC}\rangle \otimes |\psi\rangle_E$ (this holds not just for $|\Gamma_{ABC}\rangle$ but for any "pure" quantum state). We will see how to formalize this statement in Chapter 8. Thus, our three protagonists are in a product state with Evan. It follows that the bits a and b are not just random but also independent from the outcomes of any measurement that Evan can do on the E system (Exercise 3.4). All this means that the referee can use the players' answers to generate *private randomess*: by locking lock Alice, Bob, and Charlie (best thought of as quantum devices) into a laboratory, ensuring that the devices cannot communicate with each other and the outside world, and interrogating them with questions, as in the following picture:

But of course there is a catch: the referee cannot trust Alice, Bob, and Charlie to actually play the quantum strategy described above. So this observation might seem not very useful at first glance...*however, what if the optimal strategy for winning the GHZ game was actually unique?* In this case, the referee could *test* Alice, Bob, and Charlie with randomly selected questions and check that they pass the test every time. After a while, the referee become increasingly confident that the players are in fact able to win the GHZ game every time. But then, by uniqueness of the winning strategy, the referee will in fact know the precise strategy that Alice, Bob, and Charlie are pursuing! In other words, the referee would *not* have to put any trust into Alice, Bob, and Charlie – they would rather prove themselves by winning the GHZ game every time.

This remarkable idea for generating private random bits was proposed by Colbeck [Col09]. Note that we need private random bits in the first place to generate the random questions – thus this protocol proposes to achieve a task known as *randomness expansion*. Private random bits cannot be generated without an initial seed of random bits. The argument sketched so far is of course not rigorous at all: we do not take into account that Alice, Bob, Charlie may not behave the same way every time we play the game, that they may have a (quantum) memory, we ignored questions of robustness and finite statistics, etc. However, these challenges can be circumvented

and secure randomness expansion protocols using completely untrusted devices do exist [AM16]. This general line of research is known as *device-independent quantum cryptography*, since it does not rely on assumptions on the inner workings of the devices involved, but only on their observed correlations [MY98].

3.4 Rigidity of the GHZ game

In the remainder of the lecture, we will show that the perfect winning strategy for the GHZ game is indeed essentially unique, following Colbeck and Kent [CK11]. We say that the GHZ game is *rigid* or that it is a *self-test* for the three-qubit quantum strategy described above.

Let us first observe that in our three-qubit strategy, the state $|\Gamma_{ABC}\rangle$ is already uniquely determined by the measurement operators. This follows from Eq. (3.2), because any $+1$-eigenvector of $Z \otimes Z \otimes Z$ is necessarily of the form $\alpha |000\rangle + \beta |110\rangle + \gamma |101\rangle + \delta |011\rangle$, and the other three conditions are only satisfied if $\alpha = -\beta = -\gamma = -\delta$. Thus we obtain (3.4) up to an irrelevant overall phase.

Let us now consider a general quantum strategy given by a state $|\psi_{ABC}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ and observables A_x, B_y, C_z with $A_x^2 = I$, $B_y^2 = I$, and $C_z^2 = I$ such that Eq. (3.2) is satisfied. Our approach to proving the rigidity theorem will be to uncover some *hidden symmetries* that allow us to reduce to the case of three qubits:

Claim 3.1 (Informal)**.** *In any optimal strategy, the observables must anticommute: "*$\{A_0, A_1\} = 0$, $\{B_0, B_1\} = 0$, $\{C_0, C_1\} = 0$*" (see below for fine-print).*

We will prove this claim later, but let us see first see how anticommutativity allows us to identify three qubits on which the observables A_x act like the Pauli operators from our optimal quantum strategy.

How to find a qubit?

Consider, e.g., the pair of observables A_0, A_1. By assumption, they satisfy $A_0^2 = A_1^2 = I$ as well as $\{A_0, A_1\} = 0$. Since $A_0^2 = \pm I$, its eigenvalues are ± 1. If $|\phi\rangle$ be an eigenvector of A_0 with eigenvalue ± 1, i.e., $A_0 |\phi\rangle = \pm |\phi\rangle$, then

$$A_0 A_1 |\phi\rangle = -A_1 A_0 |\phi\rangle = -A_1(\pm 1 |\phi\rangle) = \mp A_1 \phi,$$

so $A_1 |\phi\rangle$ is an eigenvector of A_0 with eigenvalue ∓ 1. This means that the unitary A_1 interchanges the two eigenspaces of A_0. In particular, both must have the same dimension, which we shall denote by m_A. Moreover, if $\{|e_{0,j}\rangle\}_{j=1,\ldots,m_A}$ is an orthonormal basis of the $+1$-eigenspace then the vectors $|e_{1,j}\rangle := A_1 |e_{0,j}\rangle$ form an orthonormal basis of the -1-eigenspace. Thus, the unitary defined by

$$U_A \colon \mathcal{H}_A \to \mathbb{C}^2 \otimes \mathbb{C}^{m_A}, \quad |e_{i,j}\rangle \mapsto |i, j\rangle = |i\rangle \otimes |j\rangle .$$

maps A_0 and A_1 to the Pauli Z and X operators acting on the right-hand side qubit:

$$U A_0 U^\dagger = Z \otimes I, \quad U A_1 U^\dagger = X \otimes I.$$

Indeed,

$$\begin{aligned} U A_0 U^\dagger |i, j\rangle &= U A_0 |e_{i,j}\rangle = U(-1)^i |e_{i,j}\rangle = (-1)^i |i, j\rangle = (Z \otimes I) |i, j\rangle \\ U A_1 U^\dagger |i, j\rangle &= U A_1 |e_{i,j}\rangle = U |e_{i\oplus 1,j}\rangle = |i \oplus 1, j\rangle = (X \otimes I) |i, j\rangle \end{aligned}$$

To summarize: We found that $\mathcal{H}_A \cong \mathbb{C}^2 \otimes \mathbb{C}^{m_A}$ such that A_0, A_1 act by $Z \otimes I$, $X \otimes I$, respectively.

The same argument works for Bob and Charlie's pairs of observables. Thus the total Hilbert space decomposes as

$$\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C \cong \left(\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2\right) \otimes \left(\mathbb{C}^{m_A} \otimes \mathbb{C}^{m_B} \otimes \mathbb{C}^{m_C}\right)$$

and the measurement operators act as in the three-qubit solution on the first tensor factor. E.g.,

$$\begin{aligned} A_0 &\cong (Z \otimes I \otimes I) \otimes (I \otimes I \otimes I), \\ A_1 &\cong (X \otimes I \otimes I) \otimes (I \otimes I \otimes I), \end{aligned}$$

etc. We discussed above that in the three-qubit solution the state is uniquely determined by the measurement operators. Thus,

$$|\psi_{ABC}\rangle \cong |\Gamma\rangle \otimes |\gamma'\rangle\,,$$

where $|\Gamma\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ is the three-qubit state from Eq. (3.4) and $|\gamma'\rangle \in \mathbb{C}^{m_A} \otimes \mathbb{C}^{m_B} \otimes \mathbb{C}^{m_C}$ some auxiliary state (which is irrelevant because the observables do not act on it). This is the desired rigidity result.

Anticommutations from correlations (proof of claim 3.1)

We still need to prove Theorem 3.1. We first rewrite the optimality condition in Eq. (3.2) as

$$\begin{aligned} A_0 |\psi\rangle &= +B_0 C_0 |\psi\rangle \\ A_0 |\psi\rangle &= -B_1 C_1 |\psi\rangle \\ A_1 |\psi\rangle &= -B_1 C_0 |\psi\rangle \\ A_1 |\psi\rangle &= -B_0 C_1 |\psi\rangle\,. \end{aligned}$$

Above we used that $B_y^2 = I$ and $C_z^2 = I$, and we write A_0 instead of $A_0 \otimes I_B \otimes I_C$, etc., to make the formulas easier to read. From the first two and last two equations, respectively,

$$\begin{aligned} A_0 |\psi\rangle &= +\frac{1}{2}\left(B_0 C_0 - B_1 C_1\right)|\psi\rangle \\ A_1 |\psi\rangle &= -\frac{1}{2}\left(B_1 C_0 + B_0 C_1\right)|\psi\rangle \end{aligned}$$

Hence,

$$\begin{aligned} A_0 A_1 |\psi\rangle &= -\frac{1}{4}\left(B_1 C_0 + B_0 C_1\right)\left(B_0 C_0 - B_1 C_1\right)|\psi\rangle = -\frac{1}{4}\left(B_1 B_0 - C_0 C_1 + C_1 C_0 - B_0 B_1\right)|\psi\rangle\,, \\ A_1 A_0 |\psi\rangle &= -\frac{1}{4}\left(B_0 C_0 - B_1 C_1\right)\left(B_1 C_0 + B_0 C_1\right)|\psi\rangle = -\frac{1}{4}\left(B_0 B_1 - C_1 C_0 + C_0 C_1 - B_1 B_0\right)|\psi\rangle\,, \end{aligned}$$

where we used that A_x, B_y, and C_z operators commute pairwise (recall that these are just shorthand notation for $A_x \otimes I \otimes I$, $I \otimes B_y \otimes I$, and $I \otimes I \otimes C_z$). Note that the right-hand side is the same for both equations! Thus we have proved:

$$\{A_0, A_1\}|\psi\rangle = 0$$

This is almost what we wanted to show!

How can we show that $\{A_0, A_1\} = 0$? This is in fact not exactly true – hence the "quotes" in Theorem 3.1. But what is true is that $\{A_0, A_1\} = 0$ on a subspace $\tilde{\mathcal{H}}_A$ of $\mathcal{H}_A$ such that $|\psi\rangle_{ABC} \in \tilde{\mathcal{H}}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. Indeed, we can expand

$$|\psi\rangle_{ABC} = \sum_i s_i |e_i\rangle_A \otimes |f_i\rangle_{BC}$$

where the $|e_i\rangle$ and $|f_i\rangle$ are orthonormal and $s_i > 0$ – this is called the *Schmidt decomposition* and we will discuss it in more detail in **MW: ref**. If there are $\dim \mathcal{H}_A$ terms then the $|e_i\rangle$ form a basis of $\mathcal{H}_A$ and so $\{A_0, A_1\}|\psi\rangle = 0$ implies that $\{A_0, A_1\} = 0$. Otherwise, we can restrict to the subspace $\tilde{\mathcal{H}}_A := \mathrm{span}\{|e_i\rangle_A\}$. In the latter case, $|\psi\rangle_{ABC} \in \tilde{\mathcal{H}}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, the operators A_x are block diagonal with respect to $\tilde{\mathcal{H}}_A \oplus \tilde{\mathcal{H}}_a^{\perp}$, and $\{A_0, A_1\} = 0$ on $\tilde{\mathcal{H}}_A$. We can proceed likewise for B_y and C_z.

Statement of the rigidity theorem

What have we proved? In mathematical terms, we have established the following theorem:

Theorem 3.2 (Rigidity for the GHZ game)**.** *Consider any perfect quantum strategy for the GHZ game given by a state $|\psi_{ABC}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ and ± 1-valued observables $\{A_x\}$, $\{B_y\}$, $\{C_z\}$. Then there are isometries $V_A \colon \mathbb{C}^2 \otimes \mathbb{C}^{m_A} \to \mathcal{H}_A$, $V_B \colon \mathbb{C}^2 \otimes \mathbb{C}^{m_B} \to \mathcal{H}_B$, $V_C \colon \mathbb{C}^2 \otimes \mathbb{C}^{m_C} \to \mathcal{H}_C$, for suitable $m_A, m_B, m_C \in \mathbb{N}$, and a state $|\gamma\rangle \in \mathbb{C}^{m_A} \otimes \mathbb{C}^{m_B} \otimes \mathbb{C}^{m_C}$ such that*

$$|\psi\rangle_{ABC} = (V_A \otimes V_B \otimes V_C)(|\Gamma\rangle \otimes |\gamma\rangle)$$

and

$$\begin{aligned} V_A^\dagger A_0 V_A &= Z \otimes I, & V_A^\dagger A_1 V_A &= X \otimes I, \\ V_B^\dagger B_0 V_B &= Z \otimes I, & V_B^\dagger B_1 V_B &= X \otimes I, \\ V_C^\dagger C_0 V_C &= Z \otimes I, & V_C^\dagger C_1 V_C &= X \otimes I. \end{aligned}$$

In the coming weeks, we will revisit the techniques used above in a more systematic way. At the end of the term you will be well equipped to write up a fully formal proof of Theorem 3.2.

Outlook

There are many interesting aspects of nonlocal games beyond what we discussed in this lecture. For example, is the rigidity theorem robust, in the sense that if players win the GHZ game with probability close to 100% then their strategy must be "close" to the three-qubit strategy in some suitable sense? And how do the results that we discussed generalize to a setting where one plays many repetitions of a game – in multiple rounds (sequentially) or even at the same time (in parallel)? It is clear that if p is the optimal winning probability for a single instance then for n repetitions the winning probability is at least p^n , but perhaps one can do better by using strategies that exploit correlations or entanglement in a clever way? All this is also related to remarkable progress in quantum complexity theory, where multi-prover interactive protocols play an important role, for example on the topics of delegating and verifying quantum computations (see, e.g., [Vid20]).

Exercises

3.1 **GHZ state:** Find unitaries U_A, U_B, U_C such that

$$(U_A \otimes U_B \otimes U_C)\,|\Gamma_{ABC}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle).$$

Conclude that one can also win the GHZ game by using the right-hand side state, which is known as the *GHZ state*. Why is this no contradiction to the rigidity theorem?

3.2 **Quantum strategies:** Verify that the winning probability of a general quantum strategy for the GHZ game, specified by a state $|\psi_{ABC}\rangle$ and observables A_x, B_y, C_z, is given by Eq. (3.3).

3.3 **Classical strategies:** When they meet before the game is started, they toss a biased coin. Let π denote the probability that the coin comes up heads. Depending on the outcome of the coin toss, which we denote by $\lambda \in \{\text{HEADS}, \text{TAILS}\}$, they use one of two possible deterministic strategies $a_\lambda(x), b_\lambda(y), c_\lambda(z)$ to play the game.

(a) Suppose that Alice, Bob, and Charlie play the following randomized classical strategy: Find a formula analogous to Eq. (3.3) for the winning probability $p_{\text{win,cl}}$ of their strategy.
(b) In class we argued that the bound $p_{\text{win,cl}} \leq 3/4$ also applies to randomized classical strategies. Verify this explicitly for the strategy described above by using the formula you derived in (a).
(c) Any classical strategy can be realized by a quantum strategy. Show this explicitly for the randomized classical strategy described above, by constructing a quantum state $|\psi_{ABC}\rangle$ and observables A_x, B_y, C_z such that $p_{\text{win,cl}} = p_{\text{win,q}}$.

3.4 **Product states yield independent measurement outcomes:** Suppose that Alice and Bob share a quantum state $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. Alice performs a POVM measurement $\{Q_{A,x}\}$ and Bob a POVM measurement $\{R_{B,y}\}$, so the joint probability of their outcomes is given by

$$p(x, y) = \langle \Psi_{AB} | Q_{A,x} \otimes R_{B,y} | \Psi_{AB} \rangle \,.$$

Show that if $|\Psi_{AB}\rangle$ is a product state then the measurement outcomes of Alice and Bob are independent random variables.

3.5 **Symmetries of ebit and singlet:** Recall the ebit state $|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)$.

(a) Show that $(M \otimes I) |\Phi^+_{AB}\rangle = (I \otimes M^T) |\Phi^+_{AB}\rangle$ for every operator M.
(b) Deduce that $(U \otimes \bar{U}) |\Phi^+_{AB}\rangle = |\Phi^+_{AB}\rangle$ for every unitary U.

Now consider the *singlet* $|\Psi^-_{AB}\rangle := \frac{1}{\sqrt{2}} (|10\rangle - |01\rangle)$, which is another of the Bell states in Eq. (2.2).

(a) Show that $(M \otimes M) |\Psi^-_{AB}\rangle = \det(M) |\Psi^-_{AB}\rangle$ for every operator M.

Chapter 4

Pure state estimation, symmetric subspace

Suppose we are given a quantum system and we would like to learn about its quantum state. Is there a measurement that gives us a classical description "ψ" of the state $|\psi\rangle$? Clearly, this cannot be done perfectly – for otherwise we could distinguish non-orthogonal states (by comparing their classical descriptions), which we already know to be impossible.

How about if we are given not just one copy of a state, but in fact many copies $|\psi\rangle^{\otimes n}$? Note that if $\psi \neq \phi$ are any two distinct states, whether orthogonal or not, then $|\langle\psi|\phi\rangle| < 1$ and thus

$$\left(\langle\psi|^{\otimes n}\right)\left(|\phi\rangle^{\otimes n}\right) = \langle\psi^{\otimes n}|\phi^{\otimes n}\rangle = \langle\psi|\phi\rangle^n \to 0,$$

suggesting that we may be able to distinguish them arbitrarily well in the limit of $n \to \infty$ copies. Of course, since $|\langle\psi|\phi\rangle|$ can be arbitrarily close to one, we have to be careful. But when $|\langle\psi|\phi\rangle| \approx 1$ then the two states are almost indistinguishable by any measurement (Exercise 2.4), and so we make only a small error by conflating them. In the above discussion it is good to recall from Chapter 1 that two vectors $|\psi\rangle$ and $e^{i\eta}|\psi\rangle$ that only differ by an overall phase should be really thought of as the *same* quantum state (they cannot be distinguished by any procedure). We also discussed that a good way of getting rid of this ambiguity is by considering the projector $\psi = |\psi\rangle\langle\psi|$ in place of $|\psi\rangle$. We will always use this convention: if $|\psi\rangle$ is a unit vector then ψ refers to the corresponding projector. We call ψ or $|\psi\rangle$ a *pure quantum state*.[1]

Given the preceding discussion, it is plausible that we can achieve the following task (Fig. 4.1):

Problem 4.1 (Pure state estimation)**.** Find a POVM $\{Q_{\hat{\phi}}\}_{\hat{\phi}\in\Omega}$ on $(\mathbb{C}^d)^{\otimes n}$, with possible outcomes in the set $\Omega = \{\hat{\phi} = |\hat{\phi}\rangle\langle\hat{\phi}|\}$ of pure state on $\mathbb{C}^d$, such that when we measure $\phi^{\otimes n} = |\phi\rangle^{\otimes n}\langle\phi|^{\otimes n}$, we obtain an outcome $\hat{\phi}$ that is "close" to ϕ (on average, or even with high probability).

We will quantify "closeness" using (the square of) the *fidelity* $|\langle\hat{\phi}|\phi\rangle|$, which you know from Exercise 2.4. Of course how well we can do will depend on the number n of copies of the state that we are given (the more the easier) and the Hilbert space dimension d (the larger the harder).

[1] The name suggests that there also exist a more general notion of a quantum state. Indeed we have already seen the need for this. For example, if two qubits are in the ebit state, then states of the individual qubits *cannot* be described by pure states (i.e., vectors in $\mathbb{C}^2$). Can you see how this follows from Eq. (2.1)? Next week, in Chapter 7, we will introduce a more general notion of a quantum state which allows us to model this situation. These are the so-called *non-pure* or *mixed* quantum states. In turn, such mixed states can always be described in terms of subsystems of a larger quantum system that is in a pure state. Thus the situation is completely parallel to the case of measurements, where we identified a larger class of measurements (POVMs) which could nevertheless be reduced to ordinary projective measurements on a larger system. But for today we focus on the important case of pure states. Later in this course we will learn how to solve the state estimation or "tomography" problem for general (i.e., not necessarily pure) quantum states (Chapter 14).

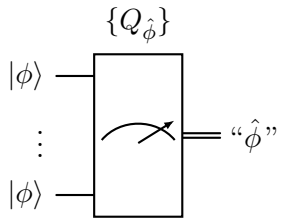


Figure 4.1: Illustration of pure state estimation (Theorem 4.1). We put the outcome "$\hat{\phi}$" in quotes to emphasize that it is the classical description of a quantum state.

4.1 Continuous POVMs and uniform measure

In the statement of the pure state estimation problem, we are faced with a formal difficulty. The set of outcomes Ω is infinite (even continuously so), but so far we have only discussed POVMs with a finite number of possible outcomes. How can we generalize the concept of a POVM to an infinite set of outcomes Ω (e.g., the set of all pure states, or the set of all real numbers $\mathbb{R}$, ...)?

For simplicity, let us assume that the space of outcomes Ω carries some natural measure dx. (E.g., if $\Omega = \mathbb{R}$, we could choose the Lebesgue measure.) If w replace the $\sum$ by this measure, we arrive at the following definition. A collection of operators $\{Q_x\}_{x\in\Omega}$ is a *(continuous) POVM* with outcomes in the measure space Ω if

(a) $Q_x \geq 0$ for all $x \in \Omega$, as before, and
(b) $\int_\Omega dx\, Q_x = I$.

Moreover, $x \mapsto Q_x$ must be a measurable function. (We will always consider Borel measures, so that measurability is ensured by continuity.) The corresponding version of the *Born rule* states that the probability *density* of the outcomes, with respect to the measure dx, is given by

$$p_\psi(x) = \langle\psi|Q_x|\psi\rangle\,. \tag{4.1}$$

Thus, probabilities and expectation values can be computed as follows:

$$\begin{aligned}\mathbf{Pr}_\psi(\text{outcome} \in S) &= \int_S dx\ \langle\psi|Q_x|\psi\rangle\,,\\ \mathbf{E}_\psi\left[f(x)\right] &= \int_\Omega dx\ \langle\psi|Q_x|\psi\rangle\, f(x).\end{aligned} \tag{4.2}$$

Remark 4.2. Given the above, we can assign to any (measurable) subset $X \subseteq \Omega$ an operator $Q(X) := \int_X dx\, Q_x$. Then it holds that (i) $Q(X) \geq 0$, (ii) $Q(\emptyset) = 0$, and (iii) $Q(\bigcup_k X_k) = \sum_k Q(X_k)$ for any collection (X_k) of disjoint subsets of Ω. Thus, Q behaves just like a measure – except that each $Q(X)$ is a positive semidefinite operator rather than a nonnegative number. This explains the term "positive semidefinite operator-valued measure (POVM)".

Note also that POVMs with finitely many outcomes as discussed in Section 2.4 are a special case of the above setup. Indeed, if Ω is discrete then we can always choose dx to be the *counting measure*, which assigns to any subset $S \subseteq \Omega$ its cardinality. Then, $\int dx = \sum_x$ and so we recognize the postulates from Section 2.4.

Just like in the discrete case, any continuous POVM is physical in the sense that it can be implemented using the laws of quantum mechanics. You might be concerned whether we need infinite-dimensional Hilbert spaces in order to implement continuous POVMs. This is not so: any continuous POVM on a finite-dimensional Hilbert space can be implemented in the following

fashion: (i) sample λ from some suitable probability distribution, and (ii) measure a finite POVM depending on λ. For the details, see [CDS07].

Returning to Theorem 4.1, we still need to specify which measure $d\phi$ we would like to put on the set of pure states. One desirable property is certainly that the measure should treat all quantum states the same. That is, if we substitute $|\psi\rangle \mapsto U|\psi\rangle$ then we would like all expectation values to remain unchanged:

$$\int d\psi\, f(\psi) = \int d\psi\, f(U\psi U^\dagger) \tag{4.3}$$

for any integrable function f and any unitary $d \times d$-matrix $U \in \mathrm{U}(d)$. One can show that there exists a *unique* probability measure $d\psi$ that is *unitarily invariant* in this sense. It is called the *uniform (probability) measure* on the set of pure quantum states. Sometimes, it is also referred to as the *Haar measure*, because it is induced by the Haar probability measure of the unitary group.

Remark 4.3. Here are three other measures that are similarly determined by their symmetries:

- For any finite set S, there exists a unique probability measure that is invariant under relabeling (permuting) the elements of S: the uniform probability distribution on S.
- There exists a unique measure on $\mathbb{R}$ that assigns unit measure to the unit interval $[0, 1]$, and which is invariant under arbitrary translations $x \mapsto x + a$: the Lebesgue measure.
- There exists a unique probability measure on the unit sphere S^2 that is invariant under rotations in $\mathrm{SO}(3)$. The same is true for higher-dimensional unit spheres.

We can think of the set of pure states as the unit sphere of $\mathbb{C}^d$ modulo overall phases in $\mathrm{U}(1)$. Thus the third example makes it quite plausible that the uniform measure $d\psi$ on the set of pure states should exist. In fact, the pure states of a qubit can be exactly identified with S^2. This is known as the Bloch sphere, see Exercise 7.3.

4.2 Symmetric subspace

In order to come up with a good POVM for estimating pure states, let us analyze the *symmetries* inherent in this problem. A first observation is that n copies of a quantum state $|\phi\rangle \in \mathbb{C}^d$ is described by

$$|\phi\rangle^{\otimes n} = |\phi\rangle \otimes \cdots \otimes |\phi\rangle \in (\mathbb{C}^d)^{\otimes n},$$

which is a state that is invariant under permuting the subsystems (copies).

To make this more precise, we can use the *symmetric group* on n symbols, which is denoted S_n. Its elements are permutations $\pi \colon \{1, \ldots, n\} \to \{1, \ldots, n\}$. Thus, S_n has $n!$ elements. This is indeed a *group*, meaning that products and inverses are again contained in S_n. Moreover, for any permutation $\pi \in S_n$, we can define a corresponding operator R_π on the Hilbert space $(\mathbb{C}^d)^{\otimes n}$:

$$R_\pi \colon (\mathbb{C}^d)^{\otimes n} \to (\mathbb{C}^d)^{\otimes n}, \quad R_\pi |\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle = |\psi_{\pi^{-1}(1)}\rangle \otimes \ldots \otimes |\psi_{\pi^{-1}(n)}\rangle \tag{4.4}$$

Then it holds that

$$R_1 = I \quad \text{and} \quad R_\tau R_\pi = R_{\tau\pi}. \tag{4.5}$$

Indeed, the latter is guaranteed by our judicious use of inverses in Eq. (4.4):

$$\begin{aligned} &R_\tau R_\pi |\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle = R_\tau |\psi_{\pi^{-1}(1)}\rangle \otimes \ldots \otimes |\psi_{\pi^{-1}(n)}\rangle \\ &= R_\tau |\psi_{\pi^{-1}(1)}\rangle \otimes \ldots \otimes |\psi_{\pi^{-1}(n)}\rangle \end{aligned}$$

$$= |\psi_{\pi^{-1}(\tau^{-1}(1))}\rangle \otimes \ldots \otimes |\psi_{\pi^{-1}(\tau^{-1}(n))}\rangle$$
$$= |\psi_{(\tau\pi)^{-1}(1)}\rangle \otimes \ldots \otimes |\psi_{(\tau\pi)^{-1}(n)}\rangle$$
$$= R_{\tau\pi} |\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle .$$

Note thatEq. (4.5) also implies that $R_\pi^{-1} = R_{\pi^{-1}}$. An assignment $\pi \mapsto R_\pi$ that satisfies Eq. (4.5) is called a *representation*. Thus, the above defines a representation of the symmetric group S_n on the vector space $(\mathbb{C}^d)^{\otimes n}$. In fact, it is a *unitary representation*, which means that the operators R_π are all unitary: $R_\pi^\dagger = R_\pi^{-1}$. Next week, we will more formally introduce the machinery of group and representation theory, but today we would like to see why it is useful.

Let us return to the states $|\phi\rangle^{\otimes n}$. Clearly, they have the property that $R_\pi |\phi\rangle^{\otimes n} = |\phi\rangle^{\otimes n}$ for all $\pi \in S_n$. Thus, the vectors $|\phi\rangle^{\otimes n}$ are elements of the so-called *symmetric subspace*

$$\mathrm{Sym}^n(\mathbb{C}^d) = \left\{ |\Phi\rangle \in (\mathbb{C}^d)^{\otimes n} : R_\pi |\Phi\rangle = |\Phi\rangle \right\}.$$

In physics this is also known as the n-particle sector of the d-mode bosonic Fock space. Given an arbitrary vector $|\Psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, we can always symmetrize it to obtain a vector in the symmetric subspace. To this end we define the *symmetrizer*:

$$\Pi_n = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi \tag{4.6}$$

Lemma 4.4. *The operator Π_n is the orthogonal projection onto $\mathrm{Sym}^n(\mathbb{C}^d) \subseteq (\mathbb{C}^d)^{\otimes n}$.*

Proof. It suffices to verify the following three properties:

(a) If $|\Psi\rangle$ is in the symmetric subspace then $\Pi_n |\Psi\rangle = |\Psi\rangle$:

$$\Pi_n |\Psi\rangle = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi |\Psi\rangle = \frac{1}{n!} \sum_{\pi \in S_n} |\Psi\rangle = |\Psi\rangle .$$

(b) For any vector $|\Psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, the vector $|\Phi\rangle := \Pi_n |\Psi\rangle$ is in the symmetric subspace:

$$R_\tau |\Phi\rangle = R_\tau \Pi_n |\Psi\rangle = R_\tau \frac{1}{n!} \sum_{\pi \in S_n} R_\pi |\Psi\rangle = \frac{1}{n!} \sum_{\pi \in S_n} R_{\tau\pi} |\Psi\rangle = \frac{1}{n!} \sum_{\pi' \in S_n} R_{\pi'} |\Psi\rangle = \Pi_n |\Psi\rangle = |\Phi\rangle .$$

Here we used that $\pi \mapsto \pi' = \tau\pi$ is a bijection.

(c) The operator Π_n is Hermitian:

$$\Pi_n^\dagger = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi^\dagger = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi^{-1} = \frac{1}{n!} \sum_{\pi \in S_n} R_{\pi^{-1}} = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi = \Pi_n .$$

Here we used that $\pi \mapsto \pi^{-1}$ is a bijection. □

In particular, we can obtain a basis of the symmetric subspace as follows: Take the standard basis $|i\rangle$ of $\mathbb{C}^d$, consider the corresponding product basis $|i_1, \ldots, i_n\rangle$ on $(\mathbb{C}^d)^{\otimes n}$, and apply the symmetrizer. The resulting vectors do not depend on the order of the indices $i_1, \ldots, i_n$, but only on the number of times

$$t_i = \#\{i_k = i\}$$

each index i appears. The t_i's are called *occupation numbers* and the tuple $(t_0, \ldots, t_{d-1})$ is called a *type*. Note that we have $t_i \geq 0$ and $\sum_{i=0}^{d-1} t_i = n$. For distinct types we obtain orthogonal

vectors, and together these vectors span the symmetric subspace. If we normalize them, we obtain the *occupation number basis* of $\mathrm{Sym}^n(\mathbb{C}^d)$:

$$\begin{aligned} \|t_0, \ldots, t_{d-1}\rangle\!\rangle &:= \sqrt{\frac{n!}{t_0! \cdots t_{d-1}!}}\, \Pi_n \big(|0\rangle^{\otimes t_0} \otimes \cdots \otimes |d-1\rangle^{\otimes t_{d-1}} \big) \\ &= \sqrt{\frac{n!}{t_0! \cdots t_{d-1}!}}\, \Pi_n \, |\underbrace{0, \ldots, 0}_{t_0 \text{ times}}, \ldots, \underbrace{d-1, \ldots, d-1}_{t_{d-1} \text{ times}}\rangle \end{aligned} \tag{4.7}$$

In particular, the dimension of the symmetric subspace is equal to the number of types:

$$\dim \mathrm{Sym}^n(\mathbb{C}^d) = \operatorname{tr} \Pi_n = \binom{n+d-1}{n} = \frac{(n+d-1)!}{n!\,(d-1)!} \tag{4.8}$$

In mathematics, types are also known as *weights* and the basis is called a *weight basis*.

Example 4.5. In the case of $\mathrm{Sym}^2(\mathbb{C}^2)$, there are three types: $(2,0)$, $(1,1)$, and $(0,2)$. The corresponding occupation number basis consists of:

$$\|2,0\rangle\!\rangle = |00\rangle\,, \quad \|1,1\rangle\!\rangle = \frac{1}{\sqrt{2}}\,(|01\rangle + |10\rangle)\,, \quad \|0,2\rangle\!\rangle = |11\rangle\,.$$

Note that we can complete this to a basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$ by adding the singlet state $(|10\rangle - |01\rangle)/\sqrt{2}$, which is *antisymmetric*. More generally, it is true that $(\mathbb{C}^d)^{\otimes 2} = \mathrm{Sym}^2(\mathbb{C}^d) \oplus \bigwedge^2(\mathbb{C}^d)$, where $\bigwedge^n(\mathbb{C}^d)$ denotes the *antisymmetric subspace*, which consists of the vectors $|\Phi\rangle \in (\mathbb{C}^d)^{\otimes n}$ that pick up a minus sign for any transposition (swap) in S_n.

A resolution of the identity for the symmetric subspace

We studied the symmetric subspace because is contains the states $|\phi\rangle^{\otimes n}$ that describe the input in our pure state estimation problem. Now, not every vector in $\mathrm{Sym}^n(\mathbb{C}^d)$ is of this form – for example, $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ isn't. Moreover, the states $|\phi\rangle^{\otimes n}$ are not orthogonal. Nevertheless, there is a very useful formula for projection onto the symmetric subspace in terms of these states:

$$\Pi_n = \binom{n+d-1}{n} \int d\phi\; |\phi\rangle^{\otimes n} \langle\phi|^{\otimes n}\,, \tag{4.9}$$

where the measure $d\phi$ is the uniform probability distribution on the set of pure states that we discussed towards the end of Section 4.1. We will prove this formula in Section 5.4 by using representation theory, see Theorem 5.7.

One way of interpreting the formula is that the vectors $|\psi\rangle^{\otimes n}$ form an "overcomplete basis" of the symmetric subspace. To see what this means, take an arbitrary vector $|\Phi\rangle \in \mathrm{Sym}^n(\mathbb{C}^d)$. Then, using Eq. (4.9), we find that

$$|\Phi\rangle = \Pi_n |\Phi\rangle = \binom{d+n-1}{n} \int d\phi\; |\phi\rangle^{\otimes n} \langle\phi^{\otimes n}|\Phi\rangle = \int d\phi\, c_\phi(\Phi)\; |\phi\rangle^{\otimes n}\,,$$

where $c_\phi(\Phi) = \binom{d+n-1}{n} \langle\phi^{\otimes n}|\Phi\rangle$. This shows that any vector in the symmetric subspace can be written as a linear combination of the vectors $|\psi\rangle^{\otimes n}$. See also Theorem 13.6.

Another way to interpret Eq. (4.9) is that it shows that

$$Q_{\hat\phi} = \binom{d+n-1}{n} |\hat\phi\rangle^{\otimes n} \langle\hat\phi|^{\otimes n} \tag{4.10}$$

defines a continuous POVM on the symmetric subspace, with outcomes in the set of pure states! Indeed, we clearly have $Q_{\hat\phi} \geq 0$ and Eq. (4.9) precisely asserts that $\int d\hat\phi\, Q_{\hat\phi} = \Pi_n$. The POVM $\{Q_{\hat\phi}\}_{\hat\phi \in \Omega}$ is called the *uniform POVM* and will now use it to solve the pure-state estimation problem.

4.3 Pure state estimation

Recall the goal of pure state estimation: we are given n copies of some arbitrary unknown quantum state $|\phi\rangle$, and we want to obtain an estimate $\hat{\phi}$ of the pure state $\phi = |\phi\rangle\langle\phi|$. We will now show that the uniform POVM defined in Eq. (4.10) gives us a good estimate, following [Chi10, BCHW16, Har13]. As discussed, we will consider the fidelity squared, $|\langle\phi|\hat{\phi}\rangle|^2$ between estimate and true state.

Thus, suppose we are in state $\phi^{\otimes n}$ and suppose that we measure the unform POVM $\{Q_{\hat{\phi}}\}$. Then then the expected value of $|\langle\phi|\hat{\phi}\rangle|^{2k}$ (which is a random quantity because the measurement outcome $\hat{\phi}$ is random) can be computed as follows using Eq. (4.2):

$$\begin{aligned}
\mathbf{E}\Big[|\langle\phi|\hat{\phi}\rangle|^2\Big] &= \int d\hat{\phi}\ \langle\phi^{\otimes n}|Q_{\hat{\phi}}|\phi^{\otimes n}\rangle\ |\langle\phi|\hat{\phi}\rangle|^2 \\
&= \binom{n+d-1}{n} \int d\hat{\phi}\ |\langle\phi|\hat{\phi}\rangle|^{2(n+1)} \\
&= \binom{n+d-1}{n} \int d\hat{\phi}\ \langle\phi|^{\otimes(n+1)}\,|\hat{\phi}\rangle^{\otimes(n+1)}\,\langle\hat{\phi}|^{\otimes(n+1)}\,|\phi\rangle^{\otimes(n+1)} \\
&= \binom{n+d-1}{n} \langle\phi^{\otimes(n+1)}|\left(\int d\hat{\phi}\,|\hat{\phi}\rangle^{\otimes(n+1)}\,\langle\hat{\phi}|^{\otimes(n+1)}\right)|\phi^{\otimes(n+1)}\rangle \\
&= \binom{n+d-1}{n}\binom{n+d}{n+1}^{-1} \langle\phi^{\otimes(n+1)}|\Pi_{n+1}|\phi^{\otimes(n+1)}\rangle \\
&= \binom{n+d-1}{n}\binom{n+d}{n+1}^{-1} = \frac{n+1}{n+d} = 1 - \frac{d-1}{n+d} \geq 1 - \frac{d}{n}.
\end{aligned} \tag{4.11}$$

The second equality follows from the definition of the POVM elements $Q_{\hat{\phi}}$ in Eq. (4.10). To get the fifth equality, use formula for the projector onto the symmetric subspace $\mathrm{Sym}^{n+1}(\mathbb{C}^d)$. The rest are some simple algebraic manipulations and inequalities that I explained in class.

Success! We have shown that the uniform POVM (4.10) gives us a good estimate of the unknown pure state ϕ as soon as $n \gg d$. We can also turn the above result into a statement about the trace distance $T(\phi, \hat{\phi})$, which was introduced in Exercise 2.4. Using the relation between fidelity and trace distance that you proved in part (f) of this exercise, it follows that the uniform POVM achieves an average error as quantified by the trace distance of

$$\mathbf{E}\Big[T(\phi,\hat{\phi})\Big] = \mathbf{E}\left[\sqrt{1-|\langle\phi|\hat{\phi}\rangle|^2}\right] \leq \sqrt{\mathbf{E}\Big[1-|\langle\phi|\hat{\phi}\rangle|^2\Big]} \leq \sqrt{\frac{d}{n}}.$$

The first inequality is Jensen's inequality for the concave square root function, and the second inequality is Eq. (4.11). This result is quite intuitive: On the one hand, ϕ has $O(d)$ degrees of freedoms (more precisely, $2(d-1)$ real degrees of freedom, if we fix the norm to one and ignore the phase), so we might expect that $n = \Theta(d)$ copies are necessary and hopefully also sufficient if our goal is to estimate ϕ to constant precision. On the other hand, if the dimension is fixed then we might expect that using n copies we can estimate each component to precision $O(1/\sqrt{n})$ (in fact, this should be true even using a more naive procedure that measures one copy of ϕ at a time). The bound that we derived agrees with both intuitions.

Exercises

4.1 **Higher moments:** Here you can generalize the proof given above from $|\langle\phi|\hat{\phi}\rangle|^2$ to $|\langle\phi|\hat{\phi}\rangle|^{2k}$. This is a more stringent similarity measure, because $|\langle\phi|\hat{\phi}\rangle|^{2k} < |\langle\phi|\hat{\phi}\rangle|^2$ unless $\phi = \hat{\phi}$.

(a) Show that for any $n, k, d \in \mathbb{N}$, it holds that $\binom{n+d-1}{n}\binom{n+k+d-1}{n+k}^{-1} \geq 1 - \frac{kd}{n}$.
(b) Generalize the proof given above to show that if we measure the uniform POVM $\{Q_{\hat{\phi}}\}$ in the state $\phi^{\otimes n}$, then the expected value of $|\langle\phi|\hat{\phi}\rangle|^{2k}$ is at least $1 - \frac{kd}{n}$.

Chapter 5

Introduction to representation theory, Schur's lemma

Last time we discussed the problem of estimating an unknown pure state $\phi = |\phi\rangle\langle\phi|$ given n copies of it, i.e., $\phi^{\otimes n} = |\phi\rangle^{\otimes n} \langle\phi|^{\otimes n}$. We approached this by focusing on the *permutation symmetry* of the input state $|\phi\rangle^{\otimes n}$, which means that it is an element of the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^d)$, and we discussed that the set of all such states forms an 'overcomplete basis' of this subspace. By the latter we meant more formally that we have the following formula (Eq. (4.9)):

$$\Pi_n = \underbrace{\binom{n+d-1}{n} \int d\phi \, |\phi\rangle^{\otimes n} \langle\phi|^{\otimes n}}_{=:\Pi'_n}. \tag{5.1}$$

We used this formula to show that the *uniform POVM* $\{Q_{\hat{\phi}} := \binom{n+d-1}{n} |\hat{\phi}\rangle^{\otimes n} \langle\hat{\phi}|^{\otimes n}\}$ gives a good solution to the quantum state estimation problem. Yet, we still need to prove Eq. (5.1). One way of going about this would be to explicitly perform the integration. See, e.g., [Har13] for this approach. We will proceed differently and show that the symmetries of the left- and right-hand side expressions alone imply that the equality $\Pi_n = \Pi'_n$ must hold.

What are these symmetries? Of course, both operators are invariant under permutations, but in fact there is an additional symmetry that we have not yet discussed: the two operators also commute with $U^{\otimes n}$, for any $d \times d$ unitary matrix U. This can be seen rather directly from the definitions. For the left-hand side, which we recall was defined as $\Pi_n = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi$ in Eq. (4.6), this follows because we have $[U^{\otimes n}, R_\pi] = 0$ (we will prove this intuitive fact below). For the right-hand side, which we denoted by Π'_n, we can use the fact (see Eq. (4.3)) that the integral is invariant under substituting $|\phi\rangle \mapsto U|\phi\rangle$ or, equivalently, $\phi \mapsto U\phi U^\dagger$, to obtain

$$\begin{aligned} U^{\otimes n} \Pi'_n (U^\dagger)^{\otimes n} &= \binom{n+d-1}{n} \int d\phi \, (U|\phi\rangle)^{\otimes n} (\langle\phi| U^\dagger)^{\otimes n} \\ &= \binom{n+d-1}{n} \int d\phi \, |\phi\rangle^{\otimes n} \langle\phi|^{\otimes n} = \Pi'_n. \end{aligned} \tag{5.2}$$

To see that this symmetry indeed suffices will take us some work, since we first have to develop the required mathematics. But this time will be well-invested, since we will be able to leverage the techniques that we will learn throughout the remainder of this course!

A good reference for the material that follows is Part 1 of the textbook [Ser77].

5.1 Groups and representations

Recall that a *group* G is given by a set together with an (associative) multiplication operation (sometimes denoted $\cdot$ but mostly omitted), an identity element (denoted 1 unless there is a more concrete notation), and inverses (denoted g^{-1}).

For example, the *symmetric group* S_n consists of all permutations of the set $\{1, \dots, n\}$. That is, its elements are the bijective functions from this set to itself. The multiplication law is given by the composition of functions, i.e., given two permutations π and and τ, we define $\pi\tau$ by $(\pi\tau)(x) := \pi(\tau(x))$ for $x \in \{1, \dots, n\}$. The identity element is the identity map, and inverses are given by the usual inverse of functions. We already know this group from Section 4.2. Any permutation can be written as a product of so-called *transpositions* $x \leftrightarrow y$. These are the permutations that swap two elements $x \neq y$, while leaving all other elements fixed. We say that the symmetric group is *generated* by the transpositions.

Example 5.1. The symmetric group S_3 has $3! = 6$ elements: the identity map, three transpositions ($1 \leftrightarrow 2$, $1 \leftrightarrow 3$, $2 \leftrightarrow 3$), and two cyclic permutations ("3-cycles"), which send $1 \to 2 \to 3 \to 1$ and $1 \to 3 \to 2 \to 1$, respectively. The latter can be written as products of two transpositions.

Another well-known example is the *unitary group* $\mathrm{U}(d)$, which consists of all unitary $d \times d$-matrices. The multiplication operation is matrix multiplication, the identity matrix serves as the identity element, and inverses are given by the matrix inverses. The unitary group contains a useful subgroup, the *special unitary group*, which consists of the matrices with unit determinant:

$$\mathrm{SU}(d) = \{U \in \mathrm{U}(d) | \det(U) = 1\}$$

Note that any unitary matrix $U \in \mathrm{U}(d)$ can be written as the product of a complex number with absolute value (which we can think of as a multiple of an identity matrix) and a matrix in $\mathrm{SU}(d)$:

$$U = \underbrace{\det(U)^{1/d}}_{\in \mathrm{U}(1)} \underbrace{\frac{U}{\det(U)^{1/d}}}_{\in \mathrm{SU}(d)}, \tag{5.3}$$

We can summarize this as $\mathrm{U}(d) = \mathrm{U}(1)\,\mathrm{SU}(d)$.

We can use groups to describe symmetries by letting them *act* on mathematical objects. For example, we can let groups operate on vector spaces and ask that the action of each group element is described by a linear map. This is called a representation. Formally, a *(unitary) representation* of a group G consists of a Hilbert space $\mathcal{H}$ (which for us will always be finite-dimensional), along with unitary operators $\{R_g\}_{g \in G}$ on $\mathcal{H}$, such that

$$R_{gh} = R_g R_h \qquad (\forall g, h \in G).$$

This requirement also implies that $R_1 = I$ (the identity element acts as the identity operator) and that $R_g^\dagger = R_g^{-1} = R_{g^{-1}}$ for all $g \in G$. In other words, a unitary representation is nothing but a group homomorphism $G \to \mathrm{U}(\mathcal{H})$. We will typically omit the term "unitary" because those are the only representations that we will meet in this course. In fact, we will usually speak of "the representation $\mathcal{H}$", omitting both the operators R_g, as long as they are clear from the context. We also say that G "acts" on $\mathcal{H}$.

Example 5.2. As discussed in Chapter 4, the n qu*d*it Hilbert space $\mathcal{H} = (\mathbb{C}^d)^{\otimes n}$ is a representation of the symmetric group S_n, with operators (Eq. (4.4))

$$R_\pi \colon (\mathbb{C}^d)^{\otimes n} \to (\mathbb{C}^d)^{\otimes n}, \quad R_\pi \left|\psi_1\right\rangle \otimes \ldots \otimes \left|\psi_n\right\rangle = \left|\psi_{\pi^{-1}(1)}\right\rangle \otimes \ldots \otimes \left|\psi_{\pi^{-1}(n)}\right\rangle \qquad (\pi \in S_n).$$

In fact, it is *also* a representation of the unitary group $\mathrm{U}(d)$, with operators

$$T_U \colon (\mathbb{C}^d)^{\otimes n} \to (\mathbb{C}^d)^{\otimes n}, \quad T_U = U^{\otimes n} \qquad (U \in \mathrm{U}(d)).$$

Importantly, both actions commute: we have

$$[R_\pi, T_U] = 0 \tag{5.4}$$

for all $\pi \in S_n$ and $U \in \mathrm{U}(d)$. While intuitively clear, we verify this by a short calculation:

$$\begin{aligned} &U^{\otimes n} R_\pi(|\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle) = \left(U |\psi_{\pi^{-1}(1)}\rangle\right) \otimes \ldots \otimes \left(U |\psi_{\pi^{-1}(n)}\rangle\right) \\ &= R_\pi \left(U |\psi_1\rangle\right) \otimes \ldots \otimes \left(U |\psi_n\rangle\right) = R_\pi U^{\otimes n} |\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle . \end{aligned}$$

The symmetrizer Π_n commutes with both group actions. For the action of $\mathrm{U}(d)$, this is a direct consequence of Eq. (5.4), while for the action of S_n it is even true that $R_\pi \Pi_n = \Pi_n R_\pi = \Pi_n$ for all $\pi \in S_n$, as follows from Theorem 4.4.

The Hilbert space $(\mathbb{C}^d)^{\otimes n}$ is actually a rather complicated representation that we still need to understand better. First we discuss some simpler examples which we can fully work out by hand.

Example 5.3. Let us study some representations of the group S_3, which discussed in Theorem 5.1. Like any group, S_3 has a one-dimensional *trivial representation*:

$$\mathcal{H} = \mathbb{C}, \qquad R_\pi |0\rangle = |0\rangle \quad (\pi \in S_n)$$

where $|0\rangle$ denotes the standard basis vector of $\mathbb{C}$. This is a maximally boring representation, because every group element $\pi \in S_n$ acts by the (1×1) identity matrix.

The *sign representation* is also one-dimensional but more interesting:

$$\mathcal{H} = \mathbb{C}, \qquad R_\pi |0\rangle = \mathrm{sign}(\pi) |0\rangle \quad (\pi \in S_n)$$

Here, we use the *sign* of a permutation, which is uniquely defined by the following two properties: $\mathrm{sign}(x \leftrightarrow y) = -1$ for any transposition $x \leftarrow y$, and $\mathrm{sign}(\pi\tau) = \mathrm{sign}(\pi)\,\mathrm{sign}(\tau)$ for any two permutations $\pi, \tau \in S_n$. Thus, $\mathrm{sign}(\pi) = +1$ if π can be written as a product of an even number of swaps, and otherwise $\mathrm{sign}(\pi) = -1$. Thus, $R_{1\leftrightarrow 2} = -I$, while $R_{1\to 2\to 3\to 1} = I$.

Finally, we will let S_3 act on three-dimensional vectors imply by permuting the vectors' coordinates. This is sometimes called the *defining representation*:

$$\mathcal{H} = \mathbb{C}^3 = \{\alpha |1\rangle + \beta |2\rangle + \beta |3\rangle : \alpha, \beta, \gamma \in \mathbb{C}\} = \{ \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \}, \qquad R_\pi |j\rangle = |\pi(j)\rangle . \tag{5.5}$$

Here, unusually, we denote the standard basis by $|1\rangle, |2\rangle, |3\rangle$ rather than $|0\rangle, |1\rangle, |2\rangle$ to get cleaner formulas. Note that R_π is simply the *permutation matrix* associated with the permutation $\pi \in S_n$. For example:

$$R_{2\leftrightarrow 3} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \alpha \\ \gamma \\ \beta \end{pmatrix} \quad \text{or} \quad R_{2\leftrightarrow 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Note that this matrix has determinant -1. It is not hard to see that the sign of a permutation is always equal to the determinant of its representation matrix.

All three representations discussed in the example naturally generalize to S_n.

5.2 Decomposing representations

A useful way to analyze a representation is to decompose it into smaller building blocks. To this end we call $\mathcal{K} \subseteq \mathcal{H}$ an *invariant subspace* of a representation $\mathcal{H}$ if it is a linear subspace such that for every vector $|\psi\rangle \in \mathcal{K}$ and for every group element $g \in G$, it holds that $R_g |\psi\rangle \in \mathcal{K}$. In short: $\mathcal{K}$ is an invariant subspace if $R_g \mathcal{K} \subseteq \mathcal{K}$ for all $g \in G$. Every representation has two invariant subspaces which are not particularly interesting: $\{0\}$ and $\mathcal{H}$ itself. We say that $\mathcal{K}$ is *nontrivial* if it is neither of the two. An *irreducible representation* or *irrep* is one that has no nontrivial invariant subspaces.

If an representation is not irreducible, then it can be decomposed into smaller building blocks. To see this, note that if $\mathcal{K} \subseteq \mathcal{H}$ is an invariant subspace, so is its orthogonal complement $\mathcal{K}^\perp$. Indeed, if $|\phi\rangle \in \mathcal{K}^\perp$ then, for all $|\psi\rangle \in \mathcal{K}$,

$$\langle \psi | R_g | \phi \rangle = \langle R_g^\dagger \psi | \phi \rangle = \langle R_{g^{-1}} \psi | \phi \rangle = 0,$$

since $R_{g^{-1}} |\psi\rangle \in \mathcal{K}$. This shows that $R_g |\phi\rangle \in \mathcal{K}^\perp$. As a consequence, the operators R_g are block diagonal with respect to the orthogonal direct sum $\mathcal{H} = \mathcal{K} \oplus \mathcal{K}^\perp$:

$$R_g = R_g^{\mathcal{K}} \oplus R_g^{\mathcal{K}^\perp} = \begin{bmatrix} R_g^{\mathcal{K}} & 0 \\ 0 & R_g^{\mathcal{K}^\perp} \end{bmatrix}, \tag{5.6}$$

where $R_g^{\mathcal{H}}$ denotes the restriction of R_g to the subspace $\mathcal{K}$ and $R_g^{\mathcal{K}^\perp}$ the restriction to $\mathcal{K}^\perp$. Note that the operators $\{\tilde{R}_g\}$ turn $\mathcal{K}$ into a representation of G; likewise for $\{\hat{R}_g\}$ and $\mathcal{K}^\perp$. Thus we have successfully decomposed the given representation $\mathcal{H}$ into "smaller" representations $\mathcal{K}$ and $\mathcal{K}^\perp$. If $\mathcal{K}$ is a *nontrivial* invariant subspace then these are indeed of lower dimension. In this case, we can separately apply the same reasoning to $\mathcal{K}$ and $\mathcal{K}^\perp$ and continue this process until we arrive at a decomposition

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \ldots \oplus \mathcal{H}_m \tag{5.7}$$

that cannot be refined further, so we must have that each $\mathcal{H}_k$ is irreducible. Note that in our construction the summands are orthogonal to each other. We conclude that every representation can be written as an orthogonal direct sum of irreducible representations. Moreover, we observe that a representation is irreducible if and only if it is indecomposable.

Example 5.4. Any one-dimensional representation is irreducible. In particular, the trivial and the sign representation in Theorem 5.3 are irreducible. However, the three-dimensional representation defined in (5.5) is *not* irreducible, since

$$\mathcal{K} = \{\alpha |1\rangle + \beta |2\rangle + \beta |3\rangle \ : \ \alpha + \beta + \gamma = 0\} = \{ \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} | \alpha + \beta + \gamma = 0\}, \tag{5.8}$$

is a two-dimensional (and hence nontrivial) invariant subspace. In Exercise 5.1 you can show that it is irreducible. Its orthogonal complement is given by

$$\mathcal{K}^\perp = \mathbb{C}(|1\rangle + |2\rangle + |3\rangle) = \{ \begin{pmatrix} \alpha \\ \alpha \\ \alpha \end{pmatrix} \}.$$

As it is one-dimensional, it must be irreducible. Note that R_π acts just like in (is "equivalent" to) the trivial representation: we have $R_\pi(|1\rangle + |2\rangle + |3\rangle) = |1\rangle + |2\rangle + |3\rangle$ for all $\pi \in S_3$.

Example 5.5 (Symmetric subspace)**.** As discussed in Theorem 5.2, we can think of $(\mathbb{C}^d)^{\otimes n}$ as a representation of both S_n and $\mathrm{U}(d)$.

From the perspective of the symmetric group S_n, $\mathrm{Sym}^n(\mathbb{C}^d)$ is clearly an invariant subspace. However, *any* subspace $W \subseteq \mathrm{Sym}^n(\mathbb{C}^d)$ is also a invariant, since $R_\pi |\phi\rangle = |\phi\rangle$ holds for *every* vector $|\phi\rangle \in \mathrm{Sym}^n(\mathbb{C}^d)$. Thus, $\mathrm{Sym}^n(\mathbb{C}^d)$ is *not* irreducible as a representation of S_n.

From the perspective of the unitary group $\mathrm{U}(d)$, the symmetric subspace is also an invariant subspace. This follows from Eq. (5.4). Indeed, for every $|\Phi\rangle \in \mathrm{Sym}^n(\mathbb{C}^d)$ and $U \in \mathrm{U}(d)$, we have

$$R_\pi(U^{\otimes n} |\Phi\rangle) = U^{\otimes n}(R_\pi |\Phi\rangle) = U^{\otimes n} |\Phi\rangle ,$$

and thus $U^{\otimes n} |\Phi\rangle \in \mathrm{Sym}^n(\mathbb{C}^d)$. In fact, $\mathrm{Sym}^n(\mathbb{C}^d)$ is an irreducible representation of $\mathrm{U}(d)$! We will prove this carefully in Chapter 6.

Composing representations

So far we decomposed representations, but we can also assemble larger representations from smaller building blocks. For example, given two representations $\mathcal{K}$ and $\mathcal{L}$ of the same group G, with operators $\{R_g^{\mathcal{K}}\}_{g\in G}$ and $\{R_g^{\mathcal{L}}\}_{g\in G}$, their *direct sum* is naturally a representation of G. Simply define

$$\mathcal{H} := \mathcal{K} \oplus \mathcal{L}, \qquad R_g := R_g^{\mathcal{K}} \oplus R_g^{\mathcal{L}} \quad (g \in G)$$

We can also turn their *tensor product* into a representation:

$$\mathcal{H} := \mathcal{K} \otimes \mathcal{L}, \qquad R_g := R_g^{\mathcal{K}} \otimes R_g^{\mathcal{L}} \quad (g \in G)$$

In particular we may apply this in the case that $\mathcal{L} = \mathbb{C}^m$ is a trivial representation of dimension m (i.e., $R_g^{\mathcal{L}} = I_m$). It is instructive to observe that

$$\mathcal{K} \otimes \mathbb{C}^m \cong \underbrace{\mathcal{K} \oplus \ldots \oplus \mathcal{K}}_{m \text{ times}}, \qquad R_g^{\mathcal{K}} \otimes I_m \cong \begin{bmatrix} R_g^{\mathcal{K}} & & & \\ & R_g^{\mathcal{K}} & & \\ & & \ddots & \\ & & & R_g^{\mathcal{K}} \end{bmatrix}.$$

Thus, $\mathcal{K} \otimes \mathbb{C}^m$ is a convenient way of denoting a direct sum of m copies of $\mathcal{K}$.

5.3 Intertwiners and Schur's lemma

An important part of representation theory is to *classify* all representations of a given group G. But how can we compare different representations? In particular, when can we say that two representations are "the same"?

Suppose that $\mathcal{H}$ and $\mathcal{H}'$ are two representations of a group G, with operators $\{R_g\}_{g\in G}$ and $\{R'_g\}_{g\in G}$, respectively. A linear map $J\colon \mathcal{H} \to \mathcal{H}'$ is called an *intertwiner* if it holds that

$$JR_g = R'_g J \qquad (\forall g \in G).$$

In other words, it "intertwines" the group action, hence the name. Such maps are also called *equivariant*. When the intertwiner is invertible, we can write the above as

$$JR_g J^{-1} = R'_g \qquad (\forall g \in G).$$

Thus, a *single* isomorphism (change of coordinates) J relates *all* the representation operators (matrices). An invertible intertwiner is called an *equivalence*, and we say that the two representations $\mathcal{H}$ and $\mathcal{H}'$ are *equivalent*. We denote this by $\mathcal{H} \cong \mathcal{H}'$ or $\{R_g\} \cong \{R'_g\}$.

An important tool in this context is known as *Schur's Lemma*. Roughly speaking, it states that there no nontrivial intertwiners between inequivalent irreps, while for equivalent ones they are unique (up to a scalar). The formal statement is as follows:

Lemma 5.6 (Schur). *Let $J\colon \mathcal{H} \to \mathcal{H}'$ be an intertwiner between* irreducible *representations.*

(a) Either J is invertible (and hence $\mathcal{H} \cong \mathcal{H}'$) or $J = 0$.
(b) If the two representations are the same (i.e., $\mathcal{H} = \mathcal{H}'$ and $R_g = R'_g$ for all $g \in G$), then J is proportional to the identity operator: $J = \lambda I$ for some $\lambda \in \mathbb{C}$.

Proof. (a) Suppose that $J \neq 0$, so we want to show that J is invertible. Both $\ker(J)$ and $\mathrm{ran}(J)$ are invariant subspaces, as is readily verified. Since $\mathcal{H}$ is irreducible, this means that either $\ker(J) = \{0\}$ or $\ker(J) = \mathcal{H}$. We must have the former, since otherwise $J = 0$ – so J is injective. Similarly, since $\mathcal{H}'$ is irreducible, either $\mathrm{ran}(J) = \{0\}$ or $\mathrm{ran}(J) = \mathcal{H}'$. We must have the latter, since otherwise $J = 0$ – thus, J is also surjective. We conclude that J is invertible.

(b) Any operator $J\colon \mathcal{H} \to \mathcal{H}$ on a complex vector space has an eigenvalue $\lambda \in \mathbb{C}$. Thus, $\ker(J - \lambda I) \neq \{0\}$. But if J is an intertwiner then so is $J - \lambda I$ (here we use that $R_g = R'_g$). Thus $\ker(J - \lambda I)$ is an invariant subspace other than $\{0\}$. Since $\mathcal{H}$ is irreducible, we must therefore have that $\ker(J - \lambda I) = \mathcal{H}$. We conclude that $J = \lambda I$. □

In part (a) of Schur's lemma, J is in fact proportional to a unitary. You can show this in Exercise 5.3. As a consequence, if $\mathcal{H} \cong \mathcal{H}'$ then there always exists a *unitary* intertwiner. This is true even if the representations are not irreducible.

5.4 Proof of the integral formula

We can use Schur's lemma to prove the integral formula for the symmetrizer. The key idea is the following: the right-hand side of Eq. (5.1) is an operator on the symmetric subspace. and Eq. (5.2) shows that it is an intertwiner with respect to the action of $\mathrm{U}(d)$. Because the symmetric subspace is irreducible (as we claimed in Theorem 5.5 and will prove in Chapter 6), part (b) of Schur's Lemma implies at once that the operator must be proportional to Π_n. Finally one can verify proportionality constant is one by comparing the trace.

We now give the formal statement and a more detailed proof.

Theorem 5.7. *The symmetrizer $\Pi_n = \frac{1}{n!}\sum_{\pi\in S_n} R_\pi$ is equal to*

$$\Pi_n = \binom{n+d-1}{n} \int d\phi \, |\phi\rangle^{\otimes n} \langle\phi|^{\otimes n},$$

where $d\phi$ denotes the uniform probability measure on the space of pure states $\{\phi\} = \{|\phi\rangle\langle\phi|\}$.

Proof. As in Eq. (5.1), we denote

$$\Pi'_n := \binom{n+d-1}{n} \int d\phi \, |\phi\rangle^{\otimes n} \langle\phi|^{\otimes n}.$$

Thus our goal is to prove that $\Pi_n = \Pi'_n$. Both the left-hand side are operators on $(\mathbb{C}^d)^{\otimes n}$. Let us abbreviate the symmetric subspace by $\mathcal{H} = \mathrm{Sym}^n(\mathbb{C}^d)$, so that $(\mathbb{C}^d)^{\otimes n} = \mathcal{H} \oplus \mathcal{H}^\perp$. If we

decompose the operators accordingly, we see that they are block diagonal:

$$\Pi_n = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \Pi'_n = \begin{bmatrix} J & 0 \\ 0 & 0 \end{bmatrix},$$

where I is the identity operator on $\mathcal{H}$ and J is some operator on $\mathcal{H}$ that we still need to characterize. The former holds because Π_n is the orthogonal projection onto the symmetric subspace (Theorem 4.4), so it acts as the identity on $\mathcal{H}$ and sends all orthogonal vectors to zero. The latter holds because every $|\phi\rangle^{\otimes n}$ is in the symmetric subspace, so Π'_n maps any vector into the symmetric subspace and it maps vectors orthogonal to the symmetric subspace to zero. We can also decompose the group action. Since $\mathcal{H}$ is an invariant subspace for the action of $\mathrm{U}(d)$, so is $\mathcal{H}^\perp$, and hence

$$U^{\otimes n} = \begin{bmatrix} T_U^{\mathcal{H}} & 0 \\ 0 & T_U^{\mathcal{H}^\perp} \end{bmatrix},$$

with $T_U^{\mathcal{H}}$ the restriction of $U^{\otimes n}$ to the symmetric subspace and $T_U^{\mathcal{H}^\perp}$ the restriction to the orthogonal complement. In Eq. (5.2) we proved that the unitary invariance of the measure $d\phi$ implies that

$$U^{\otimes n}\Pi'_n = \Pi'_n U^{\otimes n}.$$

Using the block diagonal form of the operators, it follows that

$$T_U^{\mathcal{H}} J = J T_U^{\mathcal{H}}.$$

In other words, J is an intertwiner with respect to the action of $\mathrm{U}(d)$ on the symmetric subspace $\mathcal{H} = \mathrm{Sym}^n(\mathbb{C}^d)$! Because the latter is irreducible, part (b) of Schur's lemma shows that J must be proportional to I, i.e., there exists $\lambda \in \mathbb{C}$ such that $J = \lambda I_{\mathcal{H}}$ and therefore

$$\Pi'_n = \lambda\, \Pi_n.$$

It remains to prove that $\lambda = 1$. To this end, we compute the trace of the two operators:

$$\begin{aligned} \operatorname{tr} \Pi_n &= \dim \mathrm{Sym}^n(\mathbb{C}^d) = \binom{n+d-1}{n} \\ \operatorname{tr} \Pi'_n &= \binom{n+d-1}{n} \int d\phi \underbrace{\operatorname{tr}\left[|\phi\rangle^{\otimes n} \langle\phi|^{\otimes n}\right]}_{=\langle\phi^{\otimes n}|\phi^{\otimes n}\rangle = 1} = \binom{n+d-1}{n}. \end{aligned}$$

The former is Eq. (4.8), and in the latter we used that $d\phi$ is a probability measure, so that $\int d\phi = 1$. Thus, $\operatorname{tr} \Pi'_n = \operatorname{tr} \Pi_n \neq 0$ and hence we must have $\lambda = 1$, concluding the proof. □

Remark 5.8. Above we used Schur's lemma for the action of $\mathrm{U}(d)$, but the symmetric subspace is also an invariant subspace for the action of S_n. This means that we have both

$$U^{\otimes n} = \begin{bmatrix} T_U^{\mathcal{H}} & 0 \\ 0 & T_U^{\mathcal{H}^\perp} \end{bmatrix} \quad \text{and} \quad R_\pi = \begin{bmatrix} R_\pi^{\mathcal{H}} & 0 \\ 0 & R_\pi^{\mathcal{H}^\perp} \end{bmatrix}.$$

Since $[T_U, R_\pi] = 0$, it follows that $[T_U^{\mathcal{H}}, R_\pi^{\mathcal{H}}] = 0$ for every $U \in \mathrm{U}(d)$ and $\pi \in S_n$, that is,

$$T_U^{\mathcal{H}} R_\pi^{\mathcal{H}} = R_\pi^{\mathcal{H}} T_U^{\mathcal{H}}.$$

In other words, the operators $R_\pi^{\mathcal{H}}$ are intertwiners with respect to the action of $\mathrm{U}(d)$, and the operators $T_U^{\mathcal{H}}$ are intertwiners with respect to the action of S_n. What can we learn by applying Schur's lemma in this situation?

Because the symmetric subspace is an irreducible $\mathrm{U}(d)$-representation, part (b) of Schur's lemma implies that each $R_\pi \propto I$. But indeed, each R_π acts trivially on the symmetric subspace (by its very definition), so $R_\pi = I$ and the preceding is in complete agreement with what we already know.

What if we consider the symmetric subspace as an S_n-representation? It is clearly not the case that the operators $T_U^{\mathcal{H}}$ are proportional to the identity. For example, we have $X^{\otimes 2}|0,0\rangle = |1,1\rangle$, which is not proportional to $|0,0\rangle$. Fortunately this is no contradiction: Schur's lemma is simply not applicable in this case, because the symmetric subspace is *not* irreducible as an S_n-representation. In fact, as just discussed, S_n acts trivially on the symmetric subspace and hence $\mathcal{H}$ decomposes into $\binom{n+d-1}{n}$ many one-dimensional trivial representations of S_n.

Exercises

5.1 **Irreps of S_3:** Show that the representation defined in Eq. (5.8) is irreducible.

5.2 **Defining representation of SU(2):** The group $\mathrm{U}(2)$ acts on $\mathcal{H} = \mathbb{C}^2$ by matrix-vector-multiplication. This is called the *defining representation* of $\mathrm{U}(2)$. Show that it is irreducible, even if we only act by $\mathrm{SU}(2)$.

5.3 **Schur's lemma and unitarity:** Here you can strengthen part (a) of Schur's lemma.

(a) Show that if J is an intertwiner then so is $J^\dagger$.
(b) Conclude that any intertwiner between irreducible representation must be proportional to a unitary operator.

5.4 **Schur's lemma:** Let G be a commutative group (i.e., $gh = hg$ for all $g, h \in G$). Show that any irreducible representation of G is necessarily one-dimensional.

5.5 **Spectral theorem:** Let M be a Hermitian operator acting on a Hilbert space $\mathcal{H}$.

(a) Show that $R_t := e^{\imath M t}$ defines a unitary representation of the group $G = \mathbb{R}$ (with the addition of real numbers as the group "multiplication") on $\mathcal{H}$.
(b) Decompose $\mathcal{H}$ into irreducible representations.

Chapter 6

Irreducibility of the symmetric subspace

In last lecture's introduction to representation theory, we states that the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^d)$ is an irreducible representation of $\mathrm{U}(d)$. We used this irreducibility, together with Schur's lemma, to prove the important integral formula in Eq. (4.9) for the projector onto the symmetric subspace. This week we will show that the symmetric subspace is indeed irreducible.

In the lecture we will only discuss the case of qubits ($d = 2$), but the proof strategy generalizes directly and you can prove the general case in Exercise 6.1. To start, recall from Eq. (4.7) that $\mathrm{Sym}^n(\mathbb{C}^2)$ has the following *occupation number basis*:

$$\begin{aligned}
\|n, 0\rangle\rangle &= |\underbrace{0 \ldots 0}_{n \text{ times}}\rangle = |0\rangle^{\otimes n} \\
\|n-1, 1\rangle\rangle &= \frac{1}{\sqrt{n}} \left(|\underbrace{0 \ldots 0}_{n-1 \text{ times}} 1\rangle + |\underbrace{0 \ldots 0}_{n-2 \text{ times}} 1\, 0\rangle + \cdots + |1 \underbrace{0 \ldots 0}_{n-1 \text{ times}}\rangle \right) \\
&\vdots \\
\|m, n-m\rangle\rangle &= \sqrt{\frac{n!}{m!(n-m)!}} \left(|\underbrace{0 \ldots 0}_{m \text{ times}} \underbrace{1 \ldots 1}_{n-m \text{ times}}\rangle + \text{ permutations} \right) \\
&\vdots \\
\|0, n\rangle\rangle &= |\underbrace{1 \ldots 1}_{n \text{ times}}\rangle = |1\rangle^{\otimes n}
\end{aligned} \tag{6.1}$$

Thus, the m-th basis vector $\|m, n-m\rangle\rangle$ is given by the uniform superposition of all bitstrings with m zeros and $n-m$ ones. Because $m \in \{0, \ldots, n\}$, there are $n+1$ such basis vectors.

To prove that $\mathrm{Sym}^n(\mathbb{C}^2)$ is irreducible, we will show that the following holds for any invariant subspace $\mathcal{K} \subseteq \mathrm{Sym}^n(\mathbb{C}^2)$:

(a) if $\mathcal{K} \neq \{0\}$, then $\mathcal{K}$ must contain at least one of the basis vectors $\|m, n-m\rangle\rangle$, and
(b) if $\mathcal{K}$ contains one such basis vector, then it must in fact contain *all* these basis vectors.

Clearly, (a) and (b) together imply that either $\mathcal{K} = \{0\}$ or $\mathcal{K} = \mathrm{Sym}^n(\mathbb{C}^2)$. In other words, $\mathrm{Sym}^n(\mathbb{C}^2)$ has no *nontrivial* invariant subspaces and hence it is irreducible.

6.1 Lie algebra and representation

To realize the above strategy, we need to get a better handle on invariant subspaces. Recall that an invariant subspace $\mathcal{K} \subseteq \mathrm{Sym}^n(\mathbb{C}^2)$ is one such that

$$T_U \mathcal{K} \subseteq \mathcal{K} \tag{6.2}$$

for every $U \in \mathrm{U}(2)$, where $T_U = U^{\otimes n}$ is the action of $\mathrm{U}(2)$. This is a *nonlinear* condition in U, which makes it rather difficult to work with.

We will now learn a powerful technique that can be used to linearize it. We will discuss this next for general $\mathrm{U}(d)$. The basic idea is that the exponential map sums to products. Because we deal with matrices, we need the *matrix exponential*, which for an arbitrary complex matrix A is defined, e.g., via the usual power series $e^A := \sum_{k=0}^{\infty} \frac{A^k}{k!}$. The matrix exponential has a number of useful properties:

Lemma 6.1. *For any $d \times d$-matrix A, it holds that*

(a) $(e^A)^\dagger = e^{(A^\dagger)}$.
(b) $e^{A \otimes I} = e^A \otimes I$.
(c) If $[A, B] = 0$ commute, then $e^A e^B = e^{A+B}$.
(d) $U e^A U^\dagger = e^{UAU^\dagger}$.
(e) $\det(e^A) = e^{\mathrm{tr}[A]}$.

All but the last can be directly verified from the power series. If M is Hermitian, with spectral decomposition $M = \sum_i m_i |\phi_i\rangle\langle\phi_i|$, then we can compute its exponential simply by exponentiating each eigenvalue, i.e., $e^M = \sum_i e^{a_i} |\phi_i\rangle\langle\phi_i|$. So at least in this case, the last property is also easy to see. This observation also shows that any unitary matrix can be written as the matrix exponential of an *anti*-Hermitian matrix:

$$\mathrm{U}(d) = \left\{ e^M \mid M^\dagger = -M \right\} = \left\{ e^{\imath H} \mid H^\dagger = H \right\}.$$

This generalizes the fact that any complex number of absolute value one can be written in the form $e^{\imath\theta}$ for some $\theta \in \mathbb{R}$. To understand what this means geometrically, let $M = -M^\dagger$ and consider $U_s = e^{sM}$, which is a curve of unitaries parameterized by $s \in \mathbb{R}$. If we take the derivative at $s = 0$, we get

$$\dot{U}_0 = \partial_{s=0}\, U_s = \partial_{s=0}\, e^{sM} = M, \tag{6.3}$$

so we can think of M as the *tangent vector* of the curve at $U_0 = I$, as in the following picture:

Mathematically, we have used the fact that the group $\mathrm{U}(d)$ is a *Lie group*, which means that it is not just a set but a smooth manifold and all group operations are smooth. The tangent space of a Lie group at the identity is a *Lie algebra*, meaning that it is closed under commutators $[\cdot, \cdot]$. The Lie algebra of $G = \mathrm{U}(d)$ consists of the anti-Hermitian matrices (Eq. (6.3)). Indeed, if M and N are anti-Hermitian then so is $[M, N]$, because

$$[M, N]^\dagger = (MN - NM)^\dagger = N^\dagger M^\dagger - M^\dagger N^\dagger = NM - MN = -[M, N].$$

If we have an representation $\{R_U\}$ of $\mathrm{U}(d)$ on some Hilbert space $\mathcal{H}$ then (assuming it is smooth), we can also consider the corresponding curve R_{U_s}. Then the "infinitesimal action" in direction M is by definition

$$r_M := \partial_{s=0} R_{U_s} = \partial_{s=0} R_{e^{sM}}.$$

Because derivatives depend linearly on the tangent vector, the mapping $M \mapsto r_M$ is a *linear* map from the anti-Hermitian $d \times d$-matrices to the anti-Hermitian operators on $\mathcal{H}$. It is called the *Lie algebra representation* on $\mathcal{H}$, or the *action* of the Lie algebra. We will always use upper case letters for Lie group actions and the corresponding lower case letters for the associated Lie algebra actions. In fact, because the group action is on a complex vector space, the map $M \mapsto r_M$ is well-defined not just for anti-Hermitian matrices but for general complex $d \times d$-matrices M.

Example 6.2. To convince you of these facts, let us specialize to the action $T_U = U^{\otimes n}$ of $\mathrm{U}(d)$ on $\mathcal{H} = (\mathbb{C}^d)^{\otimes n}$. Here, the Lie algebra action is

$$t_M = \partial_{s=0} \big(e^{sM}\big)^{\otimes n} = \partial_{s=0} \big(e^{sM} \otimes \cdots \otimes e^{sM}\big) = M \otimes I \otimes \cdots \otimes I + \cdots + I \otimes \cdots \otimes I \otimes M,$$

using the product rule. It is plain that $M \mapsto t_M$ is well-defined for arbitrary complex $d \times d$ matrices, and a complex linear map to the space of linear operators on $\mathcal{H}$. This can also be seen as follows: the formula for the group action makes sense not just for the unitary group but for the entire the *general linear group* $\mathrm{GL}(d)$, which consists of all invertible $n \times n$-matrices, and the Lie algebra of the latter consists of all complex $d \times d$-matrices.

In fact, one can also recover the Lie group action from the Lie algebra action. We will not need this in today's lecture, but it is still useful to know.

Lemma 6.3. *It holds that $R_{e^M} = e^{r_M}$ for every $M = -M^\dagger$.*

Proof. The curve $V_s := e^{s r_M}$ is the unique solution to the first-order ordinary differential equation $\dot{V}_s = V_s r_M$ with initial condition $V_0 = I$. It suffices to show that $W_s := R_{e^{sM}}$ solves the same ODE. Indeed, observe that $W_0 = R_I = I$ and for any s we have

$$\begin{aligned} \dot{W}_s = \partial_{\varepsilon=0} W_{s+\varepsilon} &= \partial_{\varepsilon=0} R_{e^{(s+\varepsilon)M}} \\ &= \partial_{\varepsilon=0} R_{e^{sM} e^{\varepsilon M}} \\ &= \partial_{\varepsilon=0} R_{e^{sM}} R_{e^{\varepsilon M}} \\ &= R_{e^{sM}} \partial_{\varepsilon=0} R_{e^{\varepsilon M}} = W_s r_M, \end{aligned}$$

where the second line follows from (c) of Theorem 6.1, the third uses the fact that $R_{UU'} = R_U R_{U'}$ for any representation, and the fourth line holds because $R_{e^{sM}}$ is a linear operator and hence commutes with the derivative. This concludes the proof. □

We can now state the key observation that "linearizes" Eq. (6.2).

Lemma 6.4. *Let $\mathcal{H}$ be a representation of $\mathrm{U}(d)$ and let $\mathcal{K} \subseteq \mathcal{H}$ be a subspace. Then the following are equivalent:*

(a) $\mathcal{K}$ is an invariant subspace for the action of $\mathrm{U}(d)$. That is, $R_U \mathcal{K} \subseteq \mathcal{K}$ for all $U \in \mathrm{U}(d)$.
(b) $\mathcal{K}$ is an invariant subspace for the action of the Lie algebra of $\mathrm{U}(d)$. That is, $r_M \mathcal{K} \subseteq \mathcal{K}$ for all anti-Hermitian $d \times d$-matrices M.
(c) $\mathcal{K}$ is an invariant subspace for the action of the Lie algebra of $\mathrm{GL}(d)$. That is, $r_M \mathcal{K} \subseteq \mathcal{K}$ for all complex $d \times d$-matrices M.

Proof. To prove that (a) implies (b), let $M = -M^\dagger$. Then, for every $|\phi\rangle \in \mathcal{K}$ we have that

$$r_M |\phi\rangle = \partial_{s=0} \underbrace{R_{e^{sM}} |\phi\rangle}_{\in \mathcal{K}}.$$

The underbraced expression is in $\mathcal{K}$ because $e^{sM} \in \mathrm{U}(d)$ and we assumed that $\mathcal{K}$ is an invariant subspace for the $\mathrm{U}(d)$-action. Since any (finite-dimensional) vector subspace is closed, the limit of a sequence of vectors in $\mathcal{K}$ must again be in $\mathcal{K}$. As the derivative is a limit of difference quotients, each of which is in $\mathcal{K}$, the claim follows.

Next, we note that (b) implies (c) by linearity. Indeed, we can write any matrix M in the form $M = A + \imath B$, with A and B anti-Hermitian. Then it holds that $r_M = r_A + \imath r_B$ and the claim follows.

Finally we claim that (c) implies (a). Take any $U \in \mathrm{U}(d)$ and write it as $U = e^M$ for some anti-Hermitian M. Using Theorem 6.3, we find that, for every $|\phi\rangle \in \mathcal{K}$,

$$R_U |\phi\rangle = R_{e^M} |\phi\rangle = e^{r_M} |\phi\rangle = \sum_{k=0}^{\infty} \underbrace{\frac{r_M^k}{k!} |\phi\rangle}_{\in \mathcal{K}}.$$

The underbraced expression is in $\mathcal{K}$ because $r_M \mathcal{K} \subseteq \mathcal{K}$ by assumption. As above it follows that the limit $R_U |\phi\rangle \in \mathcal{K}$ is also in $\mathcal{K}$. □

Note that it suffices to check conditions (b) and (c) for operators M in a basis of the Lie algebra.

6.2 Proof of irreducibility of the symmetric subspace

We now use the above techniques to prove that the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^2)$ is an irreducible representation of $\mathrm{U}(2)$. To implement the plan outlined at the beginning of the lecture, we will use the Lie algebra action, which in the present setting is given by

$$t_M = M \otimes I \otimes \cdots \otimes I + \ldots + I \otimes \cdots \otimes I \otimes M,$$

as discussed in Theorem 6.2. The following examples show that this is a useful idea:

- If $M = Z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ is the Pauli Z matrix, then t_Z acts as follows on the computational basis of $(\mathbb{C}^2)^{\otimes n}$:

  $$t_Z |i_1, \ldots, i_n\rangle = \sum_{k=1}^{n} (-1)^{i_k} |i_1, \ldots, i_n\rangle = \big(\#0\text{'s} - \#1\text{'s}\big) |i_1, \ldots, i_n\rangle,$$

  where $\#0$'s denotes the number of zeros in $i_1, \ldots, i_n \in \{0, 1\}$ and $\#1$'s the number of ones. As a consequence:

  $$t_Z \|m, n-m\rangle\rangle = (m - (n-m)) \|m, n-m\rangle\rangle = (2m - n) \|m, n-m\rangle\rangle. \tag{6.4}$$

  Thus, t_Z preserves the symmetric subspace and the occupation number basis vectors are precisely the eigenvectors of t_Z on that space. As $m \in \{0, 1, \ldots, n\}$, the corresponding eigenvalues are $\{-n, -n+2, \ldots, n-2, n\}$. Since these eigenvalues are all distinct, we can recover $\|m, n-m\rangle\rangle$ as the unique (up to phase) eigenvector of t_Z.

- If $M_+ = |0\rangle\langle 1| = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, then t_{M_+} acts on a computational basis vector $|i_1, \ldots, i_n\rangle$ by inspecting each bit i_k: if $i_k = 1$, then it is replaced by 0, and otherwise the term does not contribute. For example,

$$\begin{aligned} t_{M_+} |011\rangle &= (M_+ \otimes I \otimes I) |011\rangle + (I \otimes M_+ \otimes I) |011\rangle + (I \otimes I \otimes M_+) |011\rangle \\ &= |001\rangle + |010\rangle . \end{aligned}$$

  As a consequence, it is not hard to verify that, for $m < n$,

$$t_{M_+} \|m, n-m\rangle\rangle = \sqrt{(n-m)(m+1)}\, \|m+1, n-(m+1)\rangle\rangle.$$

  The precise proportionality constant is not important. What matters is that the proportionality constant is nonzero unless $m = n$, in which case the basis vector is annihilated.

- If $M_- = |1\rangle\langle 0| = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ then, similarly, one can see that

$$t_{M_-} \|m, n-m\rangle\rangle = \sqrt{m(n-m+1)}\, \|m-1, n-m+1\rangle\rangle. \tag{6.5}$$

  The proportionality constant is nonzero unless $m = 0$, in which case the basis vector is annihilated.

Thus we have found three operators, t_Z, t_{M_+}, and t_{M_-}, that allow us to *identify* and *transition between* the basis vectors $\|m, n-m\rangle\rangle$ for $m \in \{0, 1, \ldots, n\}$. Because these operators also preserve invariant subspaces (Theorem 6.4), this allows us to implement the proof strategy outlined above.

Theorem 6.5. *The symmetric subspace* $\mathrm{Sym}^n(\mathbb{C}^2)$ *is an irreducible representation of* $\mathrm{U}(2)$.

Proof. Let $\mathcal{K} \subseteq \mathrm{Sym}^n(\mathbb{C}^2)$ be an arbitrary invariant subspace. We first use that $t_Z \mathcal{K} \subseteq \mathcal{K}$ by Theorem 6.4. Because t_Z is a Hermitian operator (also when restricted to the subspace), it follows that $\mathcal{K}$ must be spanned by eigenvectors of t_Z. Now, $\mathcal{K} \subseteq \mathrm{Sym}^n(\mathbb{C}^2)$, and we saw in Eq. (6.4) that the eigenspaces of t_Z are all one-dimensional and spanned by the occupation number basis vectors. Thus, if $\mathcal{K} \neq \{0\}$, then $\mathcal{K}$ must contain at least one of the basis vectors $\|m, n-m\rangle\rangle$ for some $m \in \{0, 1, \ldots, n\}$. On the other hand, we also know that $t_{M_\pm} \mathcal{K} \subseteq \mathcal{K}$, again by Theorem 6.4. It follows that if $\mathcal{K} \neq \{0\}$ then in fact *all* basis vectors $\|m, n-m\rangle\rangle$ are contained in $\mathcal{K}$, and hence $\mathcal{K} = \mathrm{Sym}^n(\mathbb{C}^2)$. This concludes the proof that the symmetric subspace is irreducible when regarded as a representation of the group $\mathrm{U}(2)$. □

Corollary 6.6. *A subspace* $\mathcal{H} \subseteq (\mathbb{C}^2)^{\otimes n}$ *is an irreducible representation of* $\mathrm{SU}(2)$ *if and only if it is an irreducible representation of* $\mathrm{U}(2)$. *In particular,* $\mathrm{Sym}^n(\mathbb{C}^2)$ *is an irreducible representation of* $\mathrm{SU}(2)$.

Proof. Any unitary $U \in \mathrm{U}(2)$ can be writen in the form $U = \lambda U'$, where $\lambda \in \mathrm{U}(1)$ and $U' \in \mathrm{SU}(2)$ (Eq. (5.3)). Because $T_U = \lambda^n T_{U'}$, it is clear that a subspace of $(\mathbb{C}^2)^{\otimes n}$ is invariant for the action of $\mathrm{U}(2)$ if and only if is invariant for the action of $\mathrm{SU}(2)$. □

Example 6.7. One particular consequence is that the decomposition from Theorem 4.5,

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

is a decomposition into irreducible representations of $U(2)$. Indeed, we just proved that $\mathrm{Sym}^2(\mathbb{C}^2)$ is irreducible, and because $\bigwedge^2(\mathbb{C}^2)$ is one-dimensional (it is spanned by singlet state) the latter must also be irreducible. More generally, it holds that

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

is a decomposition into irreducible subspaces.

Exercises

6.1 **Irreducibility:** Show that $\mathrm{Sym}^n(\mathbb{C}^d)$ is an irreducible representation of $\mathrm{U}(d)$ and of $\mathrm{SU}(d)$, by generalizing the argument given above.

6.2 **Dual representations:** This problem introduces the concept of a *dual representation*. To start, consider a representation $\mathcal{H}$ of some group G, with operators $\{R_g\}$. Let $\mathcal{H}^*$ denote the dual Hilbert space, whose elements are "bras" $\langle\phi|$, and define operators R_g^* on $\mathcal{H}^*$ by $R_g^* \langle\phi| := \langle\phi| R_{g^{-1}}$.

(a) Verify that the operators $\{R_g^*\}$ turn $\mathcal{H}^*$ into a representation of G. This representation is called the *dual representation* of $\mathcal{H}$.

(b) Show that $\mathcal{H}$ is irreducible if and only if $\mathcal{H}^*$ is irreducible.

A representation $\mathcal{H}$ is called *self-dual* if $\mathcal{H}^* \cong \mathcal{H}$.

(c) Show that $\mathrm{Sym}^k(\mathbb{C}^2)$ is a self-dual representation of $\mathrm{SU}(2)$. We will see next week that this implies that any representation of $\mathrm{SU}(2)$ is self-dual.
(d) Is $\mathrm{Sym}^k(\mathbb{C}^2)$ self-dual as a representation of $\mathrm{U}(2)$?
(e) Show that any representation of S_3 is self-dual.

More generally, all representations of S_n are self-dual. However, for $d > 2$, most representations of $\mathrm{SU}(d)$ are *not* self-dual.

Chapter 7

Mixed states, partial traces, purifications

This week we will introduce another bit of formalism to our toolbox by generalizing from "pure" states, which are described by unit vectors in a Hilbert space, to so-called "mixed" quantum states, which are mathematically described by certain operators called "density operators". This allows us to describe classical randomness and ensembles of quantum states, as well as the state of subsystems when the global quantum state is entangled (cf. the footnote on p. 31).

7.1 Mixed states and density operators

Suppose that we have a device – a *quantum information source* – that emits certain different quantum states $|\psi_i\rangle$ with probabilities p_i each, where i ranges in some index set I:

We call $\{p_i, |\psi_i\rangle\}_{i\in I}$ an *ensemble* of quantum states. Note that the states $|\psi_i\rangle$ need *not* be orthogonal. If we measure the (random) state emitted by the source by some POVM $\{Q_x\}_{x\in\Omega}$, the probability of outcomes is given by

$$\begin{aligned}
\mathbf{Pr}(\text{outcome } x) &= \sum_i p_i \ \mathbf{Pr}_{\psi_i}(\text{outcome } x) \\
&= \sum_i p_i \ \langle\psi_i|Q_x|\psi_i\rangle \\
&= \sum_i p_i \ \operatorname{tr}\big[|\psi_i\rangle\langle\psi_i|\, Q_x\big] \\
&= \operatorname{tr}\Big[\underbrace{\big(\sum_i p_i\, |\psi_x\rangle\langle\psi_x|\big)}_{=:\rho} Q_x\Big],
\end{aligned}$$

where we first used the fact that state $|\psi_i\rangle$ is emitted with probability p_i and then the Born rule. Thus a single operator ρ captures all information that is needed to compute probabilities of outcomes! We can interpret ρ as the *average state* of the ensemble, or as the average state emitted by the source. Observe that

(a) $\rho \geq 0$, i.e., it is positive semidefinite, and

(b) $\operatorname{tr}\rho = 1$.

We call an operator satisfying these two properties a *density operator* or *density matrix* – or simply a *quantum state* on $\mathcal{H}$. From here onwards we will always use the term "quantum state" in to mean density operators. As we just computed, the *Born rule* for density operators reads

$$\mathbf{Pr}_\rho(\text{outcome } x) = \operatorname{tr}[\rho\, Q_x].$$

Likewise, the expectation value of an observable O can be computed in terms of the density operator:

$$\mathbf{E}_\rho[\text{outcome}] = \operatorname{tr}[\rho\, O].$$

In Exercise 7.1 you can also verify that if we perform a projective measurement $\{P_x\}_{x\in\Omega}$ on a quantum system in state ρ and we obtain some outcome x, then the post-measurement state should be described by the density operator

$$\rho' = \frac{P_x \rho P_x}{\operatorname{tr}[\rho\, P_x]}$$

If the ensemble consists of a only one state $|\psi\rangle$ then $\rho = |\psi\rangle\langle\psi|$. In this case we call ρ (or $|\psi\rangle$) a *pure state*; it is also common to write $\psi = |\psi\rangle\langle\psi|$. We already know this terminology and notation from the previous lectures. We record some useful characterizations of pure states:

Lemma 7.1. *For a quantum state ρ, the following are equivalent:*

(a) *ρ is pure,*
(b) $\operatorname{rk}\rho = 1$,
(c) *the eigenvalues of ρ are* $\{1, 0, \ldots, 0\}$,
(d) $\rho^2 = \rho$
(e) *the so-called purity* $\operatorname{tr}[\rho^2]$ *is equal to one.*

All but the last are easy to see; for the last see Exercise 7.2.

If ρ is not pure, it is called a *mixed state* (but this term is also used to mean "not necessarily pure"). In particular, every quantum system has a *maximally mixed state*, which is defined by

$$\tau_{\mathcal{H}} := \frac{I}{\dim \mathcal{H}}.$$

It is the analog of a uniform distribution in probability theory.

Every density operator arises from some ensemble of pure quantum states. For example, we can always write $\rho = \sum_{j=1}^{d} p_j\, |\phi_j\rangle\langle\phi_j|$, where $d = \dim(\mathcal{H})$, $\{|\phi_j\rangle\}$ is an eigenbasis, and $\{p_j\}$ are the corresponding eigenvalues. However, there are infinitely many other ways of writing a mixed state in terms of an ensemble. For example, take any two non-orthogonal states of a qubit, say $|0\rangle$ and $|+\rangle$, and consider their uniform mixture:

$$\rho = \frac{1}{2}\big(|0\rangle\langle 0| + |+\rangle\langle +|\big) = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}.$$

Observe that neither are $|0\rangle$ and $|+\rangle$ eigenvectors of ρ, nor are $\{1/2, 1/2\}$ the eigenvalues (for otherwise ρ would be equal to the maximally mixed state).

Density operators are not only useful to describe quantum information sources, but they arise in many other situations. In physics, they are used to describe statistical ensembles (e.g., *Gibbs states*). Density operators also allow us to embed probability theory into quantum theory.

The idea is simple: if $\{p_x\}_{x=1}^d$ is the probability distribution of a random variable X, we can associate with it the ensemble $\{p_x, |x\rangle\}$ on $\mathbb{C}^d$ and hence the density operator

$$\rho_X = \sum_x p_x |x\rangle\langle x| = \begin{pmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_d \end{pmatrix}. \tag{7.1}$$

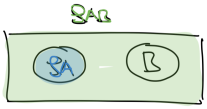
More generally, if $p(x_1, \ldots, x_n)$ is the joint probability distribution of random variables $X_1, \ldots, X_n$, the corresponding density operator is

$$\rho_{X_1 \ldots X_n} = \sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) |x_1\rangle\langle x_1| \otimes \ldots \otimes |x_n\rangle\langle x_n| . \tag{7.2}$$

We call states of the form Eqs. (7.1) and (7.2) *classical states*. Observe that if all probabilities are the same then we obtain the maximally mixed state defined earlier.

7.2 Reduced density operators and partial trace

Density operators are also useful in another situation which appears to have nothing to do with ensembles. Namely, they allow us to describe the state of subsystems if the global state is known. To see this, suppose that ρ_{AB} is a quantum state on $\mathcal{H}_A \otimes \mathcal{H}_B$, as illustrated below:



We could like to find a mathematical object (hopefully, a density operator) that describes the state of subsystem A. To this end we consider an arbitrary POVM $\{Q_{A,x}\}_{x\in\Omega}$ on $\mathcal{H}_A$. According to Axiom E, we need to consider the POVM $\{Q_{A,x} \otimes I_B\}$ when we want to perform this measurement on the joint system AB. Thus, the probability of measurement outcomes can be computed as follows:

$$\begin{aligned}
\mathbf{Pr}_{\rho_{AB}}(\text{outcome } x) &= \mathrm{tr}[\rho_{AB}\,(Q_{A,x} \otimes I_B)] \\
&= \sum_{a,b} \langle ab|\, \rho_{AB}\,(Q_{A,x} \otimes I_B)\, |ab\rangle \\
&= \sum_{a,b} \langle a|\, (I_A \otimes \langle b|)\rho_{AB}\,(Q_{A,x} \otimes I_B)(I_B \otimes |b\rangle)\, |a\rangle \\
&= \sum_{a,b} \langle a|\, (I_A \otimes \langle b|)\rho_{AB}(I_B \otimes |b\rangle)Q_{A,x}\, |a\rangle \\
&= \sum_{a} \langle a| \left(\sum_b (I_A \otimes \langle b|)\rho_{AB}(I_B \otimes |b\rangle) \right) Q_{A,x}\, |a\rangle \\
&= \mathrm{tr}\Big[\underbrace{\left(\sum_b (I_A \otimes \langle b|)\rho_{AB}(I_B \otimes |b\rangle) \right)}_{=:\rho_A} Q_{A,x} \Big]
\end{aligned}$$

The operator ρ_A just introduced is called the *reduced state* or the *reduced density operator* of ρ_{AB} on subsystem A. We also call ρ_{AB} an *extension* of ρ_A. The computation above works not just for POVM elements: for every operator N_A on $\mathcal{H}_A$, we have

$$\mathrm{tr}[\rho_{AB}(N_A \otimes I_B)] = \mathrm{tr}[\rho_A N_A]. \tag{7.3}$$

It is not hard to conclude from this that ρ_A is again density operator. As we derived above, for every POVM measurement $\{Q_{A,x}\}$ on $\mathcal{H}_A$ we have

$$\mathbf{Pr}_{\rho_{AB}}(\text{outcome } x) = \mathrm{tr}[\rho_A Q_{A,x}] = \mathbf{Pr}_{\rho_A}(\text{outcome } x)$$

Similarly, for every observable O_A on $\mathcal{H}_A$, we can compute its expectation value as

$$\mathbf{E}_{\rho_{AB}}[\text{outcome}] = \mathrm{tr}[\rho_A O_A] = \mathbf{E}_{\rho_A}[\text{outcome}]$$

Thus, the reduced state ρ_A is the appropriate object for describing the state of subsystem A if the overall state is ρ_{AB}.

To systematize the above, we define the *partial trace* of an operator M_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$ by the same formulas as above:

$$\mathrm{tr}_B[M_{AB}] = \sum_b (I_A \otimes \langle b|) M_{AB} (I_A \otimes |b\rangle).$$

In particular, $\rho_A = \mathrm{tr}_B[\rho_{AB}]$. We can also compute partial traces of operators that are *not* quantum states (but in this case we will *never* write M_A). The following rule tells us how to compute partial traces of tensor product operators, $M_{AB} = N_A \otimes O_B$:

$$\mathrm{tr}_B[N_A \otimes O_B] = N_A \,\mathrm{tr}[O_B] \tag{7.4}$$

This formula justifies the term "partial trace".[1] It follows directly from the definition,

$$\mathrm{tr}_B[M_A \otimes N_B] = \sum_b (I_A \otimes \langle b|)(M_A \otimes N_B)(I_A \otimes |b\rangle) = M_A \sum_b \langle b|N_B|b\rangle = M_A \,\mathrm{tr}[N_B].$$

and is quite useful to compute partial traces in practice. We demonstrate this in the following example, which shows that even if ρ_{AB} is a pure state, ρ_A can be mixed! This fact is an important motivation for the introduction of the density operator formalism.

Example 7.2. Consider two qubits that are in the maximally entangled ebit state

$$|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

To compute the reduced states of the individual qubits, first note that the corresponding two-qubit density operator is

$$\begin{aligned}
\rho_{AB} = |\Phi^+_{AB}\rangle\langle\Phi^+_{AB}| &= \frac{1}{2}\Big(|00\rangle + |11\rangle\Big)\Big(\langle 00| + \langle 11|\Big) \\
&= \frac{1}{2}\Big(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|\Big) \\
&= \frac{1}{2}\Big(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 0| \otimes |1\rangle\langle 0| + |0\rangle\langle 1| \otimes |0\rangle\langle 1| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|\Big).
\end{aligned}$$

We can compute the reduced state ρ_A using Eq. (7.4):

$$\rho_A = \mathrm{tr}_B[\rho_{AB}] = \frac{1}{2}\Big(|0\rangle\langle 0| + |1\rangle\langle 1|\Big) = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = \frac{I}{2},$$

noting that $\mathrm{tr}[|0\rangle\langle 0|\,0] = \mathrm{tr}[|1\rangle\langle 1|\,1] = 1$, while $\mathrm{tr}[|0\rangle\langle 1|] = \mathrm{tr}[|1\rangle\langle 0|] = 0$. Thus, the reduced state of A is the maximally mixed state, and similarly for B. Note that this matches precisely the result of our calculation in Eq. (2.1) in Chapter 2.

[1] It also characterizes the partial trace uniquely, because tr_B is a linear map from the space of operators on $\mathcal{H}_A \otimes \mathcal{H}_B$ to the space of operators on $\mathcal{H}_A$, and the former is is spanned by the tensor product operators.

We mention some other useful properties of the partial trace. The partial trace is cyclic on B:

$$\mathrm{tr}_B[(I \otimes O_B)M_{AB}] = \mathrm{tr}_B[M_{AB}(I \otimes O_B)]$$

We can pull operators on A through the partial trace:

$$\mathrm{tr}_B[(N_A \otimes I_B)M_{AB}(N'_A \otimes I_B)] = N_A \,\mathrm{tr}_B[M_{AB}]N'_B.$$

If we first perform a partial trace over one subsystem and then over the other, this is the same as performing a partial trace over both. In particular: $\mathrm{tr} \circ \mathrm{tr}_B = \mathrm{tr}$. By combining the last two properties, we obtain the following identity, which generalizes Eq. (7.3):

$$\mathrm{tr}[M_{AB}(N_A \otimes I_B)] = \mathrm{tr}[\mathrm{tr}_B[M_{AB}]\ N_A],$$

Remark 7.3. A convention that you will find in the literature is that tensor products with the identity operator are omitted. E.g., instead of $X_A \otimes I_B$ one writes X_A, since the subscript already conveys the necessary information. Using this convention, Eqs. (7.3) and (7.4) become

$$\mathrm{tr}[\rho_{AB}N_A] = \mathrm{tr}[\rho_A N_A],$$
$$\mathrm{tr}_B[N_A O_B] = N_A \,\mathrm{tr}[O_B],$$

which is possibly easier to read.

7.3 Purification and Schmidt decomposition

Above we saw that the pure ebit state has mixed reduced states. This is is not an accident.

Lemma 7.4 (Schmidt decomposition)**.** *Every pure state $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ has a so-called Schmidt decomposition:*

$$|\Psi_{AB}\rangle = \sum_{i=1}^{r} s_i \,|e_i\rangle_A \otimes |f_i\rangle_B \,,$$

where $r \in \mathbb{N}$ is called the Schmidt rank, the numbers $s_i > 0$ are called Schmidt coefficients, the $\{|e_i\rangle_A\}$ are orthonormal vectors in $\mathcal{H}_A$, and the $\{|f_i\rangle_B\}$ are orthonormal vecrtors in $\mathcal{H}_B$.

This is just a restatement of the singular value decomposition. A similar calculation as in Theorem 7.2 shows that the reduced states of $|\Psi_{AB}\rangle$ are given by

$$\rho_A = \sum_{i=1}^{r} s_i^2 \,|e_i\rangle\langle e_i| \,, \qquad \rho_B = \sum_{i=1}^{r} s_i^2 \,|f_i\rangle\langle f_i| \,. \tag{7.5}$$

Thus, the reduced states have rank r and their nonzero eigenvalues are the squares s_i^2 of the Schmidt coefficients. In particular: ρ_A and ρ_B have the *same* nonzero eigenvalues. We also see from the above that most reduced states are mixed: ρ_A and ρ_B are pure if and only if $|\Psi_{AB}\rangle$ is a product state.

We can also go the other way around. Iff ρ_A is an arbitrary density operator $\mathcal{H}_A$, then there always exists an additional Hilbert space $\mathcal{H}_B$ and a pure state $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ such that

$$\mathrm{tr}_B[|\Psi_{AB}\rangle\,\langle\Psi_{AB}|] = \rho_A.$$

We call such a state $|\Psi_{AB}\rangle$ a *purification* of ρ_A. In other words, purifications are pure states that extend the given density operator. To see that purifications always exist, consider a spectral decomposition of the density operator: $\rho_A = \sum_{i=1}^r p_i |\phi_i\rangle\langle\phi_i|$. Then,

$$|\Psi_{AB}\rangle := \sum_{i=1}^{r} \sqrt{p_i} |\phi_i\rangle_A \otimes |i\rangle_B \in \mathcal{H}_A \otimes \mathcal{H}_B \tag{7.6}$$

is a purification, where $H_B = \mathbb{C}^r$.

The existence of purifications, while not hard to prove, is an important result, as it shows that we can always replace mixed states by pure states, at the expense of adding an auxiliary Hilbert space. This is analogous to Exercise 2.3, where you showed that generalized measurements can always be implemented by ordinary measurements on a larger Hilbert space. For example, this justifies why in Chapter 3 we were allowed to only consider quantum strategies involving pure states (and observable measurements).

Purifications are not unique. We can see this already from Eq. (7.6): if we replace $|i_B\rangle$ by any other orthonormal basis, then we get another purification. However, this is essentially the only freedom that we have: any two purifications are related by a unitary or isometry:

Lemma 7.5. *Let $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and $|\Psi'_{AB'}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{B'}$ be two purifications of the same state ρ_A. If $\dim \mathcal{H}_B \leq \dim \mathcal{H}_{B'}$, then there exists an isometry $V_{B\to B'} \colon \mathcal{H}_B \to \mathcal{H}_{B'}$ such that*

$$(I_A \otimes V_{B\to B'}) |\Psi_{AB}\rangle = |\Psi'_{AB'}\rangle .$$

If $\mathcal{H}_B = \mathcal{H}_{B'}$ then V is a unitary.

You can prove this in Exercise 7.5 by using the Schmidt decomposition.

7.4 The trace distance between quantum states

In Exercise 2.4 we introduced a natural distance measure between pure states, called the *trace distance*. It can be extended to mixed states in the following way. Let ρ and σ be two density operators on some Hilbert space $\mathcal{H}$. We define their *trace distance* to be

$$T(\rho, \sigma) := \max_{0\leq Q\leq I_{\mathcal{H}}} \operatorname{tr}[Q(\rho - \sigma)] = \max_{0\leq Q\leq I_{\mathcal{H}}} |\operatorname{tr}[Q(\rho - \sigma)]|.$$

This formula generalizes the one for pure states. The equality holds because $\operatorname{tr}[(I - Q)(\rho - \sigma)] = -\operatorname{tr}[Q(\rho - \sigma)]$ and $0 \leq I - Q \leq I$. The trace distance is a metric, and so in particular satisfies the triangle inequality. It has the following alternative expression,

$$T(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1,$$

in terms of the *trace norm*, which for general Hermitian operators Δ with spectral decomposition $\Delta = \sum_i \lambda_i |e_i\rangle \langle e_i|$ is defined by $\|\Delta\|_1 = \sum_i |\lambda_i|$. The trace distance has a natural operational interpretation in terms of the optimal probability of distinguishing ρ and σ by a POVM measurement. We discussed this in Exercise 2.2 in the case of pure states, but the result described there holds in general.

The trace distance can only decrease when we trace out a system: for any two density operators ρ_{AB} and σ_{AB}, we have

$$T(\rho_A, \sigma_A) \leq T(\rho_{AB}, \sigma_{AB})$$

This is quite intuitive – it should not be easier to distinguish two states if one is only given access to a subsystem. You can prove this in Exercise 7.7. In Exercise 2.2, you also proved that for pure states $\rho = |\phi\rangle\langle\phi|$ and $\sigma = |\psi\rangle\langle\psi|$, the trace distance can be expressed in terms of the overlap:

$$T(\rho, \sigma) = \sqrt{1 - |\langle\phi|\psi\rangle|^2} \tag{7.7}$$

In Section 14.1 we will generalize the overlap to mixed states and state an analogous relation.

The trace distance also has another operational interpretation: it bounds the the difference in expectation values for any observable measurement. You can show this in Exercise 7.6.

Exercises

7.1 **Post-measurement state for density operators:** Consider a quantum system described by an ensemble of pure states $\{p_i, |\psi_i\rangle\}$, with associated density operator ρ. Suppose that we perform a projective measurement $\{P_x\}_{x\in\Omega}$.

(a) Verify that the probability of any measurement outcome is x is given by $\operatorname{tr}[\rho P_x]$.
(b) Now suppose you observe measurement outcome x. Given this outcome, compute the probability that the original state was in some specific state $|\psi_i\rangle$. *Hint: Bayes' theorem.*
(c) Given that the outcome was x, determine the ensemble of post-measurement states, and verify that the corresponding density operator is $P_x\rho P_x / \operatorname{tr}[\rho P_x]$.

7.2 **Purity:** Let ρ be a density operator on $\mathbb{C}^d$ and consider its purity $\operatorname{tr}(\rho^2)$.

(a) Show that $\operatorname{tr}[\rho^2] \in [1/d, 1]$.
(b) Show that $\operatorname{tr}[\rho^2] = 1$ if and only if ρ is a pure state.

7.3 **Bloch sphere:** This exercise gives a geometric picture of the state space of a qubit. Recall that the matrices I, X, Y, Z form a basis of the real vector space of Hermitian 2×2-matrices.

(a) Show that an operator ρ on $\mathbb{C}^2$ is a density operator if and only if it can be written in the form $\rho = \frac{1}{2}(I + r_X X + r_Y Y + r_Z Z)$ for a vector $\boldsymbol{r} = (r_X, r_Y, r_Z) \in \mathbb{R}^3$ of norm $\|\boldsymbol{r}\| \leq 1$.

Thus, set of quantum states of a qubit is a three-dimensional ball. In particular, it is convex (this is true in any dimension). The set of *pure* states is its surface, called the *Bloch sphere*:

(b) To see this, show that ρ is a pure state if and only if $\|\boldsymbol{r}\| = 1$. More generally, how is $\|\boldsymbol{r}\|$ related to the eigenvalues of ρ?
(c) Show that if U is a unitary then the density matrix $U\rho U^\dagger$ is parameterized by a vector $\boldsymbol{r}'$ of the same norm: $\|\boldsymbol{r}'\| = \|\boldsymbol{r}\|$. Conclude that for every unitary $U \in U(2)$ there exists a rotation matrix $O \in \mathrm{SO}(3)$ such that $\rho \mapsto U\rho U^\dagger$ corresponds to $\boldsymbol{r} \mapsto \boldsymbol{r}' = O\boldsymbol{r}$.

The final subproblem only relies on part (a):

(d) Re-prove Theorem 1.1 (the uncertainty relation) and show that it also holds for mixed states.

7.4 **Symmetries imply normal forms:** In this problem, you will show that quantum states that commute with U or $U^{\otimes 2}$ are tightly constrained by these symmetries. Recall that the single-qubit Hilbert space $\mathbb{C}^2$ is an irreducible representation of $\mathrm{U}(2)$.

(a) Show that if ρ is a density operator on $\mathbb{C}^2$ such that $[\rho, U] = 0$ for every unitary $U \in \mathrm{U}(2)$, then $\rho = I/2$.

While the two-qubit Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2$ is *not* irreducible, you know that it decomposes into two irreducible representations of U(2). Let $\tau_{\text{triplet}} = \Pi_2/3$ and $\tau_{\text{singlet}} = |\Psi^-\rangle\langle\Psi^-|$. As always, Π_2 denotes the projector onto the symmetric subspace, and $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|10\rangle - |01\rangle)$ denotes the singlet state.

(b) Show that if ρ is a density operator on $\mathbb{C}^2 \otimes \mathbb{C}^2$ such that $[\rho, U^{\otimes 2}] = 0$ for every $U \in \mathrm{U}(2)$, then there exists $p \in [0, 1]$ such that $\rho = p\,\tau_{\text{triplet}} + (1-p)\tau_{\text{singlet}}$.

Hint: Use Schur's lemma.

7.5 **Purifications:** In this problem, you will establish some useful facts concerning purifications that will also be helpful in the remainder of this problem set.

(a) Prove Theorem 7.5. *Hint: Use the Schmidt decomposition.*

Next, you will construct a particular purification and see how symmetries can be lifted. Let ρ_A be a density operator on a Hilbert space $\mathcal{H}_A$. For simplicity, assume that $\mathcal{H}_A = \mathbb{C}^d$.

(b) Show that $|\Psi_{AB}\rangle := (\sqrt{\rho_A} \otimes I_B) \sum_{i=1}^{d} |ii\rangle$ is a purification of ρ_A (often called the *standard purification*). Here, $\mathcal{H}_B = \mathbb{C}^d$, and $\sqrt{\rho_A}$ is the positive semidefinite square root, defined by taking the square root of each eigenvalue while keeping the same eigenspaces.

(c) Show that this purification has the following symmetry: For every unitary U_A with $[U_A, \rho_A] = 0$, we have $(U_A \otimes \bar{U}_B)\,|\Psi_{AB}\rangle = |\Psi_{AB}\rangle$. Here, $\bar{U}_B$ denotes the complex conjugate of U_A.

7.6 **Trace distance and observables:** In this problem, you can show that density operators ρ and σ with small trace distance $T(\rho, \sigma)$ have similar expectation values.

(a) Show that, for every two Hermitian operators M and N, $|\mathrm{tr}[MN]| \leq \|M\|_1 \|N\|_\infty$. Here, $\|M\|_1$ is the *trace norm* defined in class (i.e., the sum of absolute values of the eigenvalues of M) and $\|N\|_\infty := \max_{\|\phi\|=1} \|N\,|\phi\rangle\|$ is the *operator norm* (which can also be defined as the maximum of the absolute values of the eigenvalues of N).

(b) Conclude that, for every observable O, $|\mathrm{tr}[\rho O] - \mathrm{tr}[\sigma O]| \leq 2\,\|O\|_\infty\, T(\rho, \sigma)$.

(c) Find a (nonzero) observable for which the bound in part (b) is an equality.

7.7 **Monotonicity of the trace distance:** Show that $T(\rho_A, \sigma_A) \leq T(\rho_{AB}, \sigma_{AB})$ for any two states ρ_{AB} and σ_{AB}.

7.8 **Gentle measurement:** In this problem, you will derive a useful technical result known as the *gentle measurement lemma*. Let ρ be a quantum state and $0 \leq Q \leq I$ a POVM element.

(a) Show that if $\mathrm{tr}[\rho Q] \geq 1 - \varepsilon$ then $T(\rho, \frac{\sqrt{Q}\rho\sqrt{Q}}{\mathrm{tr}[\rho Q]}) \leq \sqrt{\varepsilon}$.
Hint: First prove the result for pure states.

(b) Explain in one sentence why this result is called the *gentle measurement lemma*.

Chapter 8

Entanglement of pure and mixed states, monogamy of entanglement

Yesterday, in Chapter 7 we introduced density operators, partial traces, and purifications. In particular, we learned that any bipartite pure state has a Schmidt decomposiiton. This has a number of important consequences.

8.1 Pure state entanglement

For one, it helps us to understand entanglement in pure states. For example, it shows that if $|\Psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B$ is a product state then its reduced density operators are pure. Conversely, if either of the reduced density operators of a *pure* state $|\Psi\rangle_{AB}$ is pure then $|\Psi_{AB}\rangle$ must be a product state. In other words, if ρ_A or ρ_B are mixed then this is a signature of entanglement (for pure states)! This suggests that quantities built from the eigenvalues of the reduced density operators such as the *entanglement entropy* that some of you might already know should be good entanglement measures. You will explore this further in Exercise 8.1 and we will discuss the entanglement entropy in Chapter 10.

How about if ρ_{AB} is a general density operator (not necessarily pure)? Then it is still true that

$$\rho_A \text{ pure} \quad \Rightarrow \quad \rho_{AB} = \rho_A \otimes \rho_B \tag{8.1}$$

(but ρ_B can now be mixed). To see this, choose an arbitrary purification $|\Psi_{ABC}\rangle$ of ρ_{AB}. Since $\rho_A = \mathrm{tr}_{BC}[|\Psi_{ABC}\rangle\langle\Psi_{ABC}|]$ is pure, we know from the preceding discussion that we must have

$$\Psi_{ABC} = |\psi_A\rangle \otimes |\phi_{BC}\rangle\,,$$

where $\rho_A = |\psi_A\rangle\langle\psi_A|$. But then

$$\rho_{AB} = \mathrm{tr}_C[|\Psi_{ABC}\rangle\langle\Psi_{ABC}|] = \mathrm{tr}_C[|\psi_A\rangle\langle\psi_A| \otimes |\phi_{BC}\rangle\langle\phi_{BC}|] = |\psi_A\rangle\langle\psi_A| \otimes \mathrm{tr}_C[|\phi_{BC}\rangle\langle\phi_{BC}|] = \rho_A \otimes \rho_B,$$

since necessarily $\rho_B = \mathrm{tr}_C[|\phi_{BC}\rangle\langle\phi_{BC}|]$. This is what we wanted to show.

Monogamy of entanglement is the idea that if two systems are strongly entangled then each of them cannot be entangled very much with other systems. We can get some intuition why this should be true as consequence of Eq. (8.1). For example, suppose that

$$\rho_{AB} = |\Psi\rangle\langle\Psi|_{AB}$$

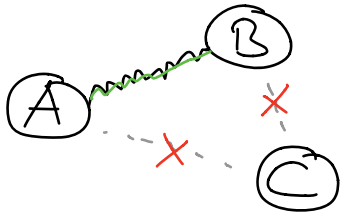


Figure 8.1: Illustration of monogamy of entanglement.

where $|\Psi\rangle_{AB}$ is in a pure state – say, a maximally entangled state. Since ρ_{AB} is pure, any extension ρ_{ABC} must factorize,

$$\rho_{ABC} = \rho_{AB} \otimes \rho_C,$$

as implied by Eq. (8.1) (with $A = AB$ and $B = C$). Thus, A and B are both completely uncorrelated with C (Fig. 8.1). In particular, $\rho_{AC} = \rho_A \otimes \rho_C$ and $\rho_{BC} = \rho_B \otimes \rho_C$ are product states.

Remark 8.1. The above analysis should perhaps be taken with a grain of salt. Since it only relied on ρ_{AB} being in a pure state, it is also applicable to, say, $\psi_{AB} = |0\rangle_A \otimes |0\rangle_B$ – which is a product state, not an entangled state! Nevertheless, the conclusion remains that also in this case ρ_{AC} and ρ_{BC} have to product states. However, this is a consequence of $\rho_A = |0\rangle\langle 0|_A$ and $\rho_B = |0\rangle\langle 0|_B$ being pure, not of entanglement between A and B.

Does monogamy hold more generally for mixed states and can it be made quantitative? Indeed this is possible – and we will see that symmetry is the key.

8.2 Mixed state entanglement

First, though, we have to define what it means for a general quantum state to be entangled. For pure states $|\Psi_{AB}\rangle$, we already know that a state is entangled if and only if it is *not* a tensor product,

$$|\psi\rangle_{AB} \neq |\psi\rangle_A \otimes |\psi\rangle_B .$$

For mixed states, however, there are non-product quantum states that should nevertheless not be considered entangled.

Example 8.2 (Classical joint distributions)**.** Let $p(x, y)$ be a probability distribution of two random variables. Following (7.2), we construct a corresponding density operator

$$\rho_{AB} = \sum_{x,y} p(x, y)\, |x\rangle\langle x|_A \otimes |y\rangle\langle y|_B .$$

In general, ρ_{AB} is not a product state (indeed, ρ_{AB} is a product state precisely when the two random variables are independent). For example, if Alice and Bob know the outcome of a fair coin flip, their state would be described by the density operator

$$\rho_{AB} = \frac{1}{2}\left(|00\rangle\langle 00|_{AB} + |11\rangle\langle 11|_{AB}\right),$$

that is not of product form. However, the "non-productness" in ρ_{AB} corresponds to classical correlations, so we do *not* want to think of ρ_{AB} as being entangled.

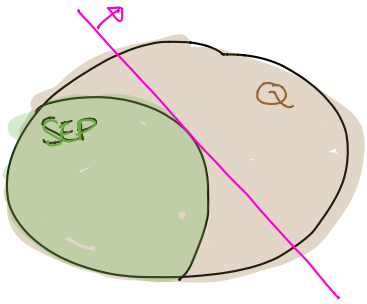


Figure 8.2: The set of separable states SEP is a convex subset of the set of all quantum states Q. Hyperplanes (such as the pink one) that contain all separable states on one side give rise to entanglement witnesses.

This suggests the following general definition: We say that a quantum state ρ_{AB} is *entangled* if and only if it is *not* a mixture of product states:

$$\rho_{AB} \neq \sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)}. \tag{8.2}$$

Here, $\{p_i\}$ is an arbitrary probability distribution and the $\rho_A^{(i)}$ and $\rho_B^{(i)}$. States of the right-hand side form are called *separable* or simply *unentangled*. If $\rho_{AB} = |\psi\rangle\langle\psi|_{AB}$ is a pure state then ρ_{AB} it is separable exactly if it is a tensor product, $|\psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B$, so this generalizes our definition of entanglement for pure states.

Remark 8.3. There are separable states other than the classical states in Theorem 8.2. This is because we do not demand the operators $\{\rho_A^{(i)}\}$ and $\{\rho_B^{(i)}\}$ in Eq. (8.2) are orthogonal.

Remark 8.4. Separable states have a pleasant operational interpretation. They are the largest class of quantum states σ_{AB} that can be created by Alice and Bob in their laboratories if allow Alice and Bob to perform arbitrary quantum operations in their laboratory but restrict their communication with each other to be classical.

Let us denote the set of all density operators on $\mathcal{H}_A \otimes \mathcal{H}_B$ by

$$Q_{AB} = \{\rho_{AB} : \rho_{AB} \geq 0, \operatorname{tr} \rho_{AB} = 1\}$$

and the subset of separable states by

$$SEP_{AB} = \{\rho_{AB} \text{ separable}\}.$$

Both sets are *convex*. As a consequence of SEP_{AB} being convex, it can be faithfully defined by a collection of separating hyperplanes, i.e., hyperplanes that contain all separable state on one side (Fig. 8.2). Any such hyperplane gives rise to an *entanglement witness* – a one-sided test that can be used to certify that a state is entangled. Formally, an *entanglement witness* for a quantum state ρ_{AB} is an observable O_{AB} such that $\operatorname{tr}[O_{AB}\,\rho_{AB}] > 0$, while $\operatorname{tr}[O_{AB}\,\sigma_{AB}] \leq 0$ for every separable state σ_{AB}. You will explore this in Exercise 8.5.

On the other hand, testing whether an arbitrary quantum state ρ_{AB} is separable or entangled is unfortunately a very difficult problem. In fact, deciding if a given density operator (given in terms of all its matrix elements) is separable is an NP-*hard* problem [Gur03]! This means that we are unlikely to ever find an efficient (as in, polynomial-time) algorithm. In practice, the situation is less bleak since we have ways of testing whe a quantum state is approximately separable (see below).

Figure 8.3: (a) In a permutation symmetric state, any pair of particles is entangled in the same way and should therefore not be entangled very much. (b) Similarly, if Alice is entangled with many Bobs in the same way then she is not entangled very much with each of them.

8.3 Monogamy and symmetry

We are now ready to study the monogamy of entanglement in more detail. We will consider two situations where we would expect monogamy to play a role:

De Finetti theorem

First, consider a permutation-symmetric state

$$|\Psi\rangle_{A_1\dots A_N} \in \mathrm{Sym}^N(\mathbb{C}^d).$$

Note that all the reduced density matrices $\rho_{A_iA_j}$ are the same. Thus, any particle is equally entangled with any other particle, and so we would expect that by monogamy each pair is therefore not "very much" entangled at all (Fig. 8.3, (a)).

The *quantum de Finetti theorem* asserts that our expectation is indeed correct:

$$\rho_{A_1\dots A_k} \approx \int d\psi\, p(\psi)\, |\psi\rangle^{\otimes k} \langle\psi|^{\otimes k} \tag{8.3}$$

as long as $k \ll n/d$, where $k+n=N$. Here, $p(\psi)$ is some probability density over the set of pure states that depends on the state ρ. In particular, $\rho_{A_1A_2}$ is approximately a mixture of product states for large n. To make "$\approx$" precise, we can use the trace distance as a distance measure (Section 7.4).

Example 8.5 (Warning)**.** The GHZ state $|\gamma\rangle_{A_1A_2A_3} = (|000\rangle + |111\rangle)/\sqrt{2}$ is a state in the symmetric subspace $\mathrm{Sym}^3(\mathbb{C}^2)$. Note that, e.g., the first particle is maximally entangled with the other two – so clearly it is *not* true that permutation symmetric states are unentangled. However, if we look at the reduced state of two particles then we find

$$\rho_{A_1A_2} = \frac{1}{2}\left(|00\rangle\langle 00| + |11\rangle\langle 11|\right) = \frac{1}{2}|0\rangle^{\otimes 2}\langle 0|^{\otimes 2} + \frac{1}{2}|1\rangle^{\otimes 2}\langle 1|^{\otimes 2},$$

which is a mixture (not a superposition) of product states. This example shows that the partial trace is indeed necessary.

Permutation symmetric states arise naturally in *mean-field systems*. The ground state $|E_0\rangle$ of a mean-field Hamiltonian $H = \sum_{1\le i<j\le n} h_{ij}$ is necessarily in the symmetric subspace – provided that the ground space is nondegenerate and that n is larger than the single-particle Hilbert space. Thus, the de Finetti theorem shows that, locally, ground states of mean field systems look like mixtures of product states – a property that is highly useful for their analysis. You will explore this in more detail in Exercise 8.3.

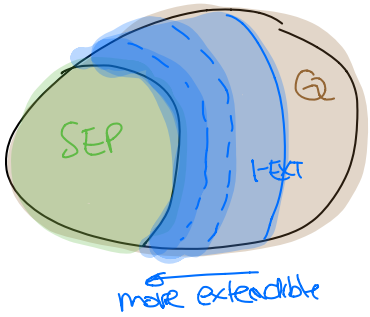


Figure 8.4: The extendibility hierarchy: If a state is n extendible then it is $O(1/n)$-close to being separable.

Extendibility hierarchy

A closely related situation is the following: Suppose that ρ_{AB} is a quantum state that has an extension $\rho_{AB_1\dots B_n}$ such that

$$\rho_{AB_i} = \rho_{AB} \qquad (\forall i, j)$$

(Fig. 8.3, (b)). We say that ρ_{AB} has an *n-extension*. Thus A is equally entangled with all B_i and so we would expect that ρ_{AB} is not entangled "very much". Indeed, it is true that, for large n,

$$\rho_{AB} \approx \sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)},$$

i.e., ρ_{AB} is again approximately a mixture of product states.

In contrast to situation (1), however, there is no longer a symmetry requirement between A and B, i.e., this reasoning applies to general states ρ_{AB}. It turns out that one in this way obtains a hierarchy of efficient approximates test for separability [DPS02, DPS04] (cf. [NOP09, HNW17]). Indeed, as you can discuss in Exercise 8.6, if a state ρ_{AB} is n-extendible then it is $O(1/n)$-close to being a separable state (Fig. 8.4).

8.4 The quantum de Finetti theorem

We will now prove the finite quantum de Finetti theorem [KR05], which establishes (8.3) in the following precise form:

Theorem 8.6 (Quantum de Finetti theorem for states on symmetric subspace)**.** *Let $|\Phi\rangle_{A_1\dots A_N} \in \mathrm{Sym}^N(\mathbb{C}^d)$ be a state on the symmetric subspace, $\rho = |\Phi\rangle\langle\Phi|$, and $N = k + n$. Then*

$$T(\rho_{A_1\dots A_k}, \int d\psi\; p(\psi)\; |\psi\rangle^{\otimes k} \langle\psi|^{\otimes k}) \leq \sqrt{\frac{dk}{n}},$$

where $p(\psi)$ is a probability density on the space of pure states on $\mathbb{C}^d$ (which depends on $|\Phi\rangle$).

Proof. We follow the proof strategy in [BCHW16]. Let

$$|\Phi\rangle_{A_1\dots A_N} \in \mathrm{Sym}^N(\mathbb{C}^d),$$

where N is the number of particles and d the dimension of the single-particle Hilbert space.

The basic idea is the following: Suppose that we measure with the uniform POVM (4.10) on the last $n := N - k$ systems of $\rho = |\Phi\rangle\langle\Phi|$. Then, if the measurement outcome is some $|\psi\rangle$, we

would expect that the first k systems are likewise in the state $|\psi\rangle^{\otimes k}$, at least on average, since the overall state is permutation symmetric among all n subsystems.

Let us try to implement this idea. Since $|\Phi\rangle \in \mathrm{Sym}^N(\mathbb{C}^d)$, it is in particular symmetric under permutations of the last $n = N - k$ subsystems. Hence, $|\Phi\rangle = (I_k \otimes \Pi_n)|\Phi\rangle$, and so

$$\begin{aligned}\rho_{A_1\ldots A_k} &= \mathrm{tr}_{A_{k+1}\ldots A_N}\left[|\Phi\rangle\langle\Phi|\right] = \mathrm{tr}_{A_{k+1}\ldots A_N}\left[(I_k \otimes \Pi_n)|\Phi\rangle\langle\Phi|\right] \\ &= \binom{n+d-1}{n}\int d\psi\,(I_k \otimes \langle\psi|^{\otimes n})|\Phi\rangle\langle\Phi|(I_k \otimes |\psi\rangle^{\otimes n}) = \int d\psi\, p(\psi)\,|V_\psi\rangle\langle V_\psi|\,.\end{aligned}$$

In the second to last step, we have inserted the resolution of identity (4.9), and in the last step, we have introduced introduced unit vectors $|V_\psi\rangle$ and numbers $p(\psi) \geq 0$ such that

$$\sqrt{p(\psi)}\,|V_\psi\rangle = \binom{n+d-1}{n}^{1/2}(I_k \otimes \langle\psi|^{\otimes n})|\Phi\rangle\,. \tag{8.4}$$

Note that $p(\psi)$ is a probability density. Indeed, $\int d\psi\, p(\psi) = \mathrm{tr}\,\rho = 1$, since the overall state is normalized. We would now like to prove that

$$\rho_{A_1\ldots A_k} = \int d\psi\, p(\psi)\;|V_\psi\rangle\langle V_\psi| \approx \int d\psi\, p(\psi)\;|\psi\rangle^{\otimes k}\langle\psi|^{\otimes k} =: \widetilde{\rho}_{A_1\ldots A_k}, \tag{8.5}$$

based on the intuition expressed above that on average the post-measurement states $|V_\psi\rangle$ are close to $|\psi\rangle^{\otimes k}$. Let us first consider the average squared overlap:

$$\begin{aligned}\int d\psi\, p(\psi)\,|\langle V_\psi|\psi^{\otimes k}\rangle|^2 &= \int d\psi\, p(\psi)\;\langle V_\psi|\psi^{\otimes k}\rangle\langle\psi^{\otimes k}|V_\psi\rangle \\ &= \binom{n+d-1}{n}\int d\psi\;\langle\Phi|\psi^{\otimes(n+k)}\rangle\langle\psi^{\otimes(n+k)}|\,|\Phi\rangle\rangle \\ &= \binom{n+d-1}{n}\binom{n+k+d-1}{n+k}^{-1}\underbrace{\langle\Phi|\Pi_{n+k}|\Phi\rangle}_{=1} \\ &= \binom{n+d-1}{n}\binom{n+k+d-1}{n+k}^{-1} \geq 1 - \frac{kd}{n}.\end{aligned}$$

In the second step, we inserted the definition of $|V_\psi\rangle$ from Eq. (8.4). Then we applied formula (4.9) to remove the integral, and the last inequality is precisely part (a) of Exercise 4.1. This is (almost) the desired result – the average squared overlap is close to one as long as $n \gg kd$.

It remains to show that the two states ρ and $\widetilde{\rho}$ in Eq. (8.5) are also close in trace distance. Indeed,

$$\begin{aligned}T(\rho_{A_1\ldots A_k}, \widetilde{\rho}_{A_1\ldots A_k}) &= \frac{1}{2}\|\rho_{A_1\ldots A_k} - \widetilde{\rho}_{A_1\ldots A_k}\| \\ &\leq \int d\psi\, p(\psi)\frac{1}{2}\|\rho_{A_1\ldots A_k} - \widetilde{\rho}_{A_1\ldots A_k}\| \\ &= \int d\psi\, p(\psi)\, T\left(|V_\psi\rangle\langle V_\psi|, |\psi\rangle^{\otimes k}\langle\psi|^{\otimes k}\right) \\ &= \int d\psi\, p(\psi)\,\sqrt{1 - |\langle V_\psi|\psi^{\otimes k}\rangle|^2} \\ &\leq \sqrt{\int d\psi\, p(\psi)\;(1 - |\langle V_\psi|\psi^{\otimes k}\rangle|^2)}\end{aligned}$$

$$= \sqrt{1 - \int d\psi \, p(\psi) \, |\langle V_\psi | \psi^{\otimes k} \rangle|^2} \leq \sqrt{\frac{kd}{n}}.$$

Here, we first applied the triangle inequality, then we used the relationship between trace distance and fidelity for pure states from Eq. (7.7), and the next inequality is Jensen's inequality (for the square root function, which is concave). (Jensen's inequality for a *concave* function f asserts that $E[f(X)] \leq f(E[X])$ for any random variable X.) Thus we have proved the quantum de Finetti theorem. □

In Exercises 8.3 and 8.4 you will explore some applications of the theorem.

Remark 8.7. From our proof we also obtain an explicit form for the density $p(\psi)$, namely $p(\psi) = \langle \Phi | I_k \otimes Q_\psi | \Phi \rangle$, where $\{Q_\psi\}$ is the uniform POVM (4.10).

Beyond the symmetric subspace

Our intuition behind the de Finetti theorem only relied on the fact that the reduced density matrices were all the same. But this is a feature that states on the symmetric subspace share with arbitrary *permutation-invariant* states, i.e., states that satisfy

$$[R_\pi, \rho_{A_1 \dots A_N}] = 0, \quad \text{or} \quad R_\pi \rho_{A_1 \dots A_N} = \rho_{A_1 \dots A_N} R_\pi$$

for all $\pi \in S_N$. Examples of permutation-invariant states are states on the *anti*symmetric subspace, or tensor powers of mixed states, such as $\rho^{\otimes N}$, which we will study in more detail in Chapter 13.

To obtain a de Finetti theorem for this situation, it is useful to prove that any permutation-invariant state $\rho_{A_1 \dots A_N}$ has a purification on a symmetric subspace: That is, there exists a pure state $|\Phi\rangle_{(A_1 B_1) \dots (A_N B_N)} \in \mathrm{Sym}^n(\mathcal{H}_A \otimes \mathcal{H}_B)$, where $\mathcal{H}_B$ is some auxiliary space, such that $\rho_{(A_1 B_1) \dots (A_N B_N)} = |\Phi\rangle\langle\Phi|$ is an extension of $\rho_{A_1 \dots A_N}$. The auxiliary space $\mathcal{H}_B$ can be chosen of the same dimension as $\mathcal{H}_A$. (You see an easy example of this in Exercise 8.4.) The point is that we can now apply the quantum de Finetti theorem proved above to the purification!

Following this strategy, you will prove in Exercise 8.2 the following version of the quantum de Finetti theorem:

Theorem 8.8 (Quantum de Finetti theorem for permutation-invariant states)**.** *Let $\rho_{A_1 \dots A_N}$ be a permutation-invariant quantum state on $(\mathbb{C}^d)^{\otimes N}$ and $N = k + n$. Then*

$$T(\rho_{A_1 \dots A_k}, \int d\mu(\sigma) \, \sigma^{\otimes k}) \leq \sqrt{\frac{d^2 k}{n}},$$

where $d\mu(\sigma)$ is a probability measure on the space of density operators on $\mathbb{C}^d$ (which depends on ρ).

Nowadays, there are many further variants of the de Finetti theorem that quantify the monogamy of entanglement in interesting and useful ways.

Exercises

8.1 **Pure state entanglement:** In class we observed that a pure state $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is *unentangled* if and only if its reduced density operators ρ_A and ρ_B are pure states. Here you will generalize this observation and show that the maximal fidelity squared between $|\Psi_{AB}\rangle$

and any product state is given by the largest eigenvalue of ρ_A, denoted $\lambda_{\max}(\rho_A)$. That is, show that

$$\max_{\|\phi_A\|=\|\psi_B\|=1} |\langle \Psi_{AB} | \phi_A \otimes \psi_B \rangle|^2 = \lambda_{\max}(\rho_A).$$

Hint: Use the Schmidt decomposition discussed.

8.2 **De Finetti theorem for permutation-invariant quantum states:** In this problem, you will extend the quantum de Finetti theorem from states on the symmetric subspace to arbitrary permutation-invariant states. A quantum state $\rho_{A_1 \dots A_N}$ is called *permutation-invariant* if $[R_\pi, \rho_{A_1 \dots A_N}] = 0$ for all $\pi \in S_N$.

(a) Give two examples of permutation-invariant quantum states that are not just states on the symmetric subspace.

Now let $\rho_{A_1 \dots A_N}$ be an arbitrary permutation-invariant quantum state on $(\mathbb{C}^d)^{\otimes N}$.

(b) Show that the reduced density operators for any fixed number of subsystems are all the same. That is, show that $\rho_{A_{i_1} \dots A_{i_k}} = \rho_{A_1 \dots A_k}$ for all $1 \leq k \leq N$ and pairwise distinct indices $i_1, \dots, i_k$.

By monogomy, we would therefore expect that a de Finetti theorem should also hold in this situation. You will prove this in the remainder of this exercise:

(c) Show that there exists a pure state $\rho_{(A_1 B_1) \dots (A_N B_N)}$ on $\mathrm{Sym}^N(\mathbb{C}^d \otimes \mathbb{C}^d) \subseteq (\mathbb{C}^d \otimes \mathbb{C}^d)^{\otimes N}$ such that $\rho_{A_1 \dots A_N} = \mathrm{tr}_{B_1 \dots B_N}[\rho_{(A_1 B_1) \dots (A_N B_N)}]$.

(d) Conclude that, for every $1 \leq k \leq N$, there exists a probability measure $d\mu$ on the set of density operators on $\mathbb{C}^d$ such that $T(\rho_{A_1 \dots A_k}, \int d\mu(\rho)\, \rho^{\otimes k}) \leq \sqrt{d^2 k/n}$, where $n = N - k$.

8.3 **De Finetti and mean field theory [BCHW16]:** In this exercise you will explore the consequences of the quantum de Finetti theorem for mean field theory. Consider a Hermitian operator h on $\mathbb{C}^d \otimes \mathbb{C}^d$ and the corresponding *mean-field Hamiltonian*, i.e., the operator

$$H = \frac{1}{n-1} \sum_{i \neq j} h_{i,j}$$

on $(\mathbb{C}^d)^{\otimes n}$, where each term $h_{i,j}$ acts by the operator h on subsystems i and j and by the identity operator on the remaining subsystems (e.g., $h_{1,2} = h \otimes I^{\otimes(n-2)}$).

(a) Show that the eigenspaces of H are invariant subspaces for the action of the symmetric group.

Now assume that the eigenspace with minimal eigenvalue (the so-called *ground space*) is nondegenerate and spanned by some $|E_0\rangle$, with corresponding eigenvalue E_0. Then part (a) implies that $R_\pi |E_0\rangle = \chi(\pi) |E_0\rangle$ for some function χ. This function necessarily satisfies $\chi(\pi\tau) = \chi(\pi)\chi(\tau)$.

(b) Show that $\chi(i \leftrightarrow j) = \chi(1 \leftrightarrow 2)$ for all $i \neq j$. Conclude that $|E_0\rangle$ is either a symmetric tensor or an antisymmetric tensor.
Hint: First show that $\chi(\pi\tau\pi^{-1}) = \chi(\tau)$.

If $n > d$, then there exist no nonzero antisymmetric tensors. Thus, in the so-called *thermodynamic limit* of large n, the ground state $|E_0\rangle$ is in the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^d)$ and so the quantum de Finetti theorem is applicable.

(c) Show that, for large n, the energy density in the ground state can be well approximated by minimizing over tensor power states. That is, show that

$$\frac{E_0}{n} \approx \min_{|\psi\rangle} \langle \psi^{\otimes 2} | h | \psi^{\otimes 2} \rangle = \frac{1}{n} \min_{|\psi\rangle} \langle \psi^{\otimes n} | H | \psi^{\otimes n} \rangle \,.$$

Hint: Exercise 7.6.

This justifies the folklore that "in the mean field limit the ground state has the form $|\psi\rangle^{\otimes \infty}$".

8.4 **The antisymmetric state:** In class, we discussed the quantum de Finetti theorem for the symmetric subspace. It asserts that the reduced density operators $\rho_{A_1 \dots A_k}$ of a state on $\mathrm{Sym}^{k+n}(\mathbb{C}^D)$ are $\sqrt{kD/n}$ close in trace distance to a separable state (in fact, to a mixture of tensor power states).

The goal of this exercise is to show that some kind of dependence on the dimension D is unavoidable in the statement of the theorem. To start, consider the *Slater determinant*

$$|S\rangle_{A_1 \dots A_d} = |1\rangle \wedge \cdots \wedge |d\rangle := \sqrt{\frac{1}{d!}} \sum_{\pi \in S_d} \mathrm{sign}(\pi) \, |\pi(1)\rangle \otimes \ldots \otimes |\pi(d)\rangle \in (\mathbb{C}^d)^{\otimes d}.$$

We define the *antisymmetric state* on $\mathbb{C}^d \otimes \mathbb{C}^d$ by tracing out all but two subsystems,

$$\rho_{A_1 A_2} = \mathrm{tr}_{A_3 \dots A_d} \left[|S\rangle \langle S| \right].$$

(a) Let $F = R_{1 \leftrightarrow 2}$ denote the swap operator on $(\mathbb{C}^d)^{\otimes 2}$. Prove the following identity, which is known as the *swap trick*:

$$\mathrm{tr}[F(\sigma \otimes \gamma)] = \mathrm{tr}[\sigma \gamma]$$

(b) Show that $T(\rho_{A_1 A_2}, \sigma_{A_1 A_2}) \geq \frac{1}{2}$ for all separable states $\sigma_{A_1 A_2}$.
Hint: Consider the POVM element $Q = \Pi_2$ (i.e., the projector onto the symmetric subspace).

Thus you have shown that the antisymmetric state is far from any separable state. However, note that $|S\rangle$ is *not* in the symmetric subspace.

(c) Show that $|S\rangle^{\otimes 2} \in \mathrm{Sym}^d(\mathbb{C}^d \otimes \mathbb{C}^d)$, while $\rho_{A_1 A_2}^{\otimes 2}$ is likewise far away from any separable state. Conclude that the quantum de Finetti theorem must have some dimension dependence.
Hint: $|S\rangle^{\otimes 2}$ is a state of $2d$ quantum systems that we might label $A_1 \dots A_d A'_1 \dots A'_d$ (the unprimed systems refer to the first copy of $|S\rangle$ and the primed to the second). Let the permutation group S_d act by simultaneously permuting unprimed and primed systems and show that $|S\rangle^{\otimes 2}$ is in the corresponding symmetric subspace. Similarly, $\rho^{\otimes 2}$ is an operator on $A_1 A_2 A'_1 A'_2$. How do you need to partition the systems so that $\rho^{\otimes 2}$ is far from being separable?

8.5 **Entanglement witness for the ebit:** Construct an entanglement witness for the ebit state $|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)$.

Hint: Exercise 8.1.

8.6 **The extendibility hierarchy:** In this problem, you will show that any quantum state that has an n-extension is close to a separable state if n is large, as discussed in class.

(a) Imitate the proof of the quantum de Finetti theorem given in class to show that, for any pure state $|\Phi\rangle_{AB_1\ldots B_n} \in \mathcal{H}_A \otimes \mathrm{Sym}^n(\mathcal{H}_B)$,

$$\mathrm{tr}_{B_2\ldots B_n}[|\Phi\rangle\langle\Phi|] \approx \int d\psi\, p(\psi)\; |W_\psi\rangle\langle W_\psi|_A \otimes |\psi\rangle\langle\psi|_{B_1}$$

for large n. Here, the integral is over the set of pure states on $\mathcal{H}_B$, $p(\psi)$ is a probability density, and the $|W_\psi\rangle$ are pure states in $\mathcal{H}_A$.

Now suppose that ρ_{AB} is an arbitrary quantum state that has an n-extension (i.e., that there exists some $\sigma_{AB_1\ldots B_n}$ such that $\sigma_{AB_k} = \rho_{AB}$ for all k).

(b) Show that ρ_{AB} also has an n-extension $\rho_{AB_1\ldots B_n}$ that is permutation-invariant on the B-systems, i.e., $[I_A \otimes R_\pi, \rho] = 0$ for all $\pi \in S_n$.

Any n-extension as in (b) admits a purification in $(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \otimes \mathrm{Sym}^n(\mathcal{H}_B \otimes \mathcal{H}_{B'})$, where $\mathcal{H}_{A'} = \mathcal{H}_A$ and $\mathcal{H}_{B'} = \mathcal{H}_B$.

(c) Conclude that any n-extendible ρ_{AB} is close to a separable state for large n.
Hint: Exercise 7.7.

8.7 **PPT criterion**: In this exercise, you will study a simple yet very useful entanglement criterion. Given an operator M_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$, we define its *partial transpose* as the operator $M_{AB}^{T_B}$ with matrix elements

$$\langle a, b|M_{AB}^{T_B}|a', b'\rangle = \langle a, b'|M_{AB}|a', b\rangle\,.$$

Note that this definition depends on the choice of basis for $\mathcal{H}_B$ (but not of the basis for $\mathcal{H}_A$).

(a) Show that $\mathrm{tr}\, M_{AB}^{T_B} = \mathrm{tr}\, M_{AB}$.
(b) Observe that if $M_{AB} = X_A \otimes Y_B$ then $M_{AB}^{T_B} = X_A \otimes Y_B^T$ and argue that this uniquely determines the partial transpose.

In particular, we can consider the partial transpose of a density operator ρ_{AB}.

(c) Show that if ρ_{AB} is separable then $\rho_{AB}^{T_B} \geq 0$.

You thus obtain the so-called *PPT criterion*, short for positive partial transpose criterion: *If the partial transpose $\rho_{AB}^{T_B}$ is* not *positive semidefinite then ρ_{AB} must be entangled.*

(d) Verify using the PPT criterion that the ebit $|\Psi_2^+\rangle$ is entangled.
(e) Consider the family of *isotropic two-qubit states*,

$$\rho_{AB}(p) := p\,\tau_{\mathrm{sym}} + (1-p)\tau_{\mathrm{anti}},$$

where τ_{sym} denotes the maximally mixed state on the symmetric subspace of two qubits and $\tau_{\mathrm{anti}} = |\psi^-\rangle\langle\psi^-|$ the singlet state. For which values of $p \in [0, 1]$ does the PPT criterion establish entanglement?

In general, the PPT criterion is only a sufficient, but not a necessary criterion for entanglement. If $\dim H_A \otimes H_B > 6$, then there exist entangled states with a positive semidefinite partial transpose.

Chapter 9

Classical and quantum data compression

Today we will discuss one of the very well-known objectives of information theory: the compression of data sources. We will start with classical data compression (i.e., the compression of bitstrings), which was solved by Shannon in the late 40s. The results obtained for classical bit strings will turn out to be directly useful for solving our main problem of interest – namely, the compression of *quantum data* (i.e., strings of qubits).

9.1 Classical 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, sheshould send over one bit. Otherwise, since both outcomes are possible, she would make an error 25% of the time! See Fig. 9.1 for an illustration of the situation.

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? In other words, what is the minimal *compression rate*, i.e., the minimal rate of bits per coin flip that Alice needs to send to Bob in order to communicate the outcomes of her coin flips (with an arbitrarily small probability of error)?

A sequence of coin flips will in general be an arbitrary string of the form

`HHTHHHTHHHHHHHHHHHHHHHHHHTHHHHHHHHHHHTHH`

Let us denote by k the number of heads (`H`) in such a sequence, so that $n - k$ is the number of tails (`T`). The probability of any such sequence is given by $p^k(1-p)^{n-k}$.

What do "typical" sequences look like? If we assume that Alice' coin flips are *independent* then we would expect that heads will come up $k \approx pn$ times for large enough n. Indeed, a version of the (weak) *law of large numbers* states that, for any fixed $\varepsilon > 0$,

$$\mathbf{Pr}(|\frac{k}{n} - p| > \varepsilon) = O(\frac{1}{n}) \to 0 \tag{9.1}$$

as $n \to \infty$. Let us thus define a *typical sequence* as a sequence of n coin flips such that $|\frac{k}{n} - p| \leq \varepsilon$. (Note that this definition depends on a choice of ε, so it might make sense to speak of an ε-typical sequence instead.) In this language, Eq. (9.1) asserts that the probability that Alice receives a typical sequence goes to one in the limit of many coin flips.

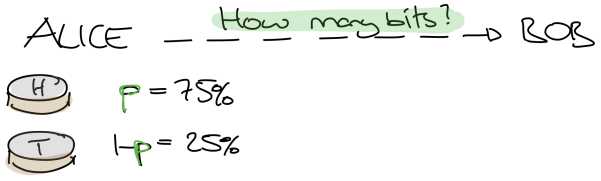


Figure 9.1: Alice wants to communicate the result of her coin flips to Bob by sending over a minimal number of bits. This an instance of a compression problem of classical data (the outcomes of Alice' coin flips).

Remark 9.1. This also gives a good way of *estimating* the bias of the coin if Alice does not know the values of p and $1-p$ beforehand. Simply flip the coin many times and output $\hat{p} := \frac{k}{n}$ as an estimate of p, where k is the number of heads. We will later learn how to similarly characterize a *quantum* data source.

This suggests the following compression scheme:

Classical data compression protocol: Let $\varepsilon > 0$ be fixed.

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

If our two protagonists agree beforehand on the lists $\mathcal{L}_k$ (you might say that they form the *codebook*), then Bob will have no trouble decoding the sequence of coin flips – he merely looks up the i-th entry in the list $\mathcal{L}_k$.

What is the probability of failure in the first step of this protocol? As a direct consequence of the law of large numbers this becomes arbitrarily small for large enough n, as we discussed above.

Remark 9.2. 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 (in terms of the *average* compression rate) and will be left as an exercise.

Is this protocol useful for compression? To send $k \in \{0, \ldots, n\}$, we need no more than $\log(n+1)$ bits. Since $\log(n+1)/n \to 0$, this does not impact the compression rate in the limit of large n. How many bits to we need to send the index i? The number of bits required depends on the number of sequences with k heads and $n-k$ tails, where $k/n \approx p$. Let us first count the number of sequences with k heads and $n-k$ tails for an arbitrary value of k. This is simply given by the binomial coefficnet $\binom{n}{k}$. To estimate this number, we use the following trick: For every $x \in [0,1]$, we have

$$1 = (x + (1-x))^n = \sum_{l=0}^{n} \binom{n}{l} x^l (1-x)^{n-l} \geq \binom{n}{k} x^k (1-x)^{n-k}.$$

Choosing $x = k/n$, we obtain the upper bound

$$\binom{n}{k} \leq x^{-k}(1-x)^{-(n-k)} = \left(\frac{k}{n}\right)^{-k} \left(1 - \frac{k}{n}\right)^{-(n-k)} = 2^{-k\log(\frac{k}{n}) - (n-k)(1-\frac{k}{n})} = 2^{nh(\frac{k}{n})}, \quad (9.2)$$

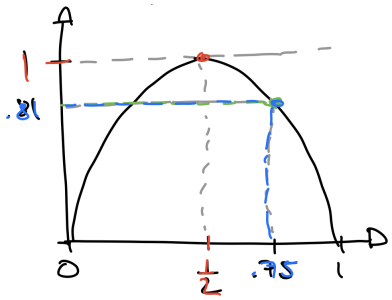


Figure 9.2: The binary entropy function $h(p)$ defined in Eq. (9.3).

where we defined the *binary (Shannon) entropy* function

$$h(p) := -p \log p - (1-p) \log(1-p). \tag{9.3}$$

Here and throughout the rest of these lecture notes, log *will always denote the logarithm to the base two.* We also define $0 \log 0 := 0$ so that $h(p)$ is a continuous function defined for all $p \in [0, 1]$. See Fig. 9.2 for a plot of the binary entropy function.

Thus, there are no more than $2^{nh(k/n)}$ many sequences with k heads and $n-k$ tails. Now, for typical sequences, $|k/n - p| \leq \varepsilon$ and so there are no more than roughly $2^{n(h(p)+\varepsilon')}$ many typical sequences for some constant $\varepsilon' > 0$ (which depends on our choice of ε and the continuity of the entropy function at p). Thus, we need no more than $n(h(p) + \varepsilon')$ bits to send over the index. In total, the compression rate of our protocol is no larger than

$$R = \frac{\#\text{ bits}}{\#\text{ coin flips}} \leq \frac{\log(n+1)}{n} + h(p) + \varepsilon'. \tag{9.4}$$

Both the first and the third term can be made arbitrarily small – the former by choosing n sufficiently large, and the latter by choosing ε sufficiently small.

In summary, the protocol sketched above will achieve a compression rate arbitrarily close to $h(p) \leq 1$ bits per coin flip. You will show in Exercise 9.2 that this compression rate $h(p)$ is optimal. The result that we proved is known as Shannon's *noiseless coding theorem* – it is called "noiseless" since we assume that the communication line from Alice to Bob is perfect. It is also known as Shannon's source coding theorem.

In our case, $h(75\%) = 0.81$ as displayed in Fig. 9.2 – so Alice achieves savings of roughly of 19% in the case of her biased coin.

Since this is a course about symmetries and information theory: *What are the symmetries in the classical data compression scenario?* One such symmetry is that the binary entropy function satisfies $h(p) = h(1-p)$, corresponding to relabeling $\texttt{H} \leftrightarrow \texttt{T}$. This is certainly expected, since merely relabeling the symbols cannot impact the optimal compression rate. However, note our compression protocol breaks this symmetry, since we explicitly compare the relative number of heads k/n to the probability p! Thus if Alice and Bob apply their compression scheme (that was designed for $p = 75\%$) to another biased coin with $p = 25\%$ then the protocol will fail with high probability in the first step. In this case there is a simple fix: We simply modify the first step of the protocol to fail only if k/n is far away from *both* p and $1-p$. It is clear that this does not impact the compression rate (we are still sending over the same information!). In Exercise 9.3 you will extend this to construct a universal classical data compression protocol at rate R that works for all data sources where $h(p) < R$.

When we discuss quantum data compression we will come back to this point and see that designing a universal quantum data compression protocol is less straightforward and requires a more careful analysis of the relevant symmetries.

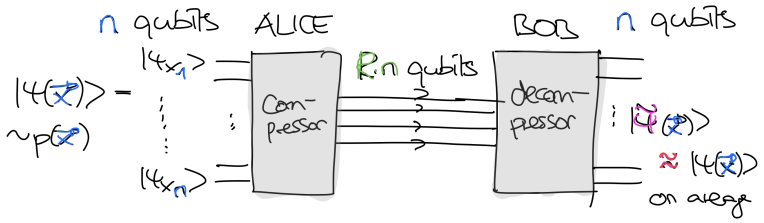


Figure 9.3: Illustration of the compression of a quantum information source.

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 we 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.

Remark 9.3. 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.

9.2 Quantum data compression

We will now discuss quantum data compression in more precise terms. Thus, we consider a *quantum information source* that emits pure states $|\psi_x\rangle \in \mathbb{C}^2$ of a qubit with probabilities p_x upon the press of a button (just like previously we obtained a random bit `H`/`T` by flipping a coin flip). We will assume that the qubit states emitted by the source are independent from each other (i.e., the source has no memory), which means that it emits sequences

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

with probabilities

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

Similarly to before, our goal in *quantum data compression* is to design a compression protocol. This protocol consists of a compressor, which encodes a sequence $|\psi(\boldsymbol{x})\rangle \in (\mathbb{C}^2)^{\otimes n}$ into some state of Rn qubits, and a corresponding decompressor. As before, we can think of R as the compression rate, but now we are sending over qubits instead of bits! Unlike in the example of the coin, we cannot in general hope to precisely recover the original state. Instead, the decompressor should produce a state $|\widetilde{\psi}(\boldsymbol{x})\rangle$ that has high overlap with the original state (say, on average):

$$\sum_{\boldsymbol{x}} p(\boldsymbol{x})\, E\left[|\langle\psi(\boldsymbol{x})|\widetilde{\psi}(\boldsymbol{x})\rangle|^2\right] \approx 1. \tag{9.5}$$

The average value $E[\ldots]$ refers to the fact that the decompressed state $|\widetilde{\psi}(\boldsymbol{x})\rangle$ for a given $|\psi(\boldsymbol{x})\rangle$ is not necessarily deterministic (since compression and decompression might involve quantum measurements, which generally have random outcomes). See Fig. 9.3 for an illustration. How could we go about solving this problem?

Let's first discuss some salient points of this setup. As discussed in Chapter 7, any ensemble such as $\{p_x, |\psi_x\rangle\}$ has a corresponding density operator $\rho = \sum_x p_x |\psi_x\rangle\langle\psi_x|$. In our case, it describes the average output of our quantum source. It is not hard to see that the density operator

corresponding to the ensemble $\{p(\boldsymbol{x}), |\psi(\boldsymbol{x})\rangle\}$, which describes n outputs of our quantum source, is given by

$$\rho^{\otimes n} = \left(\sum_x p_x |\psi_x\rangle \langle\psi_x|\right)^{\otimes n} = \sum_{\boldsymbol{x}} p(\boldsymbol{x}) |\psi(\boldsymbol{x})\rangle \langle\psi(\boldsymbol{x})| \tag{9.6}$$

It is useful to think of $\rho^{\otimes n}$ as the quantum version of an *i.i.d.* probability distribution (i.e., a probability distribution of n random variables that are independent and identically distributed). At a fundamental level, quantum information theory often reduces to questions about the asymptotic behavior of a large number of independent copies of a density operator ρ, i.e., in $\rho^{\otimes n}$ for large n (the so-called *i.i.d.* limit), similarly to what we saw for the classical coin above.

Like any density operator of a single qubit, ρ has two eigenvalues which we might denote by $\{p, 1-p\}$. We stress that the states $|\psi_x\rangle$ emitted by the source need *not* be orthogonal. This means that we *cannot* simply perform a measurement to figure out the sequence of quantum states emitted by the source, but also that the eigenvalues $\{p, 1-p\}$ of ρ need not have anything to do with the probabilities $\{p_x\}$ of the different states in the ensemble. For example, the density operator $\frac{1}{2}(|0\rangle\langle 0| + |+\rangle\langle +|)$ has eigenvalues around $\{85\%, 15\%\}$. From this perspective, it is not clear that ρ should have any significance for the compression task!

To make progress, remember that the central idea to solve classical data compression was that there was a relatively small number of *typical sequences* that occurred most of the time. In the quantum case, bits get replaced by qubits, so this suggests that we should try to look for a "small" subspace $\mathcal{H}_n \subseteq (\mathbb{C}^2)^{\otimes n}$ such that "typical" states $|\psi(\boldsymbol{x})\rangle$ have high overlap with this subspace. Let us identify on more formal level what properties this subspace should satisfy by studying the following proposal for a compression protocol:

Quantum data compression protocol: Let $\mathcal{H}_n \subseteq (\mathbb{C}^2)^{\otimes n}$, with projector P_n.

- Alice performs the projective measurement $\{P_n, I - P_n\}$. If the outcome is the latter, she sends over an arbitrary state $|\widetilde{\psi}(\boldsymbol{x})\rangle$.
- Otherwise, the post-measurement state in Alice' laboratory is

$$|\widetilde{\psi}(\boldsymbol{x})\rangle = \frac{P_n |\psi(\boldsymbol{x})\rangle}{\|P_n |\psi(\boldsymbol{x})\rangle\|} \in \mathcal{H}_n.$$

- Since this state lives in subspace $\mathcal{H}_n$ only, Alice can send it over to Bob by sending roughly $\lceil \log(\dim \mathcal{H}_n) \rceil$ qubits.
- Bob receives the state $|\widetilde{\psi}(\boldsymbol{x})\rangle$ and uses it as the decompressed state.

Remark 9.4. In step one, we send over an arbitrary state when the measurement does not "succeed" – this is not a problem since we will anyways need to inspect the average overlap squared (9.5) with the desired state. Instead, Alice could also simply fail and stop the protocol when the measurement does not succeed, just as in our classical compression protocol. Can you see how the analysis below needs to be adjusted in this case? (Exercise 7.1 could be useful.)

Remark 9.5. It might not be directly obvious how Alice and Bob can actually send over the state in the last part of the protocol. Clearly, $\dim(\mathcal{H}_n) \leq \dim(\mathbb{C}^2)^{\otimes \lceil \log(\dim \mathcal{H}) \rceil}$, so certainly $m := \lceil \log(\dim \mathcal{H}) \rceil$ qubits provide enough degrees of freedom. In practice, our two protagonists would decide on a unitary

$$U \colon (\mathbb{C}^2)^{\otimes n} \to (\mathbb{C}^2)^{\otimes n} = (\mathbb{C}^2)^{\otimes m} \otimes (\mathbb{C}^2)^{\otimes (n-m)}$$

such that any state in $\mathcal{H}_n$ gets mapped to a state into the subspace $(\mathbb{C}^2)^{\otimes m} \otimes |0\dots 0\rangle$.

In order to send over the post-measurement state, Alice would first apply U and send over the first m qubits to Bob. Upon receiving the state, Bob adds the $|0\dots 0\rangle$ back in and applies $U^\dagger$. It is clear that in this way he ends up with the state $|\widetilde{\psi}(\boldsymbol{x})\rangle$ in his laboratory.

Let us analyze the compression protocol to determine the properties that the subspace $\mathcal{H}_n$ should satisfy. Clearly, the compression rate that it achieves is

$$\frac{\log(\dim \mathcal{H}_n)}{n} \leq 1,$$

so we would like to minimize the dimension of $\mathcal{H}_n$. We will now analyze when the average overlap squared is close to one, as in (9.5): First, let us denote by

$$q(\boldsymbol{x}) := \mathbf{Pr}_{\psi(\boldsymbol{x})}(\text{outcome } P_n) = \langle\psi(\boldsymbol{x})|P_n|\psi(\boldsymbol{x})\rangle = \operatorname{tr}\left[|\psi(\boldsymbol{x})\rangle\langle\psi(\boldsymbol{x})| P_n\right] \tag{9.7}$$

the probability of passing the first step of the protocol if the state emitted by the source is $|\psi(\boldsymbol{x})\rangle$ (we used Born's rule). Then,

$$\begin{aligned}
&\sum_{\boldsymbol{x}} p(\boldsymbol{x})\, E\left[|\langle\psi(\boldsymbol{x})|\widetilde{\psi}(\boldsymbol{x})\rangle|^2\right] \\
&= \sum_{\boldsymbol{x}} p(\boldsymbol{x}) \left[q(\boldsymbol{x})|\langle\psi(\boldsymbol{x})|\frac{P_n|\psi(\boldsymbol{x})\rangle}{\|P_n|\psi(\boldsymbol{x})\rangle\|}|^2 + \dots\right] \\
&\geq \sum_{\boldsymbol{x}} p(\boldsymbol{x}) \left[q(\boldsymbol{x})|\langle\psi(\boldsymbol{x})|\frac{P_n|\psi(\boldsymbol{x})\rangle}{\|P_n|\psi(\boldsymbol{x})\rangle\|}|^2\right] \\
&= \sum_{\boldsymbol{x}} p(\boldsymbol{x}) \left[q(\boldsymbol{x})\frac{|\langle\psi(\boldsymbol{x})|P_n|\psi(\boldsymbol{x})\rangle^2|}{\|P_n|\psi(\boldsymbol{x})\rangle\|^2}\right] \\
&= \sum_{\boldsymbol{x}} p(\boldsymbol{x}) \left[q(\boldsymbol{x})\frac{q^2(\boldsymbol{x})}{q(\boldsymbol{x})}\right] \\
&= \sum_{\boldsymbol{x}} p(\boldsymbol{x}) q^2(\boldsymbol{x}) \\
&\geq \left(\sum_{\boldsymbol{x}} p(\boldsymbol{x}) q(\boldsymbol{x})\right)^2.
\end{aligned}$$

In the second line, "..." stands for the term that corresponds to the case where we abort after the first step; we simply lower bound this term by zero. The last step is Jensen's inequality for the (convex) square function. But note that

$$\sum_{\boldsymbol{x}} p(\boldsymbol{x}) q(\boldsymbol{x}) = \sum_{\boldsymbol{x}} p(\boldsymbol{x}) \operatorname{tr}\left[|\psi(\boldsymbol{x})\rangle\langle\psi(\boldsymbol{x})| P_n\right] = \operatorname{tr}\left[\rho^{\otimes n} P_n\right]$$

where we used Eqs. (9.6) and (9.7). Thus, we need that $\operatorname{tr}\left[\rho^{\otimes n} P_N\right] \approx 1$ in order for the compression protocol to achieve high fidelity in the sense of Eq. (9.5).

We thus obtain the following important result: Quantum compression is possible at rate R if we can find a sequence of subspaces $\mathcal{H}_n \subseteq (\mathbb{C}^2)^{\otimes n}$, with projectors P_n, such that

(a) $\operatorname{tr}\left[\rho^{\otimes n} P_n\right] \to 1$,
(b) $\frac{1}{n}\log(\dim \mathcal{H}_n) \leq R$.

Such subspaces are called *typical subspaces*, in analogy with the typical sequences in the classical case. Note that this condition only depends on the quantum data source in a weak way, namely through the density operator ρ. In particular, our compression protocol will work for every ensemble described by this density operator.

Tomorrow we will discuss how to construct typical subspaces that allow us to compress arbitrarily close to the optimal asymptotic rate. This rate will again be an entropy – namely, the so-called *von Neumann entropy* of the density operator ρ.

Exercises

9.1 **Fannes inequality:**

(a) Consider the function $\eta(x) = -x \log x$. Show that, for $|p - q| \leq \frac{1}{2}$,

$$|\eta(p) - \eta(q)| \leq \eta(|p - q|), \tag{9.8}$$

(b) Conclude that the binary entropy function $h(p) = -p \log p - (1 - p) \log(1 - p)$ satisfies the following inequality, which is a special case of the so-called *Fannes' inequality*:

$$|h(p) - h(q)| \leq h(|p - q|)$$

9.2 **Classical data compression:** In this exercise you will show that the Shannon entropy $h(p) = -p \log p - (1 - p) \log(1 - p)$ is the optimal compression rate for the coin flip problem discussed in class. Assume that Alice compresses her random sequence of n coin flips by applying a function $\mathcal{E}_n \colon \{H, T\}^n \to \{0, 1\}^{\lfloor nR \rfloor}$, and Bob decompresses by applying a corresponding function $\mathcal{D}_n \colon \{0, 1\}^{\lfloor nR \rfloor} \to \{H, T\}^n$.

(a) Which are the coin flip sequences that are transmitted correctly? Find an upper bound on their cardinality in terms of R.

(b) Show that, if $R < h(p)$, then the probability of success tends to zero for large n.
Hint: Distinguish between typical and atypical sequences of coin flips.

9.3 **Universal classical data compression:** Given $R > 0$, construct a data compression protocol at asymptotic rate R that works for every classical data source that emits bits with probabilities $\{p, 1 - p\}$ such that $h(p) < R$.

Chapter 10

Construction of typical subspace, compression and entanglement

Yesterday, we discussed the compression of classical and quantum data sources. Let us briefly revisit the results. We first studied classical data sources that emits bits (coin flips) with probabilities p and $1-p$ and found that the optimal compression rate is given by the Shannon entropy $h(p) = -p\log p - (1-p)\log(1-p)$. To achieve this, we restricted our consideration to *typical sequences* $\boldsymbol{b} = b_1 \dots b_n \in \{0,1\}^n$, with $k = n(p \pm \varepsilon)$ zeros (heads) for some fixed $\varepsilon > 0$. By the law of large numbers,

$$\mathbf{Pr}(\boldsymbol{b} \text{ typical}) \to 1, \tag{10.1}$$

and we found that there were at most

$$\sum_{k:|\frac{k}{n}-p|\leq\varepsilon} 2^{nh(k/n)} \leq (n+1)2^{n(h(p)+\varepsilon')} \tag{10.2}$$

typical sequences, and this is what led to a compression rate arbitrarily close to $h(p)$ for sufficiently small ε and large n.

We then considered quantum data sources, specified in terms of some ensemble with corresponding density operator $\rho = \sum_x p_x |\psi_x\rangle\langle\psi_x|$. Our main result here was that in order to compress at rate R, we wanted *typical subspaces* $\mathcal{H}_n \subseteq (\mathbb{C}^2)^{\otimes n}$, with projectors P_n, such that

(a) $\operatorname{tr}[\rho^{\otimes n} P_n] \to 1$,
(b) $\frac{1}{n}\log(\dim \mathcal{H}_n) \leq R$ for large enough n.

The first condition can be interpreted as requiring that typical states emitted by the source have high overlap with the subspace P_n, and the second condition states that the compression protocol will use no more than nR qubits to compress n samples of the source.

10.1 Construction of typical subspaces

How should we go about constructing such typical subspaces? A natural approach is to take the spectrum decomposition of ρ,

$$\rho = p\,|\phi_0\rangle\langle\phi_0| + (1-p)\,|\phi_1\rangle\langle\phi_1|,$$

and define

$$\mathcal{H}_n := \operatorname{span}\{|\phi_{b_1}\rangle \otimes \dots \otimes |\phi_{b_n}\rangle : \boldsymbol{b} \in \{0,1\}^n \text{ a typical sequence}\}$$

where we include only basis vectors corresponding to typical bitstrings for a *classical* data source with probabilities $\{p, 1-p\}$.

This is a natural definition, since the vectors $|\phi_{b_1}\rangle \otimes \ldots \otimes |\phi_{b_n}\rangle$ is the eigenbasis of $\rho^{\otimes n}$, which makes it easy to evaluate the trace $\mathrm{tr}[\rho^{\otimes n} P_n]$:

$$\mathrm{tr}[\rho^{\otimes n} P_n] = \sum_{\boldsymbol{b} \text{ typical}} \langle \phi_{b_1} \otimes \ldots \otimes \phi_{b_n} | \rho^{\otimes n} | \phi_{b_1} \otimes \ldots \otimes \phi_{b_n} \rangle = \sum_{\boldsymbol{b} \text{ typical}} p^{\#0\text{'s}} (1-p)^{\#1\text{'s}}$$
$$= \mathbf{Pr}(\boldsymbol{b} \text{ is typical}) \to 1.$$

In the third step, we recognized the probability of the classical data source emitting a typical sequence, which goes to one according Eq. (10.1)!

We still need to bound the dimension of these subspaces. But clearly $\dim(\mathcal{H}_n)$ is just the number of typical sequences, so it follows from Eq. (10.2) that

$$\frac{1}{n} \log(\dim \mathcal{H}_n) \leq R := \frac{\log(n+1)}{n} + h(p) + \varepsilon'.$$

As discussed below Eq. (9.4), the first term goes to zero for large n and we can make the third term arbitrarily small by choosing ε small enough. Thus we have construct typical subspaces that allow us to compress a quantum data source at a rate R arbitrarily close to $h(p)$. In Exercise 10.1 you will show that this is the optimal rate.

To summarize: Quantum data compression is possible at an asymptotic qubit rate arbitrarily close to the *von Neumann entropy*

$$S(\rho) := h(p)$$

which is simply the Shannon entropy of the eigenvalues of the density operator. We can also write

$$S(\rho) = -\mathrm{tr}[\rho \log \rho]$$

using the matrix logarithm. The rate $S(\rho)$ is also optimal. This important result is due to Schumacher (as well as the result in the next section). As mentione last time, the quantum data compression protocol that we described last lecture works for all quantum sources described by the density operator ρ.

Again, we may ask about the symmetries of the quantum data compression problem. Instead of relabeling zeros and ones, we could perform an arbitrary unitary transformation U on the states emitted by the source. Such a transformation is reversible and hence should not impact the compression rate. Indeed, $S(\rho) = S(U\rho U^\dagger)$, since the von Neumann entropy only depends on the eigenvalues of the density operator. But, again, our compression protocol breaks these symmetries because the subspaces $\mathcal{H}_n$ refer explicitly to the eigenbasis of ρ. This means that we cannot we apply a protocol constructed for a source described by ρ to a source described by $U\rho U^\dagger$ and expect that it works with high fidelity. We had a similar issue in the classical case and found an easy fix. In the quantum case, it is less obvious what to do.

Next week, we will undertake a more careful study of the symmetries of $(\mathbb{C}^2)^{\otimes n}$ and of $\rho^{\otimes n}$ and overcome this challenge. This will not only allow us to construct a universal compression protocol, but also solve other problems of interest. Specifically, it will allows us to estimate the estimate the eigenvalues of an unknown density operator, the corresponding von Neumann entropy, and, finally, the entire density operator.

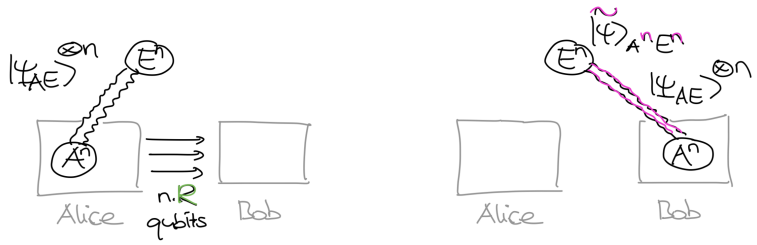


Figure 10.1: Alice wants to send half of her entangled states $|\Psi_{AE}\rangle^{\otimes n}$ over to Bob at qubit rate R.

10.2 Compression and entanglement

At a high level, compression is about minimizing communication. There are other situations in which we would like to minimize communication, such as in the following task: Suppose we start out with a large number of copies of a bipartite pure state $|\Psi_{AE}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_E$. Alice would like to transfer her A-systems (which we assume are qubits) over to Bob by sending a minimal number of qubits. Importantly, they would like to preserve all correlations with the E-systems, but neither Alice nor Bob have access to the E-systems, but they belong to another party (or the "environment") that we will call Eve. See Fig. 10.1 for an illustration.

We will call this task *quantum state transfer* (sadly, this term is usually used with a different connotation). It is often referred to as *Schumacher compression*. Thus, if $|\widetilde{\psi}\rangle_{A^nE^n}$ is the state after compression and decompression, we would like that

$$|\widetilde{\psi}\rangle_{A^nE^n} \approx |\Psi_{AE}\rangle^{\otimes n}$$

(say, on average).

Since our goal is to preserve the correlations, we might intuitively expect that the more entangled the states $|\Psi_{AE}\rangle$ are, the more communication will be required. Indeed, suppose that $|\Psi_{AE}\rangle = |\Psi\rangle_A \otimes |\Psi\rangle_E$ is a product state. In this case, Alice needs not send over *any* quantum information at all, since Bob can simply prepare the pure state $|\Psi\rangle$ on his side. However, if $|\Psi_{AE}\rangle$ is entangled then it is not hard to see that communication will be required. (Any state that Bob prepares on his end alone will necessarily be in a tensor product with Eve's state.)

Interestingly, quantum state transfer can be implemented by a protocol that is very similar to our quantum data compression protocol. The key idea is to use typical subspaces for the reduced density operator

$$\rho_A = \mathrm{tr}_E[|\Psi_{AE}\rangle \langle\Psi_{AE}|]$$

and we describe the protocol next:

Protocol for quantum state transfer: Let $\mathcal{H}_{A,n} \subseteq (\mathbb{C}^2)^{\otimes n}$ be typical subspace, with projectors $P_{A,n}$.

- Alice performs the projective measurement $\{P_{A,n}, I_{A^n} - P_{A,n}\}$. If the outcome is the latter, she signals failure.
- Otherwise, the post-measurement state is

$$|\widetilde{\psi}_{A^nE^n}\rangle = \frac{(P_{A,n} \otimes I_{E^n})\,|\Psi_{AE}\rangle^{\otimes n}}{\|(P_{A,n} \otimes I_{E^n})\,|\Psi_{AE}\rangle^{\otimes n}\|} \in \mathcal{H}_{A,n} \otimes \mathcal{H}_E^{\otimes n}.$$

- Alice sends over her subsystem $\mathcal{H}_{A,n}$ using approximately $nS(\rho_A)$ qubits (see Theorem 9.5 for).

It is straightforward to analyze this protocol. Using Born's rule, the probability of passing the first step of the protocol only depends on the reduced density operator and is given by

$$\mathbf{Pr}(\text{success}) = \langle \Psi_{AE}^{\otimes n} | P_{A,n} \otimes I_{E^n} | \Psi_{AE}^{\otimes n} \rangle = \text{tr}[\rho_A^{\otimes n} P_{A,n}] \to 1, \tag{10.3}$$

since the $P_{A,n}$ are projectors onto typical subspaces for $\rho_A^{\otimes n}$. And assuming we did not fail in the first step, the overlap between the post-measurement state and the target state is given by

$$|\langle \Psi_{AE}^{\otimes n} | \widetilde{\psi}_{A^n E^n} \rangle|^2 = |\langle \Psi_{AE}|^{\otimes n} \frac{(P_{A,n} \otimes I_{E^n}) |\Psi_{AE}\rangle^{\otimes n}}{\|(P_{A,n} \otimes I_{E^n}) |\Psi_{AE}\rangle^{\otimes n}\|}|^2 = \frac{|\langle \Psi_{AE}^{\otimes n} | P_{A,n} \otimes I_{E^n} | \Psi_{AE}^{\otimes n} \rangle|^2}{\|(P_{A,n} \otimes I_{E^n}) |\Psi_{AE}^{\otimes n}\rangle\|^2}$$
$$= \langle \Psi_{AE}^{\otimes n} | P_{A,n} \otimes I_{E^n} | \Psi_{AE}\rangle^{\otimes n} = \text{tr}[\rho_A^{\otimes n} P_{A,n}] \to 1$$

where the last step is the same calculation as in Eq. (10.3)!

To summarize: Alice can transfer her system to Bob at an asymptotic qubit rate that can be arbirarily close to $S(\rho_A)$. This quantity is often called the *entanglement entropy* of the pure state $|\Psi_{AE}\rangle$, denoted

$$S_E(\Psi) := S(\rho_A) = S(\rho_E).$$

Here we used that $S(\rho_A) = S(\rho_E)$ as a consequence of the Schmidt decomposition (see Eq. (7.5)).

Remark 10.1. The notation here is very unfortunate – the E in S_E is short for "entanglement" and not for Eve's system. E.g., for a state $|\Phi_{AB}\rangle$ we would write $S_E(\Phi) = S(\rho_A) = S(\rho_B)$.

Example 10.2. If $|\Psi_{AE}\rangle = |0\rangle_A \otimes |0\rangle_E$ then $S_E(\Psi) = 0$ – as it should be, given our discussion above. If $|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}} (|00\rangle_{AB} + |11\rangle_{AB})$ is the ebit state, however, then $S_E = 1$, which means that Alice has to send qubits at a trivial rate of 1 qubit/qubit – in agreement with our intuition that the ebit is a maximally entangled state.

We thus obtain a second operational interpretation of the von Neumann entropy: It not only characterizes the optimal quantum compression rate for a quantum data source, but it also characterizes the minimal rate of qubits that we need to send when transferring part of a bipartite pure state.

The state transfer problem is a special case of the more general (and more difficult) problem of *quantum state merging*, where the receiver already possesses part of the state. We might have a peek at this in the last week of class.

Remark 10.3. It is possible to show that any protocol for the state transfer task can be used to compress arbitrary quantum sources described by the density operator ρ_A.

10.3 Entanglement transformations

At the end of this lecture, we briefly talked some more about entanglement more generally. For pure states, $|\Psi_{AB}\rangle \neq |\psi_A\rangle \otimes |\psi_B\rangle$ means that the state is entangled. But how can be compare and quantify different states in their entanglement? One approach is to assign to each state some arbitrary numbers that we believe reflect aspects of their entanglement properties – e.g., the entanglement entropy S_E from above, the largest eigenvalue of the reduced density matrix from Exercise 8.1, or simply the collection of all eigenvalues of ρ_A or ρ_B (sometimes called the

entanglemen spectrum). Yet, this approach might perhaps seem somewhat *ad hoc* and so is (a priori) not completely satisfactory.

A more operational approach would be to compare two states $|\Phi_{AB}\rangle$ and $|\Psi_{AB}\rangle$ by studying whether one can be transformed into the other: What family of operations should we consider in such a transformation? Since our goal is compare entanglement, we should only allow for operations that cannot create entanglement from unentangled states. We already briefly mentioned such a class of operations in Theorem 8.4. It is LOCC, short for *Local Operations and Classical Communication*. Here, we imagine that Alice and Bob each have their separate laboratory and we allow the following operations:

- Local operations, i.e., arbitrary quantum operations that can be done on Alice' and Bob's subsystems. We allow any combination of unitaries, adding auxiliary systems, performing partial traces, and measurements.
- Classical communication, i.e., Alice and Bob are allowed to exchange measurement outcomes. Thus, Bob's local operations can depend on Alice's previous measurement outcomes, and vice versa.

Thus we are interested in whether

$$|\Psi_{AB}\rangle \overset{LOCC}{\longrightarrow} |\Phi_{AB}\rangle\,.$$

If yes, then we could say that $|\Psi_{AB}\rangle$ is at least as entangled as $|\Phi_{AB}\rangle$ – indeed, the former is as useful as the latter for any nonlocal quantum information processing task, since we can always convert first $|\Psi_{AB}\rangle$ into $|\Phi_{AB}\rangle$ when required.

Remark 10.4. Note that the setup here is very different from quantum data compression – there, we wanted to minimize the amount of quantum communication sent. Here, we do not allow *any* quantum communication, and classical communication comes for free.

The *exact* interconversion problem for pure states was solved by Nielsen. However, there are many parameters – namely all the eigenvalues of ρ_A and of ρ_B matter. It turns out that the asymptotic theory simplifies tremendously, and we will very briefly discuss the main results.

The key idea is to reduce the problem to studying the conversion between a given state $|\Psi_{AB}\rangle$ and a single resource state (a "universal currency" of entanglement of sorts). This resource state is the *ebit* state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$!

Thus we are interested in the following two problems: First, given n copies of a state $|\Psi_{AB}\rangle$, convert them by LOCC into as many ebits as possible:

$$|\Psi_{AB}\rangle^{\otimes n} \overset{LOCC}{\longrightarrow} \approx |\Phi^+\rangle^{\otimes Rn}$$

Just as in the case of data compression, we are interested in the maximal rate R that can be achieved with error going to zero for $n \to \infty$. This is called the *distillable entanglement* $E_D(\Psi)$ of the state $|\Psi_{AB}\rangle$.

Second, given as few ebits as possible, convert them by LOCC into n copies of $|\Psi_{AB}\rangle$:

$$|\Phi^+\rangle^{\otimes Rn} \overset{LOCC}{\longrightarrow} \approx |\Psi_{AB}\rangle^{\otimes n}$$

Here we are interested are interested in the minimal rate R that can be achieved with error going to zero for $n \to \infty$. This is called the *entanglement cost* $E_C(\Psi)$ of the state $|\Psi_{AB}\rangle$.

It is intuitively plausible that $E_C(\Psi) \geq E_D(\Psi)$, i.e., that we cannot "create entanglement out of nothing". The main result of the theory is the following: The entanglement cost and the distillable entanglement are equal, and given by the entanglement entropy discussed above!

$$E_C(\Psi) = E_D(\Psi) = S_E(\Psi)$$

Remark 10.5. You might wonder how the above story generalizes to mixed states ρ_{AB}. It turns out that in this case the entanglement theory is much more complicated. We already saw hints of this in Section 8.2 where we mentioned that even deciding whether a given state ρ_{AB} is separable or entangled is in general an NP-hard problem. In addition, while the same definitions can be made as above, there are many new phenomena. For example, in general we have that $E_C(\rho) > E_D(\rho)$, meaning that the conversion via ebits is in general asymptotically irreversible! In fact, there are entangled mixed states states such that $E_C(\rho) > 0$ while $E_D(\rho) = 0$. We call them *bound entangled states* – these are states that are entangled but no ebits can be distilled from them!

Exercises

10.1 **Quantum data compression:** In this problem you will show that there cannot exist typical subspaces with rates smaller than the von Neumann entropy. Thus, let ρ be a density operator on $\mathbb{C}^2$ and $\mathcal{H}_n \subseteq (\mathbb{C}^2)^{\otimes n}$ an arbitrary sequence of subspaces, with corresponding projectors P_n, such that $\dim(\mathcal{H}_n) \leq 2^{nR}$ for all n. Show that either $R \geq S(\rho)$ or $\mathrm{tr}[\rho^{\otimes n} P_n] \to 0$.

Chapter 11

Representation theory of U(2) and SU(2)

MW: Move this chapter to Chapter 12? MW: Let's give a proof that any irrep is of the indicated type. MW: Let's also discuss the action of U(2), not just of SU(2).

Last week, we learned the basic concepts of group representation theory (Chapter 5) and we proved that the symmetric subspaces are irreducible representations of SU(2) (Chapter 6). Today, we will discuss how the symmetric subspaces fit in the representation theory of SU(2) more generally, and we will discuss how to decompose an arbitrary representation of SU(2) into irreducibles.

11.1 Representation theory of SU(2)

We start by introducing some notation. For reasons that will become clear soon, it will be convenient to use k instead of n. So we will write $\mathrm{Sym}^k(\mathbb{C}^2)$ for the symmetric subspace of the k-th tensor power. Let us also denote by $T_U^{(k)}$ the restriction of $T_U = U^{\otimes k}$ to the symmetric subspace. That is, $T_U^{(k)}$ is given by the same formula $U^{\otimes k}$, but we only plug in vectors in the symmetric subspace and remember that the result will automatically by in the symmetric subspace. For $k = 0$, we define $\mathrm{Sym}^0(\mathbb{C}^2) = \mathbb{C}$ as the trivial representation, with $T_U^{(0)} = I$. Thus, the Hilbert space $\mathrm{Sym}^k(\mathbb{C}^2)$ together with the operators $\{T_U^{(k)}\}_{U\in\mathrm{SU}(2)}$ defines a representation of SU(2), and it is this representation that we proved to be irreducible in Chapter 6.

A basic question in the representation theory of any group is to ask about the possible irreducible representations, up to equivalence. For the group SU(2), one can show that *every* irreducible representation is equivalent to a symmetric subspace (we will not prove this fact). **MW: It would be nice if we did!** That is, if $\mathcal{H}$ is an arbitrary irreducible representation of SU(2), with corresponding operators $\{R_U\}$, then there exists $k \geq 0$ and a unitary intertwiner $J\colon \mathcal{H} \to \mathrm{Sym}^k(\mathbb{C}^2)$ such that

$$JR_UJ^\dagger = T_U^{(k)} \qquad \forall U \in \mathrm{SU}(2).$$

We will abbreviate this situation by the notation $\mathcal{H} \cong \mathrm{Sym}^k(\mathbb{C}^2)$ and $R_U \cong T_U^{(k)}$ introduced last lecture. Moreover, the symmetric subspaces are inequivalent for $k \neq l$, i.e., $\mathrm{Sym}^k(\mathbb{C}^2) \ncong \mathrm{Sym}^l(\mathbb{C}^2)$. This follows directly from the fact that $\dim \mathrm{Sym}^k(\mathbb{C}^2) = k+1$, so there cannot be a unitary map between different symmetric subspaces.

To summarize, any irreducible representation $\mathcal{H}$ of SU(2) is equivalent to exactly one of the symmetric subspaces $\mathrm{Sym}^k(\mathbb{C}^2)$, up to equivalence, and can therefore by labeled by an integer k. We can determine k directly from the dimension formula as $k = \dim H - 1$. You may know from your quantum mechanics class that the irreducible representations can also be labeled by their spin j, which is a *half-integer*. As you might expect, the connection is precisely that $j = k/2$.

Let us discuss some examples. A good source of SU(2)-representations are the various tensor powers of $\mathbb{C}^2$, i.e., $(\mathbb{C}^2)^{\otimes n}$, so this is what we shall consider. For $n = 0$, we have the trivial representation

$$(\mathbb{C}^2)^{\otimes 0} = \mathrm{Sym}^0(\mathbb{C}^2),$$

and for $n = 1$, we have

$$(\mathbb{C}^2)^{\otimes 1} = \mathrm{Sym}^1(\mathbb{C}^2) = \mathbb{C}^2$$

so this is again irreducible (and not very interesting). The first interesting examples is $n = 2$, since here we know that $(\mathbb{C}^2)^{\otimes 2}$ is *not* irreducible. In fact:

$$(\mathbb{C}^2)^{\otimes 2} = \mathbb{C} \otimes \mathbb{C} = \mathrm{Sym}^2(\mathbb{C}^2) \overset{\perp}{\oplus} \mathbb{C}\,|\Psi^-\rangle\,,$$

where $|\Psi^-\rangle = \sqrt{\frac{1}{2}}\,(|10\rangle - |01\rangle)$ is the singlet state. Both summands are the irreducible – the former because it is a symmetric subspace, and the latter since it is a one-dimensional invariant subspace. Which symmetric subspace is the latter isomorphic to? Clearly, this must be the one-dimensional $\mathrm{Sym}^0(\mathbb{C}^2)$. To see this more concretely, recall that in Exercise 3.5 you showed that

$$(U \otimes U)\,|\Psi^-\rangle = \det(U)\,|\Psi^-\rangle$$

for all unitaries U. If $U \in \mathrm{SU}(2)$ then $\det(U) = 1$, so $|\Psi^-\rangle$ spans indeed a trivial representation. We can summarize this as follows: As representations of SU(2),

$$\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathrm{Sym}^2(\mathbb{C}^2) \oplus \mathrm{Sym}^0(\mathbb{C}^2). \tag{11.1}$$

Is there a systematic way of decomposing higher tensor powers $(\mathbb{C}^2)^{\otimes n}$ for $n > 2$? We will discuss this next.

11.2 Decomposing representations of SU(2)

In fact, let us consider a more general question: Suppose we are given an arbitrary SU(2)-representation $\mathcal{H}$, with operators $\{R_U\}_{U \in \mathrm{SU}(2)}$. We know that we can always decompose a representation into irreducibles, so that

$$\mathcal{H} \cong \mathrm{Sym}^{k_1}(\mathbb{C}^2) \oplus \mathrm{Sym}^{k_2}(\mathbb{C}^2) \oplus \ldots \oplus \mathrm{Sym}^{k_m}(\mathbb{C}^2),$$

but how can we determine the numbers $k_1, \ldots, k_m$ that appear? In other words, how can we figure out how many times a certain irreducible representation $\mathrm{Sym}^k(\mathbb{C}^2)$ appears in $\mathcal{H}$? We can solve this by a similar procedure as we used last time in class. Start by defining the operator

$$r_Z := -i\partial_{s=0}\,[R_{e^{\imath sZ}}]\,. \tag{11.2}$$

Note that $e^{\imath sZ} = \left(\begin{smallmatrix} e^{\imath s} & 0 \\ 0 & e^{-\imath s} \end{smallmatrix}\right) \in \mathrm{SU}(2)$, so this definition makes sense assuming R_U is differentiable as a function of U. In general, the operator r_Z will always be Hermitian. (As mentioned in the previous lecture, this definition can be understood more conceptually in terms of the action of the Lie algebra of SU(2).)

For example, if $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ with $R_U = U^{\otimes n}$, then $r_Z = \widetilde{Z} = Z \otimes I \otimes \ldots \otimes I + \cdots + I \otimes \ldots \otimes I \otimes Z$ in the notation of yesterday's lecture, which was one of the ingredients for proving that the symmetric subspaces are irreducible. In particular, we proved that the operator $\widetilde{Z}$ preserves the

symmetric subspace. Let us denote its restriction by $t_Z^{(k)}$. Yesterday, we proved that each of the basis vectors $|\omega_{m,k-m}\rangle$ for $m = 0, \ldots, k$ are eigenvectors of $t_Z^{(k)}$, with associated eigenvalue $2m - k$. Thus, the operator $t_Z^{(k)}$ has eigenvalues $\{-k, -k+2, \ldots, k-2, k\}$, each with multiplicity one.

Now assume that $\mathcal{H}$ is irreducible and equivalent to some $\mathrm{Sym}^k(\mathbb{C}^2)$ by a unitary intertwiner $J \colon \mathcal{H} \to \mathrm{Sym}^k(\mathbb{C}^2)$. Then,

$$Jr_Z J^\dagger = -i\partial_{s=0} \left[JR_{e^{\imath s Z}} J^\dagger \right] = -i\partial_{s=0} \left[T^{(k)}_{e^{\imath s Z}} \right] = t_Z^{(k)},$$

and so we see that r_Z has likewise eigenvalues $\{-k, -k+2, \ldots, k-2, k\}$, each with multiplicity one.

How about the general case, where

$$\mathcal{H} \cong \mathrm{Sym}^{k_1}(\mathbb{C}^2) \oplus \mathrm{Sym}^{k_2}(\mathbb{C}^2) \oplus \ldots \oplus \mathrm{Sym}^{k_m}(\mathbb{C}^2)$$

? Here we have a unitary intertwiner J such that

$$JR_U J^\dagger = \begin{pmatrix} T_U^{(k_1)} & & & \\ & T_U^{(k_2)} & & \\ & & \ddots & \\ & & & T_U^{(k_m)} \end{pmatrix}$$

and hence

$$Jr_Z J^\dagger = \begin{pmatrix} t_Z^{(k_1)} & & & \\ & t_Z^{(k_2)} & & \\ & & \ddots & \\ & & & t_Z^{(k_m)} \end{pmatrix}$$

for the same reason as above. It follows that the eigenvalue spectrum of r_Z is given by the multiset

$$\{-k_1, -k_1+2, \ldots, k_1-2, k_1\} \sqcup \{-k_2, -k_2+2, \ldots, k_2-2, k_2\} \sqcup \cdots \sqcup \{-k_m, -k_m+2, \ldots, k_m-2, k_m\}.$$

It is not hard to see that one can inductively reverse-engineer the numbers $k_1, k_2, \ldots, k_m$ from this multiset: Start by taking the largest number; it must be one of the k_i's. Remove the corresponding $\{-k_i, -k_i+2, \ldots, k_i-2, k_i\}$ from the set, and repeat the procedure. Let us discuss some examples.

First, we can use this to reprove the decomposition in Eq. (11.1). Here, $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$ and $r_Z = \widetilde{Z} = Z \otimes I + I \otimes Z$ as explained above. Thus, r_Z is diagonal in the computational basis and the eigenvalues of r_Z are

$$\{2, 0, 0, -2\} = \{2, 0, -2\} \sqcup \{0\}.$$

This decomposition makes it clear that

$$(\mathbb{C}^2)^{\otimes 2} = \mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathrm{Sym}^2(\mathbb{C}^2) \oplus \mathrm{Sym}^0(\mathbb{C}^2), \tag{11.3}$$

which confirms our previous decomposition.

Next, let us consider $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$, where $r_Z = \widetilde{Z} = Z \otimes I \otimes I + I \otimes Z \otimes I + I \otimes I \otimes Z$. Here the eigenvalues are

$$\{3, 1, 1, 1, -1, -1, -1, -3\} = \{3, 1, -1, -3\} \sqcup \{1, -1\} \sqcup \{1, -1\},$$

which implies that

$$(\mathbb{C}^2)^{\otimes 3} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathrm{Sym}^3(\mathbb{C}^2) \oplus \mathrm{Sym}^1(\mathbb{C}^2) \oplus \mathrm{Sym}^1(\mathbb{C}^2). \tag{11.4}$$

At least in principle it is now clear how to proceed for arbitrary tensor powers $(\mathbb{C}^2)^{\otimes n}$. However, the counting gets more involved the larger n, so it is desirable to figure out an *inductive* way of computing this decomposition. The basic problem that we have to solve is the following. Suppose that we have an irreducible representation $\mathrm{Sym}^k(\mathbb{C}^2)$ and we tensor it with an additional qubit $\mathbb{C}^2$, i.e., we consider the representation

$$\mathcal{H} = \mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^2, \qquad R_U = T_U^{(k)} \otimes U.$$

How does it decompose into irreducibles? The answer is the following:

$$\mathcal{H} = \mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^2 \cong \begin{cases} \mathrm{Sym}^{k+1}(\mathbb{C}^2) \oplus \mathrm{Sym}^{k-1}(\mathbb{C}^2) & \text{if } k > 0 \\ \mathbb{C}^2 & \text{if } k = 0. \end{cases} \tag{11.5}$$

To confirm this formula, note that $r_Z = t_Z^{(k)} \otimes I + I \otimes Z$, so that the eigenvalues are

$$\{-k \pm 1, -k + 2 \pm 1, \ldots, k - 2 \pm 1, k \pm 1\} = \{-(k+1), -(k-1), \ldots, k-1, k+1\} \sqcup \{-(k-1), \ldots, k-1\};$$

the second set is empty if $k = 0$. See Fig. 15.3 for an illustration.

Equation (11.5) is as special case of the so-called *Clebsch-Gordan rule* that you might know from a quantum mechanics class. It tells you more generally how to decompose $\mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathrm{Sym}^l(\mathbb{C}^2)$. We will not need the general result but it can be proved just like above.

Let's quickly check that Eq. (11.5) reproduces the same results that we derived above. We start by

$$(\mathbb{C}^2)^{\otimes 2} = \mathrm{Sym}^1(\mathbb{C}^2) \otimes \mathbb{C}^2 \cong \mathrm{Sym}^2(\mathbb{C}^2) \oplus \mathrm{Sym}^0(\mathbb{C}^2).$$

The last step is using the Clebsch-Gordan rule and the result is in agreement with Eqs. (11.1) and (11.3). Next, we decompose the third tensor power by tensoring with an additional qubit:

$$\begin{aligned} (\mathbb{C}^2)^{\otimes 3} &= (\mathbb{C}^2)^{\otimes 2} \otimes \mathbb{C}^2 \cong \big(\mathrm{Sym}^2(\mathbb{C}^2) \oplus \mathrm{Sym}^0(\mathbb{C}^2)\big) \otimes \mathbb{C}^2 \\ &\cong \big(\mathrm{Sym}^2(\mathbb{C}^2) \otimes \mathbb{C}^2\big) \oplus \big(\mathrm{Sym}^0(\mathbb{C}^2) \otimes \mathbb{C}^2\big) \\ &\cong \big(\mathrm{Sym}^3(\mathbb{C}^2) \oplus \mathrm{Sym}^1(\mathbb{C}^2)\big) \oplus \big(\mathrm{Sym}^1(\mathbb{C}^2)\big) \\ &= \mathrm{Sym}^3(\mathbb{C}^2) \oplus \mathrm{Sym}^1(\mathbb{C}^2) \oplus \mathrm{Sym}^1(\mathbb{C}^2), \end{aligned}$$

which confirms Eq. (11.4). Here we first used the two-qubit result, then the distributivity law, and finally the Clebsch-Gordan rule. Similarly,

$$\begin{aligned} (\mathbb{C}^2)^{\otimes 4} &\cong \big(\mathrm{Sym}^3(\mathbb{C}^2) \oplus \mathrm{Sym}^1(\mathbb{C}^2) \oplus \mathrm{Sym}^1(\mathbb{C}^2)\big) \otimes \mathbb{C}^2 \\ &\cong \mathrm{Sym}^4(\mathbb{C}^2) \oplus \mathrm{Sym}^2(\mathbb{C}^2) \oplus \mathrm{Sym}^2(\mathbb{C}^2) \oplus \mathrm{Sym}^2(\mathbb{C}^2) \oplus \mathrm{Sym}^0(\mathbb{C}^2) \oplus \mathrm{Sym}^0(\mathbb{C}^2). \end{aligned}$$

It is now clear how to decompose $(\mathbb{C}^2)^{\otimes n}$ for arbitrary n in an inductive fashion. We will use this to great effect in two weeks in Chapters 12 and 13. There, we will also learn how to extend our considerations from $\mathrm{SU}(2)$ to $U(2)$.

Chapter 12

Spectrum estimation, i.i.d. quantum information

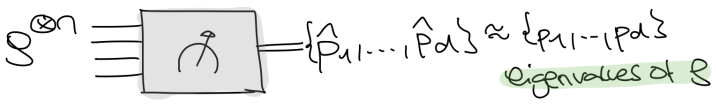
Today, we will start developing some new machinery for working with i.i.d. copies of a quantum state, i.e.,

$$\rho^{\otimes n} \text{ on } (\mathbb{C}^d)^{\otimes n}$$

where ρ is an arbitrary density operator.

12.1 Spectrum estimation

Our motivation throughout today's lecture will be the following 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 outcomes $\hat{p}_1 \geq \cdots \geq \hat{p}_d$ that are a good estimate for the true eigenvalues, as illustrated below:



This task is known as the *spectrum estimation* problem and it was first solved by Keyl and Werner [KW01] (cf. [CM06]). 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. As a direct corollary, we will be able to estimate the von Neumann entropy $S(\rho)$ of an unknown quantum source (since this is a function of the eigenvalues only). We will spend the rest of today's lecture 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 Chapter 13, we will use them to construct *universal* typical subspaces, which work for any density operator ρ with given spectrum. This will allow us to derive universal protocols for quantum data compression and quantum state transfer – the two problems discussed last week in Chapters 9 and 10. In Chapter 14, we will also see how one can estimate an arbitrary unknown quantum state ρ from $\rho^{\otimes n}$, thereby solving a task that is 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 Theorem 5.2, 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 $T_U = U^{\otimes n}$.

Now, 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 can verify this explicitly on a product basis:

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

Remark 12.1 (Warning)**.** Only when $\rho = |\psi\rangle\langle\psi|$ is a pure state is $\rho^{\otimes n} = |\psi\rangle^{\otimes n} \langle\psi|^{\otimes n}$ an operator on the symmetric subspace. We explored this at lengths in Chapters 4 to 6, 8 and 11. However, as soon as ρ is a mixed state, this is no longer the case! A simple example is the maximally mixed state $\tau = I/d$. Clearly, $\tau^{\otimes n} = I/d^n$ is supported on all of $(\mathbb{C}^d)^{\otimes n}$.

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

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

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 $\rho^{\otimes n}$, 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.

12.2 Warmup: The swap test

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

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

Both the symmetric and the antisymmetric subspace are irreducible representations. (for the symmetric subspace, we discussed this in Chapter 6; 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 operator corresponding to the latter is known as the *swap operator*

$$F = R_{1\leftrightarrow 2} = \sum_{a,b} |a, b\rangle\langle b, a| .$$

which you will recognize from Exercise 8.4. It commutes both with the action of $U(d)$ (since we know that $[U^{\otimes n}, R_\pi] = 0$ for all U and π) as well as with the action of S_2 (any operator commutes with itself and with the identity matrix). Since the projector onto the symmetric subspace can be written as $\Pi_2 = \frac{1}{2}(I + F)$, it follows that the projective measurement

$$\{P_1 := \Pi_2, \quad P_0 := I - \Pi_2\}$$

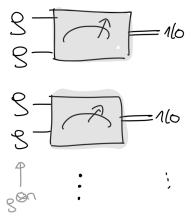


Figure 12.1: By measuring $\{P_1, P_0\}$ on $N = n/2$ independent copies of $\rho^{\otimes 2}$, we can estimate the purity of the quantum state via Eq. (12.1).

likewise commutes with the actions of $U(d)$ and S_2 – so we have identified a projective measurement with the desired symmetries!

Note that $F = P_1 - P_0$ is just the spectral decomposition of the swap operator. Using Schur's lemma as in Exercise 7.4, you can verify that there is no more fine-grained measurement with these symmetries.

Is the measurement $\{P_1, P_0\}$ at all informative? To see this, we calculate the probability of the "1" outcome:

$$\mathbf{Pr}_{\rho^{\otimes 2}}(\text{outcome } 1) = \operatorname{tr}\left[\rho^{\otimes 2}\Pi_2\right] = \operatorname{tr}\left[\rho^{\otimes 2}\frac{1}{2}\left(I + F\right)\right] = \frac{1}{2}\left(1 + \operatorname{tr}\left[\rho^{\otimes 2}F\right]\right) = \frac{1}{2}\left(1 + \operatorname{tr}\rho^2\right),$$

where we used the "swap trick" $\operatorname{tr}[F(\sigma \otimes \gamma)] = \operatorname{tr}[\sigma\gamma]$ from Exercise 8.4 in the last step. 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 (Exercise 7.2).

The important point, however, is that since ρ has eigenvalues $p_1, \ldots, p_d$ then $\operatorname{tr}\rho^2 = \sum_{i=1}^d p_i^2$, so

$$\mathbf{Pr}_{\rho^{\otimes 2}}(\text{outcome } 1) = \frac{1}{2}\left(1 + \sum_{i=1}^d p_i^2\right)$$

and we conclude that this simple measurement already allows us to learn something nontrivial about the eigenvalues of ρ. It is also known as the *swap test*. Note that for qubits ($d = 2$) the swap test provides a *complete* solution (since $p_1 + p_2 = 1$ we can determine p_1 and $p_2 = 1 - p_1$ from $\operatorname{tr}\rho^2 = p_1^2 + p_2^2$)!

Just to be perfectly clear about the interpretation of this result: When performing the projective measurement $\{P_1, P_0\}$, the measurement outcome is either 1 or 0. Only when repeated N times on independent copies of $\rho^{\otimes 2}$ will we find that

$$\frac{\#\{\text{outcome}{=}1\}}{N} \approx \mathbf{Pr}_{\rho^{\otimes 2}}(\text{outcome } 1) = \frac{1}{2}\left(1 + \sum_{i=1}^d p_i^2\right) \tag{12.1}$$

up to error $O(1/\sqrt{N})$. Thus we only obtain a good estimate when we apply the swap test to a number N of pairs $\rho^{\otimes 2}$, i.e., when given $\rho^{\otimes n}$ for large $n = 2N$ (Fig. 12.1).

While the swap test is perfectly fine for the purposes of estimating the purity, it is somewhat unsatisfactory in two regards: (i) it only works for $d > 2$ and (ii) measuring on "blocks of $\rho^{\otimes 2}$" breaks the permutation symmetry of the problem.

In the following, we will discuss a different solution which fully exploits the symmetries of the problem and generalizes readily to any d. Along the way we will discover some important tools that will have further application in the remainder of this course. For simplicity, we restrict to the case of qubits ($d = 2$) since we studied the representation theory of SU(2) before in Chapter 11.

12.3 Decomposing the n-qubit Hilbert space

We start by decomposing the Hilbert space of n qubits into irreducible representations of SU(2). From Section 11.2 we know that

$$\left(\mathbb{C}^2\right)^{\otimes n} \cong \mathrm{Sym}^{k_1}(\mathbb{C}^2) \oplus \mathrm{Sym}^{k_2}(\mathbb{C}^2) \oplus \ldots \oplus \mathrm{Sym}^{k_m}(\mathbb{C}^2)$$

for certain integers $k_1, \ldots, k_m \geq 0$ that we still need to determine (one of them should be $k_i = n$, corresponding to the symmetric subspace $\mathrm{Sym}^n(\mathbb{C}^2) \subseteq (\mathbb{C}^2)^{\otimes n}$). It is convenient to repackage this in the following way:

$$\left(\mathbb{C}^2\right)^{\otimes n} \cong \bigoplus_k \left(\underbrace{\mathrm{Sym}^k(\mathbb{C}^2) \oplus \ldots \oplus \mathrm{Sym}^k(\mathbb{C}^2)}_{m(n,k) \text{ times}} \right) \cong \bigoplus_k \mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^{m(n,k)}$$

In the first step, we reordered the symmetric subspaces according to their type (k), and in the second step we used that, for any representation $\mathcal{H}$, $\mathcal{H} \otimes \mathbb{C}^m \cong \mathcal{H} \oplus \ldots \oplus \mathcal{H}$ (m copies). Just to be sure that you remember: The above notation means that there exist unitary intertwiners that map the representation operators as follows:

$$U^{\otimes n} \cong \bigoplus_k \left(T_U^{(k)} \oplus \ldots \oplus T_U^{(k)} \right) \cong \bigoplus_k T_U^{(k)} \otimes \mathbb{C}^{m(n,k)} = \left[\begin{array}{c|c|c} T_U^{(0)} \otimes I_{\mathbb{C}^{m(n,0)}} & & \\ \hline & T_U^{(1)} \otimes I_{\mathbb{C}^{m(n,1)}} & \\ \hline & & \ddots \end{array}\right].$$

Importantly, the above considerations only hold for $U \in \mathrm{SU}(2)$. How about a general unitary $U \in U(2)$? In this case, $U/\sqrt{\det U} \in \mathrm{SU}(2)$, so it is easy to deduce the action. We find that

$$\begin{aligned} U^{\otimes n} &= (\det U)^{n/2} \left(\frac{U}{\sqrt{\det U}} \right)^{\otimes n} \\ &\cong (\det U)^{n/2} \bigoplus_k T^{(k)}_{\frac{U}{\sqrt{\det U}}} \otimes I_{m(n,k)} = (\det U)^{n/2} \bigoplus_k (\det U)^{-k/2} T_U^{(k)} \otimes I_{m(n,k)} \\ &= \bigoplus_k \underbrace{(\det U)^{(n-k)/2} T_U^{(k)}}_{=: T_U^{(n,k)}} \otimes I_{m(n,k)}. \end{aligned}$$

Here we used that, since $T_U^{(k)}$ is given by the restriction of $U^{\otimes k}$ to the symmetric subspace, it is homogeneous of degree k in U.

Let us write $V_{n,k} := \mathrm{Sym}^k(\mathbb{C}^2)$ for the symmetric subspace equipped with the operators $\{T_U^{(n,k)}\}$. This defines a representation of $U(2)$ which is irreducible (since it is even irreducible if we restrict to SU(2)). Importantly, $V_{n,k} \ncong V_{n',k}$ if $n \neq n'$ (since in this case operators with nonzero determinant will in general act in a different way).

Example 12.2. For $n = 2$, we have that

$$\left(\mathbb{C}^2\right)^{\otimes 2} = \mathrm{Sym}^2(\mathbb{C}^2) \oplus \mathbb{C}\,|\Psi^-\rangle \cong V_{2,2} \oplus V_{2,0}.$$

Indeed, $V_{2,2} = \mathrm{Sym}^2(\mathbb{C}^2)$ as a $U(2)$-representation, while you showed in Exercise 3.5 that $(U \otimes U)\,|\Psi^-\rangle = \det(U)\,|\Psi^-\rangle$; the latter is just the way that $T_U^{(2,0)}$ acts on $V_{2,0}$.

We thus obtain the following decomposition of the n-qubit Hilbert space as a representations of $U(2)$:

$$\begin{aligned}\left(\mathbb{C}^2\right)^{\otimes n} &\cong \bigoplus_k V_{n,k} \otimes \mathbb{C}^{m(n,k)}, \\ U^{\otimes n} &\cong \bigoplus_k T_U^{(n,k)} \otimes I_{m(n,k)}.\end{aligned} \tag{12.2}$$

Note that both the left-hand and the right-hand side of Eq. (12.2) make syntactical sense for arbitrary operators, not just for unitaries U. In fact, the equality is true for arbitrary operators! We summarize this important fact: For every operator A on $\mathbb{C}^2$,

$$A^{\otimes n} \cong \bigoplus_k T_A^{(n,k)} \otimes I_{m(n,k)}, \tag{12.3}$$

where

$$T_A^{(n,k)} := (\det A)^{(n-k)/2}\, T_A^{(k)}. \tag{12.4}$$

We will briefly sketch how Eq. (12.3) follows from Eq. (12.2). First, since the set of invertible matrices is dense and both sides of the equation are continuous, we may assume without loss of generality that X is invertible, so we can write $A = e^{\imath M}$. Now parametrize $M = z_1 I + z_2 X + z_3 Y + z_4 Z$ by a complex vectors $z \in \mathbb{C}^4$. Then both the left-hand side and the right-hand side of Eq. (12.3) are holomorphic functions of $z \in \mathbb{C}^4$. Note note that, for $z \in \mathbb{R}^4$, M is Hermitian, so $e^{\imath M}$ is unitary, and hene Eq. (12.3) reduces to Eq. (12.2). But any two multivariate holomorphic functions that agree on the reals must be equal – this concludes the proof of Eq. (12.3). (Another approach would be to work with the groups SL(2) and GL(2) throughout.)

In particular, we can apply Eq. (12.3) to density operators. We restate the resulting formula, since provides us with a very useful normal form of an i.i.d. quantum state $\rho^{\otimes n}$:

$$\rho^{\otimes n} \cong \bigoplus_k T_\rho^{(n,k)} \otimes I_{m(n,k)}, \tag{12.5}$$

We will use this momentarily.

12.4 Solution of the spectrum estimation problem

How does this help us to solve the spectrum estimation problem? Recall that we are looking for a measurement that commutes with both the action of SU(2) and S_n. Let us write $P_{n,k}$ for the orthogonal projection onto the k-th direct summand in Eq. (12.2). This seems like a plausible candidate! Indeed, it is plain from Eq. (12.2) that $P_{n,k}$ commutes with the action of the unitary group. Does $P_{n,k}$ also commute with the action of S_n? Yes, this in fact follows from Schur's lemma – we will discuss this next time in a more general context. Thus, we have found the desired candidate measurement!

Remark 12.3. Note that this measurement generalizes the swap test discussed in Section 12.2, since for $n = 2$ we have that $P_{2,2} = \Pi_2$ and $P_{2,0} = I - \Pi_2$ (see Theorem 12.2).

Remark 12.4. In physics terminology, the measurement $\{P_{n,k}\}$ measures the total spin $j = k/2$. In your quantum mechanics class you might have discussed the quadratic Casimir operator of SU(2) – this is an observable with eigenvalues proportional to $j(j + 1/2)$, so it can also be used to measure j.

In the remainder of today's lecture, we will analyze the projective measurement $\{P_{n,k}\}$ on $\rho^{\otimes n}$. That is, we would like to compute the probabilities

$$\mathbf{Pr}_{\rho^{\otimes n}}(\text{outcome k}) = \operatorname{tr}\left[\rho^{\otimes n} P_{n,k}\right]. \tag{12.6}$$

Note that these probabilities remain unchanged if we substitute $\rho \mapsto U\rho U^\dagger$ – this holds because $P_{n,k}$ commutes with $U^{\otimes n}$. Since we can always diagonalize ρ by a unitary, we may therefore assume that ρ already a diagonal matrix,

$$\rho = \begin{pmatrix} p & \\ & 1-p \end{pmatrix} \tag{12.7}$$

with $p \geq 1-p$, i.e., $p \in [\frac{1}{2}, 1]$. Our goal will be to show that (12.6) is exponentially small in n for most outcomes k – unless when we can obtain a good estimate of the spectrum from k. We will later see that $\hat{p} := \frac{1}{2}\left(1 + \frac{k}{n}\right)$ will provide such an estimate.

In view of Eq. (12.5), we may compute the probability of measurement outcomes in the following way:

$$\operatorname{tr}\left[\rho^{\otimes n} P_{n,k}\right] = \operatorname{tr}\left[T_\rho^{(n,k)} \otimes I_{m(n,k)}\right] = m(n,k) \operatorname{tr}\left[T_\rho^{(n,k)}\right], \tag{12.8}$$

where we used that by definition $P_{n,k}$ projects onto the k-th direct summand. We will now explain how to bound both factors in Eq. (12.8).

First we consider the number $m(n,k)$, which we remember denote the *multiplicity* of $V_{n,k}$ in $(\mathbb{C}^2)^{\otimes n}$. Equivalently, we can work with SU(2); then $m(n,k)$ denotes the number of times that $\operatorname{Sym}^k(\mathbb{C}^2)$ appears in $(\mathbb{C}^2)^{\otimes n}$. We discussed this problem already in Chapter 11 and saw that we could solve this in a recursive fashion. The key ingredient was the Clebsch-Gordan rule (11.5), which states that

$$\operatorname{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^2 \cong \begin{cases} \operatorname{Sym}^{k+1}(\mathbb{C}^2) \oplus \operatorname{Sym}^{k-1}(\mathbb{C}^2) & \text{if } k > 0 \\ \mathbb{C}^2 = \operatorname{Sym}^1(\mathbb{C}^2) & \text{if } k = 0, \end{cases} \tag{12.9}$$

and this allowed us to successively decompose $(\mathbb{C}^2)^{\otimes n}$:

$$\begin{aligned}
&(\mathbb{C}^2)^{\otimes 1} = \mathbb{C}^2 = \operatorname{Sym}^1(\mathbb{C}^2), \text{ so} \\
&(\mathbb{C}^2)^{\otimes 2} = \operatorname{Sym}^1(\mathbb{C}^2) \otimes \mathbb{C}^2 \cong \operatorname{Sym}^2(\mathbb{C}^2) \oplus \operatorname{Sym}^0(\mathbb{C}^2), \text{ so} \\
&(\mathbb{C}^2)^{\otimes 3} = \left(\operatorname{Sym}^2(\mathbb{C}^2) \oplus \operatorname{Sym}^0(\mathbb{C}^2)\right) \otimes \mathbb{C}^2 = \operatorname{Sym}^3(\mathbb{C}^2) \oplus \operatorname{Sym}^1(\mathbb{C}^2) \oplus \operatorname{Sym}^1(\mathbb{C}^2), \text{ etc.}
\end{aligned}$$

E.g., for $n = 3$, we find that $m(3,3) = 1$ and $m(3,1) = 2$, while all other $m(3,k) = 0$.

This process is visualized in Fig. 12.2 and the general result is as follows: The multiplicity $m(n,k)$ of $V_{n,k}$ in $(\mathbb{C}^2)^{\otimes n}$ is precisely equal to the number of paths from $(0,0)$ to (n,k) in Fig. 12.2. In particular, we see that $m(n,n) = 1$ (there is only a single path). Moreover, $m(n,k) > 0$ iff $n-k$ is an nonnegative even number (so that the exponent of the determinant in Eq. (12.4) is always a nonnegative integer).

How can we estimate the number of paths? Any path can be specified by a sequence of in total n "ups" and "downs". If u is the number of "ups" then $n-u$ is the number of "downs". Therefore, we must have that $u - (n-u) = k$ in order for the path to end at (n,k). Thus, $u = (n+k)/2$ is fixed and we see that there are at most $\binom{n}{(n+k)/2}$ many paths. (This provides only an upper bound, because paths that go below zero are invalid.) As a consequence, we find that

$$m(n,k) \leq \binom{n}{\frac{n+k}{2}} \leq 2^{nh(\frac{n+k}{2n})} = 2^{nh(\hat{p})}, \tag{12.10}$$

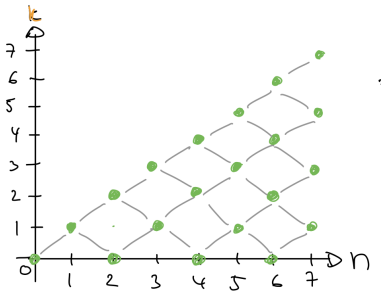


Figure 12.2: By iterating the Clebsch-Gordan rule, we obtain a decomposition of $(\mathbb{C}^2)^{\otimes n}$ into irreducible representations of $U(2)$. The multiplicity $m(n,k)$ is equal to the number of paths from $(0,0)$ to (n,k), where at each step we move to the right and either up or down (unless $k=0$).

where we introduced

$$\hat{p} := \frac{n+k}{2n} = \frac{1}{2}\left(1+\frac{k}{n}\right) \in [\tfrac{1}{2}, 1].$$

The last inequality in Eq. (12.10) is precisley the upper bound (9.2) on the binomial coefficients in terms of the binary Shannon entropy that we derived when compressing coin flips in Chapter 9. Thus, the multiplicites $m(n,k)$ grow at most exponentially, with exponent is given by precisely by the binary Shannon entropy of $\hat{p}$!

We still need to compute the right-hand side trace in Eq. (12.10). In view of Eq. (12.4), this reduces to a trace over the symmetric subspace, which we can compute in our favorite basis (6.1):

$$\begin{aligned}
\operatorname{tr}\left[T_\rho^{(n,k)}\right] &= (\det\rho)^{(n-k)/2}\operatorname{tr}\left[T_\rho^{(k)}\right] = p^{(n-k)/2}(1-p)^{(n-k)/2}\sum_{m=0}^{k}\underbrace{\langle\omega_{m,k-m}|\rho^{\otimes k}|\omega_{m,k-m}\rangle}_{=p^m(1-p)^{k-m}\leq p^k} \\
&\leq (k+1)p^{(n+k)/2}(1-p)^{(n-k)/2} \leq (n+1)p^{(n+k)/2}(1-p)^{(n-k)/2} \\
&= (n+1)2^{n(\hat{p}\log p+(1-\hat{p})\log(1-p))}
\end{aligned} \tag{12.11}$$

For the underbraced inequality, we used that $\rho = \operatorname{diag}(p, 1-p)$ with $p \geq 1-p$ (Eq. (12.7)).

If we plug Eqs. (12.10) and (12.11) back into Eq. (12.8) then we obtain the following bound on the probability of outcomes:

$$\mathbf{Pr}_{\rho^{\otimes n}}(\text{outcome k}) = \operatorname{tr}\left[\rho^{\otimes n}P_{n,k}\right] \leq (n+1)2^{-n\delta(\hat{p}\|p)}, \tag{12.12}$$

where we have introduced the *binary relative entropy*

$$\delta(\hat{p}\|p) = \hat{p}\log\frac{\hat{p}}{p} + (1-\hat{p})\log\frac{1-\hat{p}}{1-p}. \tag{12.13}$$

The relative entropy is an important quantity in information theory and statistics. The point now is that the relative entropy is a distance measure between probability distributions: It is nonnegative and $\delta(\hat{p}\|p) = 0$ if and only if $p = \hat{p}$. (Note however that it is not a metric – e.g., it is *not* symmetric under exchanging $p \leftrightarrow \hat{p}$.) More quantitatively, you will show in Exercise 12.1 that the relative entropy satisfies the following inequality, a special case of the so-called *Pinsker's inequality*:

$$\delta(\hat{p}\|p) \geq \frac{2}{\ln 2}(\hat{p}-p)^2 \tag{12.14}$$

As a consequence, the probability in Eq. (12.12) is exponentially small unless $\hat{p} \approx p$!

This allows us to solve the spectrum estimation problem for qubits: Given $\rho^{\otimes n}$, perform the projective measurement $\{P_{n,k}\}$. Upon outcome k, output $\hat{p} := \frac{1}{2}\left(1 + \frac{k}{n}\right)$ as the estimate of the maximal eigenvalue of ρ. Then:

$$\begin{aligned}\mathbf{Pr}(|\hat{p} - p| \geq \varepsilon) &= \sum_{k:|\hat{p}-p|\geq\varepsilon} \mathbf{Pr}_{\rho^{\otimes n}}(\text{outcome k}) \leq \sum_{k:|\hat{p}-p|\geq\varepsilon} (n+1)2^{-n\delta(\hat{p}\|p)} \\ &\leq \sum_{k:|\hat{p}-p|\geq\varepsilon} (n+1)2^{-n\frac{2}{\ln 2}\varepsilon^2} \leq (n+1)^2 2^{-n\frac{2}{\ln 2}\varepsilon^2},\end{aligned}$$

where we used Eqs. (12.12) and (12.14) and the fact that there are certainly no more than $n+1$ possible values for k. The right-hand side decreases exponentially with n. This means that $\hat{p} \approx p$ with very high probability. Success at last!

Remark 12.5. In Chapter 15, we will discuss how to implement the spectrum estimation measurement concretely by a quantum circuit (see also Theorem 13.1).

Exercises

12.1 **Pinsker inequality:** Show that the binary relative entropy $\delta(p\|q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$ satisfies the following inequality, which is a special case of the so-called *Pinsker inequality*:

$$\delta(p\|q) \geq \frac{2}{\ln 2}(p-q)^2$$

Hint: Remember that $\log x = \ln x / \ln 2$ *is the logarithm to the base two.*

Chapter 13

Universal typical subspaces, Schur-Weyl duality

Yesterday we solved the quantum estimation task by studying the symmetries of the problem. We found that the n-qubit Hilbert space can be decomposed as

$$\left(\mathbb{C}^2\right)^{\otimes n} \cong \bigoplus_k V_{n,k} \otimes \mathbb{C}^{m(n,k)} \tag{13.1}$$

$$X^{\otimes n} \cong \bigoplus_k T_X^{(n,k)} \otimes I_{m(n,k)} \tag{13.2}$$

not only for unitaries but in fact for arbitrary operators X on $\mathbb{C}^2$. We then considered the orthogonal projections $P_{n,k}$ onto the summands in Eq. (13.1). For large n, we found that if we perform the projective measurement $\{P_{n,k}\}$ on $\rho^{\otimes n}$ then

$$\hat{p} := \frac{1}{2}\left(1 + \frac{k}{n}\right) \tag{13.3}$$

provides a good estimate of p, the largest eigenvalue of the unknown density operator ρ. In quantitative terms,

$$\mathbf{Pr}(|\hat{p} - p| \geq \varepsilon) \leq (n+1)^2 2^{-n\delta(\hat{p}\|p)} \leq (n+1)^2 2^{-n\frac{2}{\ln 2}\varepsilon^2}, \tag{13.4}$$

where $\delta(\hat{p}\|p)$ denotes the relative entropy (12.13).

13.1 Universal typical subspaces and protocols

There is another interpretation of what we achieved above. For fixed $\varepsilon > 0$, consider the orthogonal projection

$$P_n := \sum_{k:|\hat{p}-p|<\varepsilon} P_{n,k} \tag{13.5}$$

on all summands k in Eq. (13.1) for which $|\hat{p} - p| < \varepsilon$ (recall from Eq. (13.3) that we think of $\hat{p}$ as a function of k). Then Eq. (13.4) implies that

$$\operatorname{tr}\left[P_n \rho^{\otimes n}\right] = 1 - \mathbf{Pr}(|\hat{p} - p| \geq \varepsilon) \geq 1 - (n+1)^2 2^{-n\frac{2}{\ln 2}\varepsilon^2} \to 1$$

for large n. This means that the $\mathcal{H}_n$ are typical subspaces!

What is the corresponding rate? On the other hand, P_n is a projector onto a subspace $\mathcal{H}_n \subseteq \left(\mathbb{C}^2\right)^{\otimes n}$ of dimension

$$\begin{aligned}\dim \mathcal{H}_n &= \sum_{k:|\hat{p}-p|<\varepsilon} \dim(V_{n,k}) m(n,k) \leq \sum_{k:|\hat{p}-p|<\varepsilon} (k+1) 2^{nh(\hat{p})} \leq \sum_{k:|\hat{p}-p|<\varepsilon} (k+1) 2^{n(h(p)+\varepsilon')} \\ &\leq (n+1)^2 2^{n(h(p)+\varepsilon')}.\end{aligned}$$

The first inequality is Eq. (12.10) and in the second we used that $|\hat{p} - p| < \varepsilon$ ensures that $|h(\hat{p}) - h(p)| < \varepsilon'$ for some ε' that depends only on ε (and which can be made arbitrarily small by choosing ε sufficiently small, by continuity of the binary entropy function). Thus, the rate of the typical subspaces, $\frac{1}{n} \log \dim \mathcal{H}_n$, is arbitrarily close to $h(p) = S(\rho)$, the von Neumann entropy of ρ. This is of course something that we already achieved in Chapter 10. But note that the only input to the construction was p, as is plain from Eq. (13.5). This means that we have constructed *universal typical subspaces*, which can be used for any quantum state whose eigenvalues are $\{p, 1-p\}$!

As a direct consequence, we obtain *universal protocols* for quantum compression and quantum state transfer that work for any quantum state with fixed spectrum. Simply take the protocols in Chapters 9 and 10 and replace the typical subspaces used therein (which were constructed in terms of the eigenbasis of ρ) by the universal typical subspaces constructed above! In fact, one can even obtain compression protocols that, for a given target rate R, work for any qubit source whose density operator satisfies $S(\rho) < R$ (and similarly for quantum state transfer). You discussed this in Exercise 9.3 for classical data compression, and you can work out the quantum case in Exercise 13.1). This universality is one of the main advantages of the symmetries-based approach.

13.2 Schur-Weyl duality

Let us discuss the mathematical machinery that we developed yesterday in some more detail. Our start point is the decomposition (13.1) of the n-qubit Hilbert space as a $U(2)$-representation, restated for your convenience:

$$\left(\mathbb{C}^2\right)^{\otimes n} \cong \bigoplus_k V_{n,k} \otimes \mathbb{C}^{m(n,k)} \tag{13.6}$$

$$X^{\otimes n} \cong \bigoplus_k T_X^{(n,k)} \otimes I_{m(n,k)} \tag{13.7}$$

So far, the Hilbert spaces $\mathbb{C}^{m(n,k)}$ were simply vectors spaces.

Remark 13.1. So far, we have simply argued on abstract grounds that the Hilbert space of n qubits can be decomposed in the form (13.6). 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

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

to a new basis (the "Schur basis")

$$|k, i, j\rangle$$

where $k \in \{\ldots, n-2, n\}$, $i \in \{-k, \ldots, k-2, k\}$, $j \in \{1, \ldots, m(n,k)\}$. Note that the right-hand side is *not* a tensor product of three spaces, because the allowed values for i and j depend on k.

However, we can certainly embed it into a larger space where $|k,i,j\rangle \mapsto |k\rangle \otimes |i\rangle \otimes |j\rangle$ gets mapped to a product basis vector. In Chapter 15 we will learn how to implement this transformation – called the *quantum Schur transform* – by a quantum circuit (see also Theorem 13.4 below).

However, 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 (Theorem 5.6) implies that

$$R_\pi \cong \bigoplus_k I_{V_{n,k}} \otimes R_\pi^{(n,k)} \tag{13.8}$$

for some operators $R_\pi^{(n,k)}$ on $\mathbb{C}^{m(n,k)}$. This is a consequence of the following result, which generalizes part (b) of Schur's lemma:

Lemma 13.2. *Let $\{V_\lambda\}_{\lambda\in\Lambda}$ a collection of pairwise inequivalent irreps of some group G, with Λ an arbitrary index set, and $m(\lambda)$ and $n(\mu)$ nonnegative integers for $\lambda, \mu \in \Lambda$.*

(a) Let $M\colon V_\lambda \otimes \mathbb{C}^{m(\lambda)} \to V_\mu \otimes \mathbb{C}^{n(\mu)}$ be an intertwiner. If $\lambda \neq \mu$, then $M = 0$. If $\lambda = \mu$, then M is of the form $M = I_{V_\lambda} \otimes M_\lambda$ for some operator $M_\lambda\colon \mathbb{C}^{m(\lambda)} \to \mathbb{C}^{n(\lambda)}$.
(b) Any intertwiner $M\colon \bigoplus_\lambda V_\lambda \otimes \mathbb{C}^{m(\lambda)} \to \bigoplus_\mu V_\mu \otimes \mathbb{C}^{n(\mu)}$ is of the form $M = \bigoplus_\lambda I_{V_\lambda} \otimes M_\lambda$, with M_λ as above.

Proof. This is a somewhat painful exercise in applying Schur's lemma.

(a) For every $i = 1, \ldots, n(\mu)$ and $j = 1, \ldots, m(\lambda)$, consider the "block"

$$M_{ij} := \left(I_{V_\mu} \otimes \langle i|\right) M \left(I_{V_\lambda} \otimes |j\rangle\right).$$

This is an operator (!), and in fact an intertwiner $V_\lambda \to V_\mu$. These are irreducible representations, so Schur's lemma applies. If $\lambda \neq \mu$ then the irreps are inequivalent, hence $M_{ij} = 0$, hence $M = 0$. If $\lambda = \mu$ then part (b) of Schur's lemma shows that $M_{ij} \propto I_{V_\lambda}$. Define an operator $M_\lambda\colon \mathbb{C}^{m(\lambda)} \to \mathbb{C}^{n(\lambda)}$ by $M_{ij} = \langle i|M_\lambda|j\rangle\, I_{V_\lambda}$. Then

$$M = \sum_{i,j} M_{ij} \otimes |i\rangle\langle j| = \sum_{i,j} I_{V_\lambda} \otimes |i\rangle\langle i|M_\lambda|j\rangle\langle j| = I_{V_\lambda} \otimes M_\lambda.$$

(b) Apply part (a) to each "block" of M. □

Remark 13.3. In class we only discussed the special case where $m(\lambda) = n(\lambda)$ for all λ (but the more general statement is proved identically, as you saw above).

If we apply part 13.2 of the lemma to $G = U(2)$, $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ then we obtain Eq. (13.8). In particular, this verifies that the R_π commute with the projections $P_{n,k}$ onto the different sectors, as we claimed in the last lecture. Moreover, since the $\{R_\pi\}$ form a representation, the operators $\{R_\pi^{(n,k)}\}$ turn the spaces $\mathbb{C}^{m(n,k)}$ into representations of S_n. Let us denote these representations by $W_{n,k}$. It turns out that the $W_{n,k}$ are irreducible and pairwise inequivalent representations of S_n! We will prove this at the end of this section.

Remark 13.4. Note that we gave no intrinsic definition of the S_n-representations $W_{n,k}$. While the dimensions $m(n,k)$ are uniquely determined, there is more than one intertwiner (13.6) (how many? see the variant of Schur's lemma that we derive in Theorem 13.2 below). However, any choice of intertwiner will yield an *equivalent* S_n-representation. This is because once the intertwiner was fixed, the operators $R_\pi^{(n,k)}$ were uniquely defined in terms of the permutation action on $(\mathbb{C}^2)^{\otimes n}$. It is a useful exercise to work this out in some more detail. The representations $W_{n,k}$ can also be defined without reference to $(\mathbb{C}^2)^{\otimes n}$ – they are called *Specht modules*.

Note, however, that the way that we counted $m(n,k)$ in Section 12.4 gives rise to a less ambiguous definition of an intertwiner (13.6). Indeed, recall that $m(n,k)$ counts the number of paths in Fig. 12.2, and that each path corresponds to following the Clebsch-Gordan decomposition (11.5) such that we arrive at a copy of the irreducible representation $V_{n,k}$. For different paths, these are orthogonal copies are orthogonal (as follows from the unitarity of the Clebsch-Gordan decomposition). Moreover, note that the intertwiner in the Clebsch-Gordan decomposition is unique up to phases (this again follows by Theorem 13.2 below). As a consequence, this procedure identifies an intertwiner (13.6) which is uniquely determined up to a diagonal matrix. We will explain this more clearly in Chapter 15 and use it to derive a quantum circuit for this intertwiner, called the quantum Schur transform!

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

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_k V_{n,k} \otimes W_{n,k} \tag{13.9}$$

which holds as a representation of both $U(2)$ and S_n. The spaces $\{V_{n,k}\}$ and $\{W_{n,k}\}$ are pairwise inequivalent, irreducible representations of $U(2)$ and of S_n, respectively. Equation (13.9) shows that they are "paired up" perfectly in the n-qubit Hilbert space. This is a famous result known as *Schur-Weyl duality*. In Exercise 13.2 you will see how to explicitly realize this isomorphism and construct an intertwiner that implements (13.9) for $n = 3$.

Schur-Weyl duality has a number of important consequences. For one, it implies that any operator that commutes with both the action of $U(2)$ and the action of S_n is necessarily a linear combination of the projections

$$P_{n,k} \cong \bigoplus_{k'} \delta_{k,k'} I_{V_{n,k}} \otimes I_{W_{n,k}}.$$

You can see this by applying Theorem 13.2 to each of the two group actions and comparing the result: Any operator that commutes with the $U^{\otimes n}$ must have the form $\bigoplus_k I_{V_{n,k}} \otimes Y_k$, while any operator that commutes with the R_π must have the form $\bigoplus_k X_k \otimes I_{W_{n,}}$. But $X_k \otimes I_{W_{n,k}} = I_{V_{n,k}} \otimes Y_k$ holds if and only if $X_k \propto I_{V_{n,k}}$ and $Y_k \propto I_{W_{n,k}}$. It follows that an operator that commutes with both group actions is necessarily a linear combination of the $P_{n,k}$, as we claimed. In particular, this means that $\{P_{n,k}\}$ is the most fine-grained projective measurement that has both symmetries of the spectrum estimation problem!

Remark 13.5. We can also interpret Eq. (13.9) as the decomposition of $(\mathbb{C}^2)^{\otimes n}$ with respect to the product group $G = U(2) \times S_n$. Each $V_{n,k} \otimes W_{n,k}$ is an irreducible representation of G (this follows from the argument just given). Conversely, any irreducible representation of the product group is a tensor product of an irreducible $U(2)$-representation with an irreducible S_n-representation (a pleasant exercise using Schur's lemma).

Proof of Schur-Weyl duality

We still need to show that the $W_{n,k}$ are irreducible and pairwise inequivalent. We first prove a useful lemma (for general d, not just $d = 2$):

Lemma 13.6. *Let Y be an operator on $(\mathbb{C}^d)^{\otimes n}$ that commutes with R_π for every $\pi \in S_n$. Then Y can be written as a linear combination of operators of the form $X^{\otimes n}$.*

We will give two proofs – one concrete and one abstract proof.

First proof. Since $Y = \sum_{\pi \in S_n} R_\pi Y R_\pi^\dagger$, it suffices to show that any operator of the form

$$\sum_{\pi \in S_n} R_\pi Z R_\pi^\dagger$$

can be writen as a linear combination of $X^{\otimes n}$'s. Since any operator Z can be written as a linear combination of operators of the form $Z_1 \otimes \ldots \otimes Z_n$, it suffices to prove the claim for a single such $Z = Z_1 \otimes \ldots \otimes Z_n$. Now we can use the following trick

$$\partial_{s_1=0} \ldots \partial_{s_n=0} \left(\sum_{i=1}^n s_i Z_i \right)^{\otimes n} = \sum_{\pi \in S_n} R_\pi \left(Z_1 \otimes \ldots \otimes Z_n \right) R_\pi^\dagger, \tag{13.10}$$

and the claim follows because the left-hand side is a limit of linear combinations of operators of the form $X^{\otimes n}$, and hence also a linear combination of such operators (finite-dimensional vector spaces are closed; we used a similar argument in Chapter 6). □

Example 13.7. It might be instructive to consider an example to clarify why Eq. (13.10) holds. For $n = 2$,

$$\partial_{s_1=0} \partial_{s_2=0} \left(s_1 Z_1 + s_2 Z_2 \right)^{\otimes 2} = \partial_{s_1=0} \left(Z_2 \otimes (s_1 Z_1 + s_2 Z_2) + (s_1 Z_1 + s_2 Z_2) \otimes Z_2 \Big|_{s_2=0} \right)$$
$$= \partial_{s_1=0} \left(Z_2 \otimes (s_1 Z_1) + (s_1 Z_1) \otimes Z_2 \right) = Z_2 \otimes Z_1 + Z_1 \otimes Z_2$$

and now it is clear how to prove the general case.

Second proof. Write $L(\mathcal{H})$ for the complex vector space of linear operators on some $\mathcal{H}$. We have a canonical isomorphism $L(\mathcal{H})^{\otimes k} \cong L(\mathcal{H}^{\otimes k})$. Permuting the tensor factors of $L(\mathcal{H})^{\otimes k}$ corresponds precisely to conjugating an operator $Y \in L(\mathcal{H}^{\otimes k})$ with the corresponding permutation operator R_π! Therefore, $\mathrm{Sym}^k(L(\mathcal{H})) \cong \{Y : [Y, R_\pi] = 0\}$. But we know that the vectors (operators!) $X^{\otimes k}$ form an overcomplete basis of the symmetric subspace (from Eq. (4.9)), so the claim follows. □

Theorem 13.6 gives us a way of producing contradictions by exhibiting operators that commute with S_n but which are not linear combination of $X^{\otimes n}$'s, i.e., not of the form

$$\sum_i z_i X_i^{\otimes n} = \bigoplus_k \left(\sum_i z_i T_X^{(n,k)} \right) \otimes I_{W_{n,k}}. \tag{13.11}$$

We will use this to prove that the $W_{n,k}$ are irreducible and pairwise equivalent.

First, assume for sake of finding a contradiction that $W_{n,k}$ was not irreducible. Then we could decompose

$$W_{n,k} = W_{n,k,1} \oplus W_{n,k,2}$$

as an orthogonal direct sum of two nontrivial invariant subspaces. Let $Q^{(n,k)}$ denote the projector onto the first summand. Then

$$\bigoplus_{k'} \delta_{k,k'} I_{V_{n,k}} \otimes Q^{(n,k)}$$

is an intertwiner for the S_n action which is clearly not of the form (13.11) – this is the desired contradiction!

We now show that no two $W_{n,k}$ are equivalent. Again, we assume for sake of finding a contradiction that W_{n,k_1} and W_{n,k_2} are equivalent, where $k_1 \neq k_2$. This means that there exists a nontrivial intertwiner $J \colon W_{n,k_1} \to W_{n,k_2}$. We can lift this to obtain intertwiner for the S_n-action on $(\mathbb{C}^2)^{\otimes n}$ by sending a copy of W_{n,k_1} onto a copy of W_{n,k_2}, say

$$|0\rangle_{V_{n,k_2}} \langle 0|_{V_{n,k_1}} \otimes J.$$

Again this is not of the form (13.11) – in this case because the latter operators have no "off-diagonal blocks" with respect to k. This is the desired contradiction. □

It is also true that any operator that commutes with every $U^{\otimes n}$ is necessarily a linear combination of the operators R_π (compare this with Theorem 13.6). Mathematically, we say that the two representations span each other's *commutants*. We will prove this momentarily after a preparatory lemma.

Lemma 13.8. *Let Y be an operator on $(\mathbb{C}^d)^{\otimes n}$ that commutes with $U^{\otimes n}$ for every $U \in U(d)$. Then Y commutes with $X^{\otimes n}$ for every operator X on $\mathbb{C}^d$.*

Proof. Let M be a Hermitian operator.

$$e^{is\widetilde{M}} Y e^{-is\widetilde{M}} = (e^{isM})^{\otimes n} Y (e^{-isM})^{\otimes n} = Y$$

for every $s \in \mathbb{R}$. Taking the derivative at $s = 0$, it follows that $i\widetilde{M}Y - iY\widetilde{M} = 0$, i.e., $[\widetilde{M}, Y] = 0$. Clearly, this implies that $[\widetilde{M}, Y] = 0$ for *arbitrary* operator M, whether Hermitian or not. But then

$$[(e^M)^{\otimes n}, Y] = [e^{\widetilde{M}}, Y] = 0$$

(write the matrix exponential $e^{\widetilde{M}}$ as a power series; it commutes term by term with Y). Any invertible operator can be written in the form $X = e^M$, and we can extend the claim by continuity to arbitrary X. □

Lemma 13.9. *Let Y be an operator on $(\mathbb{C}^d)^{\otimes n}$ that commutes with $U^{\otimes n}$ for every $U \in U(d)$. Then Y can be written as linear combination of the operators R_π for $\pi \in S_n$.*

Proof. Let $\mathcal{H} := (\mathbb{C}^d)^{\otimes n}$ and consider the maximally entangled state in the doubled Hilbert space,

$$|\Phi\rangle := \sum_x |x\rangle \otimes |x\rangle \in \mathcal{H} \otimes \mathcal{H},$$

where $|x\rangle$ denotes some basis of $\mathcal{H}$ (perhaps the computational basis). It is enough to show that $(Y \otimes I)|\Phi\rangle$ can be written as a linear combination of the vectors $(R_\pi \otimes I)|\Phi\rangle$, since we can always recover Y from $(Y \otimes I)|\Phi\rangle$ by using that $(I \otimes \langle\Phi|)(|\Phi\rangle \otimes I) = I$, as in the proof of teleportation.

Why should the above be true? Let us consider $\mathcal{H} \otimes \mathcal{H}$ as a representation of S_n by $R_\pi \otimes I$. Then

$$\mathcal{H}_0 := \operatorname{span}\{(R_\pi \otimes I)|\Phi\rangle : \pi \in S_n\}$$

is an invariant subspace, so the orthogonal projector onto $\mathcal{H}_0$ – let us denote it by P – commutes with $R_\pi \otimes I$ for every $\pi \in S_n$ (a fact that we used many times throughout this course). As a consequence, each block $(I \otimes \langle x|)P(I \otimes |y\rangle)$ commutes with R_π. By Theorem 13.6, this means that

$$P = \sum_{x,y} P_{xy} \otimes |x\rangle\langle y| \quad \text{for certain } P_{xy} \in \operatorname{span}\{X^{\otimes n}\}.$$

At last, we can use the assumption. Since Y commutes with every $U^{\otimes n}$ and hence, by Theorem 13.8, with any $X^{\otimes n}$, it commutes with each P_{xy}, and so $(Y \otimes I)P = P(Y \otimes I)$. As a consequence,

$$(Y \otimes I) |\Phi\rangle = (Y \otimes I)P |\Phi\rangle = P(Y \otimes I) |\Phi\rangle \in \mathcal{H}_0,$$

which is what we wanted to show. □

It is instructive to compare Theorem 13.9 with the situation that you analyzed in Exercise 7.4, which was a very special case. Theorem 13.9 is highly useful to compute averages with respect to the uniform probability distribution on pure states (Eq. (4.3)) or with respect to the Haar measure of the unitary group, which we will introduce next week (Eq. (14.4)). For example, for any operator Z on $(\mathbb{C}^d)^{\otimes n}$, $Y := \int dU\, U^{\otimes n} Z U^{\dagger,\otimes n}$ has these symmetries and hence can be written as a linear combination of the permutation operators R_π.

Exercises

13.1 **Universal quantum data compression:** In class, we discussed a quantum compression protocol that works for all qubit ensembles $\{p_x, |\psi_x\rangle\}$ for which the associated density operator $\rho = \sum_x p_x |\psi_x\rangle \langle\psi_x|$ has given eigenvalues $\{p, 1-p\}$.

Your task in this exercise is to design a *universal compression protocol* that works for all qubit ensembles with $S(\rho) < S_0$, where $S_0 > 0$ is a given target compression rate.

(a) Show that, for all $S_0 > 0$, there exist projectors $\tilde{P}_n$ on subspaces $\tilde{\mathcal{H}}_n$ of $(\mathbb{C}^2)^{\otimes n}$ such that:
   (a) For all density operators ρ with $S(\rho) < S_0$, $\operatorname{tr}\left[\tilde{P}_n \rho^{\otimes n}\right] \to 1$ as $n \to \infty$,
   (b) The dimension of $\tilde{\mathcal{H}}_n$ is at most $2^{n(S_0+\delta(n))}$ for some function δ with $\delta(n) \to 0$ as $n \to \infty$.

   Hint: Use the spectrum estimation projectors P_j in a clever way.

(b) Use the projectors $\tilde{P}_n$ to construct a compression protocol with compression rate S_0 that works for all qubit ensembles with $S(\rho) < S_0$ (i.e., show that in the limit of large block length n, the average squared overlap between the original state and the decompressed state goes to one).
   Hint: Follow the same construction as in Chapter 9.

13.2 **Schur-Weyl duality:** Your goal in this exercise is to concretely identify irreducible representations of $U(2)$ and of S_n in the n-qubit Hilbert space, and to explicitly realize the Schur-Weyl duality in a special case. Let $k \in \{0, 1, \ldots, n\}$ be an integer such that $n - k$ is even.

(a) Show that the invariant subspace

$$V'_{n,k} := \left\{ |\phi\rangle \otimes |\Psi^-\rangle^{\otimes (n-k)/2} : \ |\phi\rangle \in \operatorname{Sym}^k(\mathbb{C}^2) \right\} \subseteq (\mathbb{C}^2)^{\otimes n}$$

is an irreducible $U(2)$-representation equivalent to $V_{n,k}$. As always, $|\Psi^-\rangle$ denotes the singlet, and $U(2)$ acts on $(\mathbb{C}^2)^{\otimes n}$ by $U^{\otimes n}$. How can you obtain further $U(2)$-representations in $(\mathbb{C}^2)^{\otimes n}$ that are equivalent to $V_{n,k}$?
   Hint: Recall the symmetry of the singlet state from Exercise 3.5.

(b) Show that the invariant subspace

$$W'_{n,k} := \mathrm{span}\Big\{ R_\pi \left(|0\rangle^{\otimes k} \otimes |\Psi^-\rangle^{\otimes (n-k)/2} \right) : \pi \in S_n \Big\} \subseteq (\mathbb{C}^2)^{\otimes n}$$

is an irreducible S_n-representation equivalent to $W_{n,k}$. How can you obtain further S_n-representations in $(\mathbb{C}^2)^{\otimes n}$ equivalent to $W_{n,k}$?
Hint: You are allowed to use the statement of Schur-Weyl duality.

Now consider the case of three qubits. Here, $n = 3$, so the only two options for k are $k = 1, 3$.

(c) Show that $W_{3,3}$ is equivalent to the trivial representation $\mathbb{C}$, while $W_{3,1}$ is equivalent to the two-dimensional irreducible representation $\mathcal{H} = \{(\alpha, \beta, \gamma) : \alpha + \beta + \gamma = 0\}$ from Exercise 5.1.

(d) Construct a unitary operator $(V_{3,3} \otimes \mathbb{C}) \oplus (V_{3,1} \otimes \mathcal{H}) \to (\mathbb{C}^2)^{\otimes 3}$ that is an intertwiner for the actions of both $U(2)$ and S_3.

Hint: In (c), construct an explicit intertwiner $\mathcal{H} \cong W'_{3,1}$ that you can re-use in (d).

Chapter 14

Quantum state tomography

Today, we will solve the task of estimating an unknown quantum state given many copies – a task that is also known as *quantum state tomography*. We previously solved this for pure states (Chapter 4), but now we allow arbitrary density operator ρ, which is significantly more challenging. Thus, given $\rho^{\otimes n}$, we would like to design a POVM measurement that yields an estimate $\hat{\rho} \approx \rho$ with high probability,

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

First, however, we will generalize the fidelity from pure states to arbitrary density operators. It will be convenient in the analysis of our tomography measurement.

14.1 The fidelity between quantum states

In Section 7.4 we defined the *trace distance*

$$T(\rho, \sigma) = \max_{0 \leq Q \leq I_{\mathcal{H}}} \operatorname{tr}[Q(\rho - \sigma)]$$

as a distance measure between density operators (whether pure or mixed).

Another very useful measure was the *fidelity*, which we defined for pure states as the overlap $|\langle\phi|\psi\rangle|$ and used numerous times in our analyses. The *fidelity* also generalizes nicely to mixed states. For arbitrary density operators ρ and σ on $\mathcal{H} =: \mathcal{H}_A$, we define it by

$$F(\rho, \sigma) := \sup_{R, |\Psi_{AR}\rangle, |\Phi_{AR}\rangle} |\langle\Psi_{AR}|\Phi_{AR}\rangle|, \tag{14.1}$$

where we optimize over arbitrary Hilbert spaces $\mathcal{H}_R$ such that there exist purifications Ψ_{AR} of ρ as well as Φ_{AR} of σ. The fidelity is well-defined since you know from Chapter 8 that such purifications always exist for $\mathcal{H}_R := \mathcal{H}$. Thus, $0 \leq F(\rho, \sigma) \leq 1$, just as for pure states. Moreover, $F(\rho, \sigma) = 1$ if and only if $\rho = \sigma$ (the "only if" follows from the upper bound in Eq. (14.2) below). Note that, *by definition*, the fidelity has a nice operational interpretation: It is close to one if and only if there exist two purifications with overlap close to one.

When $\rho = |\phi\rangle\langle\phi|$ and $\sigma = |\psi\rangle\langle\psi|$ are themselves pure, then any purification is a tensor product (Eq. (8.1)). Using this observation, it is not hard to see that in this case $F(\rho, \sigma) = |\langle\phi|\psi\rangle|$, so we recover our definition for pure states.

The fidelity is *monotonic* with respect to partial traces:

$$F(\rho_A, \sigma_B) \geq F(\rho_{AB}, \sigma_{AB})$$

This follows directly from the observation that any purification of ρ_{AB} can be interpreted as a purification of ρ_A, and likewise for σ_{AB} and σ_A. (In Exercise 7.7 you proved that the trace distance satisfies a similar monotonicity property, but with "$\leq$".)

When ρ or σ is mixed, it is not longer the case that there is a one-to-one relation between fidelity and trace distance. In general, the trace distance and fidelity are related by the following *Fuchs-van de Graaf inequalities*:

$$1 - F(\rho, \sigma) \leq T(\rho, \sigma) \leq \sqrt{1 - F^2(\rho, \sigma)} \tag{14.2}$$

The upper bound is easy to prove: For any two purifications $|\Psi_{AR}\rangle$ of ρ and $|\Phi_{AR}\rangle$ of σ, we have $T(\rho\sigma) \leq T(\Psi_{AR}, \Phi_{AR}) = \sqrt{1 - |\langle\Psi|\Phi\rangle|^2}$ by the relationship between trace distance and fidelity for pure states that you proved in (f) of Exercise 2.4. If we optimize over all purifications we obtain the upper bound in Eq. (14.2). We will not prove (nor need) the lower bound.

A highly useful property that makes the fidelity more amenable to calculations is the fact that in Eq. (14.1) we can in fact restrict to a single Hilbert space $\mathcal{H}_R$ such that there exist purificiations of both ρ and σ on $\mathcal{H}_A \otimes \mathcal{H}_R$. You can prove this using the results of Exercise 7.5, from where you also know that $\mathcal{H}_R = \mathcal{H}_A$ is a valid such choice. In particular, it follows that the supremum is in fact a maximum! Using this fact, it is not too hard to establish the following alternative formula for the fidelity:

$$F(\rho, \sigma) = \operatorname{tr} \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} = \operatorname{tr} \sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}. \tag{14.3}$$

As in Exercise 7.5, $\sqrt{M}$ denotes the square root of a positive semidefinite operator M, defined by taking the square root of all eigenvalues.

Remark 14.1. This can also be written as $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$, where $\|X\|_1 := \operatorname{tr}[\sqrt{X^\dagger X}] = \operatorname{tr}[\sqrt{XX^\dagger}]$ is the trace norm for arbitrary (not necessarily Hermitian) operators. It can be calculated as the sum of the singular values of X (for a Hermitian operator, the singular values are the absolute values of the eigenvalues, so this is a proper generalization).

14.2 The measurement

The spectrum estimation measurement $\{P_{n,k}\}$ on $(\mathbb{C}^2)^{\otimes n}$ had a single outcome k, corresponding to the estimate $\hat{p} := \frac{1}{2}\left(1 + \frac{k}{n}\right)$. The key idea is that we would like to refine this measurement and design a POVM measurement $\{Q_{k,U}\}$ with *two outcomes* – k and U – such that our estimate for the unknown density operator is

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

Thus, the outcome U is a unitary operator that determines the eigenbasis of $\hat{\rho}$. (We should perhaps write $Q_{n,k,U}$ instead of $Q_{k,U}$ to indicate that these are operators on $(\mathbb{C}^2)^{\otimes n}$. But the notation as is is already quite a mouthful so we will keep n implicit in the notation.)

The POVM $\{Q_{k,U}\}$ has both a discrete and a continuous outcome, so we know from Section 4.1 that we need to choose a reference measure on the space of outcomes. For k we will use the counting measure ($\int dk = \sum_k$, see Theorem 4.2), but which measure should we choose on $U(2)$? Guided by symmetry, we will choose the *Haar probability measure* dU, which is the unique probability measure such that

$$\int dU\, f(U) = \int dU\, f(VUW) \tag{14.4}$$

for any two unitaries $V, W \in U(2)$ (we say that the measure is "left-invariant" and "right-invariant"). In other words, if U is a Haar-random unitary (i.e., a random unitary with distribution the Haar measure dU) then so is VUW, which can be interpreted as saying that we do not privilege any unitary over any other.

Remark 14.2. We asked a similar question in the case of the POVM for pure state estimation. There, we chose the "uniform" probability distribution $d\psi$ on the set of pure states, which was likewise natural. In mathematical terms, if ψ is a random pure state drawn from $d\psi$ and V an arbitrary fixed unitary then $V\psi V^\dagger$ has the same distribution as ψ (see Equation (4.3)), and we said that $d\psi$ is the uniquely probability measure with this property. It is not hard to verify that if U is a Haar-random unitary then $U |0\rangle \langle 0| U^\dagger$ is a random pure state with distribution $d\psi$.

Thus, in order for $\{Q_{k,U}\}$ to be a POVM, we need that $Q_{k,U} \geq 0$ as well as

$$\sum_k \int dU \, Q_{k,U} = I. \tag{14.5}$$

Moreover, we would like for the POVM $\{Q_{k,U}\}$ to be a refinement of $\{P_{n,k}\}$, so that the k 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 k as if we performed the measurement $\{P_{n,k}\}$. Since $\mathbf{Pr}_\sigma(\text{outcome } k) = \int dU \operatorname{tr}[Q_{k,U}\sigma]$, this means that we would like to demand that

$$\int dU \, Q_{k,U} = P_{n,k} \tag{14.6}$$

which clearly implies Eq. (14.5) (since we know that $\{P_{n,k}\}$ is a measurement).

The ansatz

What could such a POVM look like? We will make the following ansatz:

$$Q_{k,U} \propto P_{n,k}\hat{\rho}^{\otimes n}P_{n,k} = P_{n,k}U^{\otimes n}\begin{pmatrix}\hat{p} & \\ & 1-\hat{p}\end{pmatrix}^{\otimes n} U^{\dagger,\otimes n}P_{n,k} \tag{14.7}$$

for a proportionality constant that we still need to determine.

To see that this is natural, we observe that, for $k = n$, $P_{n,n} = \Pi_n$, the projector onto the symmetric subspace $\operatorname{Sym}^n(\mathbb{C}^2)$. Moreover, in this case $\hat{p} = 1$, so $\hat{\rho} = U |0\rangle \langle 0| U^\dagger =: |\hat{\psi}\rangle \langle\hat{\psi}|$ is a pure state, so $|\hat{\psi}\rangle^{\otimes n}$ is already contained in the symmetric subspace, hence

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

The right-hand side is exactly proportional to the uniform POVM (4.10) that we used for pure state estimation in Chapter 4 – that's already an encouraging sign!

Moreover, note that $Q_{k,U}$ has permutation symmetry (i.e., $[R_\pi, Q_{k,U}] = 0$) and that it is *covariant* with respect to the unitary group in the following sense: For all $V \in U(2)$,

$$= \operatorname{tr}\left[\rho^{\otimes n}Q_{k,U}\right] = \operatorname{tr}\left[V^{\otimes n}\rho^{\otimes n}V^{\dagger,\otimes n}V^{\otimes n}Q_{k,U}V^{\dagger,\otimes n}\right] \operatorname{tr}\left[(V\rho V^\dagger)^{\otimes n}Q_{k,VU}\right].$$

Note that if $Q_{k,U}$ corresponds to $\hat{\rho}$ then $Q_{k,VU}$ corresponds to $V\hat{\rho}V^\dagger$. What this means is that the following two experiments produce the same result:

(a) Prepare $(V\rho V^\dagger)^{\otimes n}$ and measure the POVM $\{Q_{k,U}\}$.
(b) Prepare $\rho^{\otimes n}$, measure the POVM $\{Q_{k,U}\}$, with outcome $\hat{\rho}$, and report $V\hat{\rho}V^\dagger$.

We could summarize this as

$$\rho \mapsto V\rho V^\dagger \quad \rightsquigarrow \quad \hat{\rho} \mapsto V\hat{\rho}V^\dagger.$$

The proportionality constant

We now show that we can choose a suitable normalization constant in Eq. (14.7) so that Eq. (14.6) holds true. The key observation is that with respect to the Schur-Weyl duality

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_k V_{n,k} \otimes W_{n,k}$$

we can use our usual equation Eq. (12.3) (with $A = \hat{\rho}$) to write

$$Q_{k,U} \propto P_{n,k} \hat{\rho}^{\otimes n} P_{n,k} \cong T_{\hat{\rho}}^{(n,k)} \otimes I_{W_{n,k}}$$

(we omit the $\bigoplus_{k'} \delta_{k,k'}$). We can thus calculate

$$\int dU\, P_{n,k} \hat{\rho}^{\otimes n} P_{n,k} \cong \underbrace{\int dU\, T_{\hat{\rho}}^{(n,k)}}_{\propto I_{V_{n,k}}} \otimes I_{W_{n,k}}. \tag{14.8}$$

The underbraced equation is a consequence of Schur's lemma! Indeed, the indicated operator is a self-intertwiner on the irreducible representation $V_{n,k}$, since

$$\begin{aligned} T_V^{(n,k)} \int dU\, T_{\hat{\rho}}^{(n,k)} &= T_V^{(n,k)} \int dU\, T_U^{(n,k)} T^{(n,k)}_{\left(\begin{smallmatrix} \hat{p} & \\ & 1-\hat{p} \end{smallmatrix}\right)} T_{U^\dagger}^{(n,k)} = \int dU\, T_{VU}^{(n,k)} T^{(n,k)}_{\left(\begin{smallmatrix} \hat{p} & \\ & 1-\hat{p} \end{smallmatrix}\right)} T_{U^\dagger}^{(n,k)} \\ &= \int dU\, T_U^{(n,k)} T^{(n,k)}_{\left(\begin{smallmatrix} \hat{p} & \\ & 1-\hat{p} \end{smallmatrix}\right)} T_{U^\dagger V}^{(n,k)} = \int dU\, T_U^{(n,k)} T^{(n,k)}_{\left(\begin{smallmatrix} \hat{p} & \\ & 1-\hat{p} \end{smallmatrix}\right)} T_{U^\dagger}^{(n,k)} T_V^{(n,k)} = \int dU\, T_{\hat{\rho}}^{(n,k)} T_V^{(n,k)} \end{aligned}$$

Here we used repeatedly that $T_{XY}^{(n,k)} = T_X^{(n,k)} T_Y^{(n,k)}$, which is clear from Eq. (12.4). In the third step we used that the integral is invariant under the substitution $U \mapsto V^\dagger U$.

Equation (14.8) shows that

$$\int dU\, P_{n,k} \hat{\rho}^{\otimes n} P_{n,k} \propto P_{n,k}, \tag{14.9}$$

so it remains to figure out the correct normalization constant to turn this into an equality. As usual, we only need to compare traces. On the one hand, we have

$$\operatorname{tr}\left[P_{n,k} \hat{\rho}^{\otimes n} P_{n,k}\right] = \operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right] \dim W_{n,k}$$

This trace not depend on U, so it is equal to the trace of the left-hand side operator in Eq. (14.9). On the other hand, the trace of the right-hand side operator simply

$$\operatorname{tr}\left[P_{n,k}\right] = \dim V_{n,k} \dim W_{n,k} = (k+1) \dim W_{n,k}$$

We conclude that the appropriately normalized POVM elements are given by

$$Q_{k,U} = \frac{k+1}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right]} P_{n,k} \hat{\rho}^{\otimes n} P_{n,k}. \tag{14.10}$$

14.3 Analysis of the measurement

We follow the approach of [HHJ+16] (cf. [Key06, OW16, OW17a] and the wonderful survey [OW17b]). Similarly to when we analyzed the spectrum estimation measurement, we will show that the probability density $\operatorname{tr}[Q_{k,U}\rho^{\otimes n}]$ is exponentially small unless $\rho \approx \hat{\rho}$. We will need to use the full strength of the Schur-Weyl toolbox.

We start with

$$\begin{aligned}
\operatorname{tr}\left[Q_{k,U}\rho^{\otimes n}\right] &= \frac{k+1}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right]} \operatorname{tr}\left[P_{n,k}\hat{\rho}^{\otimes n}P_{n,k}\rho^{\otimes n}\right] = \frac{k+1}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right]} \operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}T_{\rho}^{(n,k)} \otimes I_{W_{n,k}}\right] \\
&= \frac{(k+1)m(n,k)}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right]} \operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}T_{\rho}^{(n,k)}\right] = \frac{(k+1)m(n,k)}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right]} \operatorname{tr}\left[T_{\sqrt{\rho}\rho\sqrt{\rho}}^{(n,k)}\right] \\
&= \frac{(k+1)m(n,k)}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right]} \operatorname{tr}\left[T_{\sqrt{\sqrt{\rho}\rho\sqrt{\rho}}^{2}}^{(n,k)}\right] \leq \frac{(k+1)2^{nh(\hat{p})}}{\operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right]} \operatorname{tr}\left[T_{\sqrt{\sqrt{\rho}\rho\sqrt{\rho}}^{2}}^{(n,k)}\right]
\end{aligned} \tag{14.11}$$

We first used Eq. (14.10), then Eq. (12.3), then that $T_{XY}^{(n,k)} = T_X^{(n,k)}T_Y^{(n,k)}$ as well as the cyclicity of the trace, and finally the upper bound $m(n,k) \leq 2^{nh(\hat{p})}$ from Eq. (12.10).

We need to find a lower bound on $\operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right]$ and an upper bound on $\operatorname{tr}\left[T_{X^2}^{(n,k)}\right]$, where $X := \sqrt{\sqrt{\rho}\rho\sqrt{\rho}}$ is the operator whose trace is the fidelity (Eq. (14.3))! (We cannot use the upper bound (12.11) since X^2 is not necessarily a density operator.) To obtain these, we proceed as in Eq. (12.11):

$$\begin{aligned}
\operatorname{tr}\left[T_{\hat{\rho}}^{(n,k)}\right] &= (\det\hat{\rho})^{(n-k)/2}\, T_{\hat{\rho}}^{(k)} = (\hat{p}(1-\hat{p}))^{(n-k)/2}\, T^{(k)}_{\begin{pmatrix}\hat{p} & \\ & 1-\hat{p}\end{pmatrix}} \\
&= \hat{p}^{(n-k)/2}(1-\hat{p})^{(n-k)/2} \sum_{m=0}^{k} \hat{p}^m(1-\hat{p})^{k-m} \\
&\geq \hat{p}^{(n-k)/2}(1-\hat{p})^{(n-k)/2}\hat{p}^k = \hat{p}^{(n+k)/2}(1-\hat{p})^{(n-k)/2} = 2^{-nh(\hat{p})}
\end{aligned} \tag{14.12}$$

(In contrast to Eq. (12.11), we now evaluate the trace for $\hat{\rho}$, and we now *lower bound* the sum by a single term.) For the upper bound, let us write $\{q, 1-q\}$ for the eigenvalues of $X/\operatorname{tr}[X]$.

$$\begin{aligned}
\operatorname{tr}\left[T_{X^2}^{(n,k)}\right] &= \operatorname{tr}\left[T_{(X/\operatorname{tr}X)^2}^{(n,k)}\right](\operatorname{tr}X)^{2n} = \left(q^2(1-q)^2\right)^{(n-k)/2} T^{(k)}_{\begin{pmatrix}q^2 & \\ & (1-q)^2\end{pmatrix}} (\operatorname{tr}X)^{2n} \\
&= q^{n-k}(1-q)^{n-k} \sum_{m=0}^{k} q^{2m}(1-q)^{2(k-m)}\,(\operatorname{tr}X)^{2n} \\
&\leq q^{n-k}(1-q)^{n-k}(k+1)q^{2k}\,(\operatorname{tr}X)^{2n} \leq (k+1)q^{n+k}(1-q)^{n-k}\,(\operatorname{tr}X)^{2n} \\
&= (k+1)2^{-2n(h(\hat{p})+\delta(\hat{p}\|q))}\,(\operatorname{tr}X)^{2n} \\
&\leq (k+1)2^{-2nh(\hat{p})}F(\hat{\rho},\rho)^{2n}.
\end{aligned} \tag{14.13}$$

We now use Eqs. (14.12) and (14.13) in Eq. (14.11) and obtain:

$$\operatorname{tr}\left[Q_{k,U}\rho^{\otimes n}\right] \leq \frac{(k+1)2^{nh(\hat{p})}}{2^{-nh(\hat{p})}}(k+1)2^{-2nh(\hat{p})}F(\hat{\rho},\rho)^{2n} \leq (n+1)^2F(\hat{\rho},\rho)^{2n}$$

This is the desired upper bound! Indeed, it implies that, for ever y$\varepsilon > 0$,

$$\mathbf{Pr}_{\rho^{\otimes n}}(F(\hat{\rho},\rho) \leq 1-\varepsilon) = \sum_k \int dU\, 1_{[F(\hat{\rho},\rho)\leq 1-\varepsilon]} \operatorname{tr}\left[Q_{k,U}\hat{\rho}^{\otimes n}\right]$$

$$\leq \sum_k \int dU \, 1_{[F(\hat{\rho},\rho)\leq 1-\varepsilon]} (n+1)^2 \, (1-\varepsilon)^{2n} \leq (n+1)^3 \, (1-\varepsilon)^{2n} \, ,$$

($1_{[...]}$ denotes the characteristic function, which is equal to one when the condition is satisfied, and zero otherwise). This expression converges to zero exponentially with n!

We can also express this in terms of the trace distance. E.g.,

$$\mathbf{Pr}_{\rho^{\otimes n}}(T(\hat{\rho},\rho) \geq \varepsilon) = \mathbf{Pr}_{\rho^{\otimes n}}(F(\hat{\rho},\rho) \leq 1-\varepsilon^2) \leq (n+1)^3 \left(1-\varepsilon^2\right)^{2n}$$

where we have used the (easy) upper bound in Eq. (14.2) and the result that we just proved.

14.4 The Schur-Weyl toolbox

Below we assemble 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 two slight generalizations of formulas that we discussed today:

- The lower bound in Eq. (14.14), which is proved just like in Eq. (14.12) except for a general density operator ρ.
- The upper bound in Eq. (14.15), which is proved just like Eq. (14.13) but for general κ.

Schur-Weyl duality:

$$\begin{aligned}
(\mathbb{C}^2)^{\otimes n} &\cong \bigoplus_{k=\ldots,n-2,n} V_{n,k} \otimes W_{n,k}, \\
X^{\otimes n} &\cong \bigoplus_k T_X^{(n,k)} \otimes I_{W_{n,k}}, \quad \text{where} \quad T_X^{(n,k)} := (\det X)^{(n-k)/2} \, T_X^{(k)}, \\
R_\pi &\cong \bigoplus_k I_{V_{n,k}} \otimes R_\pi^{(n,k)}.
\end{aligned}$$

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

Dimensions:

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

There are $\leq n+1$ possible values of k.

Estimates:

$$2^{-n\left[h(\hat{p})+\delta(\hat{p}\|p)\right]} \leq \operatorname{tr}\left[T_\rho^{(n,k)}\right] \leq (k+1) 2^{-n\left[h(\hat{p})+\delta(\hat{p}\|p)\right]} \quad \text{where } \rho \text{ has eigenvalues } \{p, 1-p\}, \tag{14.14}$$

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

$$\operatorname{tr}\left[T^{(n,k)}_{X^\kappa}\right] \leq (k+1)2^{-n\kappa\left[h(\hat{p})+\delta(\hat{p}\|q)\right]}\,(\operatorname{tr} X)^{\kappa n}\,, \quad \text{where } \frac{X}{\operatorname{tr} X} \text{ has eigenvalues } \{q, 1-q\}. \tag{14.15}$$

Spectrum estimation:

$$P_{n,k} \cong \bigoplus_{k'} \delta_{k,k'} I_{V_{n,k}} \otimes I_{W_{n,k}},$$
$$\rho^{\otimes n} \cong \bigoplus_{k} T^{(n,k)}_{\rho} \otimes I_{W_{n,k}} =: \bigoplus_{k} p_k\, \rho_{V_{n,k}} \otimes \tau_{W_{n,k}},$$

and so

$$p_k = \operatorname{tr}\left[P_{n,k}\rho^{\otimes n}\right] \leq (n+1)2^{-n\delta(\hat{p}\|p)} \leq (n+1)2^{-n\frac{2}{\ln 2}(\hat{p}-p)^2}$$
$$\operatorname{tr}\left[P_n\rho^{\otimes n}\right] \geq 1-(n+1)^2 2^{-n\frac{2}{\ln 2}\varepsilon^2}$$

where $P_n := \sum_{k:|\hat{p}-p|<\varepsilon} P_{n,k}$ is the projector onto the universal typical subspace with parameter ε.

Beyond qubits

How does the Schur-Weyl toolbox generalize beyond qubits? This is best explained by making a simple coordinate change and instead of by (n,k) parametrizing all representations by

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

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 $k \geq 0$. Note that the total number of boxes is $\lambda_1 + \lambda_2 = n$, while $k = \lambda_1 - \lambda_2$ is the difference of row lengths.

If we write $V_\lambda := V_{n,k}$ and $W_\lambda := W_{n,k}$, then the Schur-Weyl duality (13.9) becomes

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_{\lambda} V_\lambda \otimes W_\lambda, \tag{14.16}$$

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

Remark 14.3. In Theorems 5.3 and 5.4 and Exercise 5.1 we already discussed the irreducible representations of S_3. In the Young diagram notation, $W_{\ydiagram{3}}$ is the trivial representation and $W_{\ydiagram{2,1}}$ is the two-dimensional representation that you proved to be irreducible in Exercise 5.1. You will verify this in Exercise 13.2. Note that these dimensions agree precisely with $m(3,3) = 1$ and $m(3,1) = 2$, as they should. Together with the sign representation, $W_{\ydiagram{1,1,1}}$, these are all the irreducible representations of S_3 (up to equivalence). Since its Young diagram has three rows, the sign representation does not occur in $(\mathbb{C}^2)^{\otimes 3}$. Indeed, it would correspond to antisymmetric tensors – but the antisymmetric subspace $\bigwedge^3 \mathbb{C}^2 = \{0\}$ is zero-dimensional.

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

$$\bar{\lambda}_1 = \frac{1}{2}\left(1 + \frac{k}{n}\right) = \hat{p}, \quad \bar{\lambda}_2 = \frac{1}{2}\left(1 - \frac{k}{n}\right) = 1 - \hat{p}.$$

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. Similarly, we can describe our POVM measurement by the POVM elements $\{P_{\lambda,U} := P_\lambda \hat{\rho}^{\otimes n} P_\lambda\}$, where $\hat{\rho} = U \operatorname{diag}(\bar{\lambda}) U^\dagger$.

The key point now is the following: Eq. (14.16) generalizes quite directly from qubits to arbitrary d. This is because the relevant irreducible representations of $U(d)$ are labeled by Young diagrams with now (at most) d rows, while the irreps of S_n are labeled by Young diagrams with n boxes. We thus obtain:

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

where we now sum over all Young diagrams with n boxes and at most d rows. All results obtained in this course generalize appropriately. The technical ingredients required for this are, e.g., the Weyl dimension formula (for $\dim V_\lambda$) and the hook length formula (for $\dim W_\lambda$). The trace $\operatorname{tr}[T_X^{(\lambda)}]$ is a so-called character which can be estimated in the same fashion as above (or evaluated more precisely using the Weyl character formula).

Remark 14.4. In fact, note that the core statement of the duality – that pairwise inequivalent irreducible representations of $U(d)$ and of S_n are lined up in "diagonal" fashion – follows from basically identical reasoning as for $d = 2$. Remember that the two main ingredients were that (i) $X^{\otimes n}$ acts block-diagonally with respect to λ and nontrivially on the tensor factors V_λ only (whatever this action looks like), and (ii) that every operator that commutes with all permutations is necessarily in the span of operators of the form $X^{\otimes n}$. Our proof of (i) generalizes readily and both proofs that we gave for (ii) work for arbitrary d (see Theorem 13.6; the first proof does not even rely on the fact that $\operatorname{Sym}^n(\mathbb{C}^d)$ is irreducible).

See, e.g., [FH13, EGH+11, Har05, Chr06, Wal14] for further detail that expand on our very heuristic discussion.

Chapter 15

Quantum circuits, swap test, quantum Schur transform

In the past two weeks we used Schur-Weyl duality as an important tool to solve various information theoretic tasks (Chapters 12 to 14). In particular we often switched back and forth between

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_k V_{n,k} \otimes W_{n,k}, \tag{15.1}$$

using a unitary intertwiner implied by the notation "$\cong$". Mathematically, this is a straightforward operation – but how can we actually realize this transformation in practice? (We posed this question already in Theorems 13.1 and 13.4.)

For our purposes it will be sufficient to worry about the action of the unitary group and ignore the action of permutation group. Indeed, the projections $\{P_{n,k}\}$ that were relevant for spectrum estimation and compression as well as the tomography POVM $\{Q_{k,U}\}$ each act by the identity operator on the S_n-irreps $W_{n,k}$. Moreover, we may restrict to SU(2), since we always know that scalars act by the n-th tensor power (indeed, we derived Eq. (15.1) in Chapter 12 by reasoning about SU(2) alone). Thus what we would like to do is to construct a unitary operator

$$(\mathbb{C}^2)^{\otimes n} \to \bigoplus_k \mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^{m(n,k)} \tag{15.2}$$

that is an intertwiner for SU(2). The n-qubit Hilbert space on the left-hand side has the (computational) product basis

$$|b_1, \ldots, b_n\rangle = |b_1\rangle \otimes \ldots \otimes |b_n\rangle\,,$$

while the right-hand side likewise has a natural basis that we could label

$$|k, m, \boldsymbol{p}\rangle := |\omega_{k,m-n}\rangle \otimes |\boldsymbol{p}\rangle \in \mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^{m(n,k)} \subseteq \bigoplus_k \mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^{m(n,k)}.$$

Here, $k \in \{\ldots, n-2, n\}$ labels the sector, $m \in \{0, 1, \ldots, k\}$ our favorite basis vectors $|\omega_{m,k-m}\rangle$ of the symmetric subspace (Eq. (6.1)), and $\boldsymbol{p}$ the different copies of $\mathrm{Sym}^k(\mathbb{C}^2)$. Why is there a vector sign in $\boldsymbol{p}$? Recall that $m(n,k)$ was precisely the number of paths from $(0,0)$ to (n,k) in Fig. 12.2. We can label any such path by a string $\boldsymbol{p} = p_1 \ldots p_n$, where each $p_i = \pm$ corresponding to making a step to the right and going either up (+) or down (-). (Note that not all such strings correspond to valid paths: some do not arrive at the right endpoint, others go below zero.)

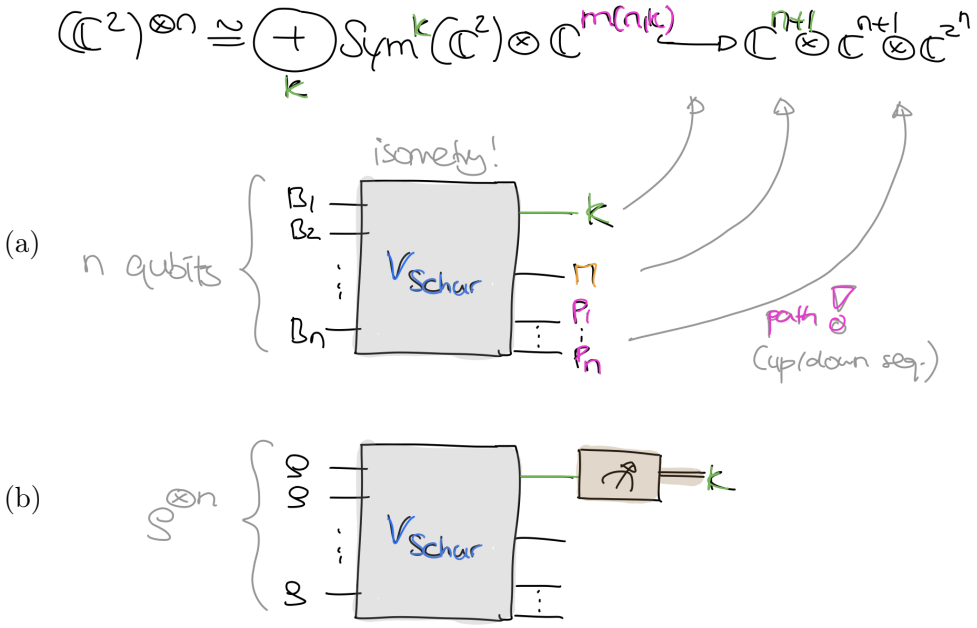


Figure 15.1: (a) The Schur transform (15.3). As usual we label subsystems by upper-case symbols. (b) We can implement the measurement $\{P_{n,k}\}$ by first applying the Schur transform and then measuring the K-system.

Now, since the values of m and $\boldsymbol{p}$ are constrained by k, the vectors $|k, m, \boldsymbol{p}\rangle$ do *not* naturally live in a tensor product space! However, we can safely think of it as a *subspace* of the tensor product space

$$\mathbb{C}^{n+1} \otimes \mathbb{C}^{n+1} \otimes (\mathbb{C}^2)^{\otimes n}$$

since (i) there are at most $n+1$ options for k, (ii) the dimension of $\mathrm{Sym}^k(\mathbb{C}^2)$ is $k+1 \leq n+1$, and (iii) each path $\boldsymbol{p}$ gives rise to a computational basis state $|\boldsymbol{p}\rangle$. Thus, what we will be after is an *isometry*

$$V_{\text{Schur}} \colon (\mathbb{C}^2)^{\otimes n} \longrightarrow \mathbb{C}^{n+1} \otimes \mathbb{C}^{n+1} \otimes (\mathbb{C}^2)^{\otimes n} \tag{15.3}$$

This transformation is called the *quantum Schur transform* (Fig. 15.1, (a)).

Why is this convenient? The isometry nicely separates the three pieces of information that we care about – the sector k and the corresponding data in $V_{n,k}$ and in $\mathbb{C}^{m(n,k)}$ – into three different subsystems. For example, we can now implement the spectrum estimation measurement $\{P_{n,k}\}$ by first applying V_{Schur} and then measuring the K-subsystem. In other words,

$$P_{n,k} = V_{\text{Schur}}^\dagger \left(|k\rangle\langle k|_K \otimes I_M \otimes I_P\right) V_{\text{Schur}}.$$

This is visualized in Fig. 15.1, (b). The goal of today's lecture will be to design a *quantum circuit* for the quantum Schur transform.

15.1 Quantum circuits

Just like we typically describe computer programs or algorithms in terms of simple elementary instructions, in quantum computing we are interested in describing "quantum software" in terms of "simple" building blocks. These building blocks are *quantum gates*, i.e., operations that involve only a smaller number of qubits (or qu*d*its). We obtain a *quantum circuit* by connecting the

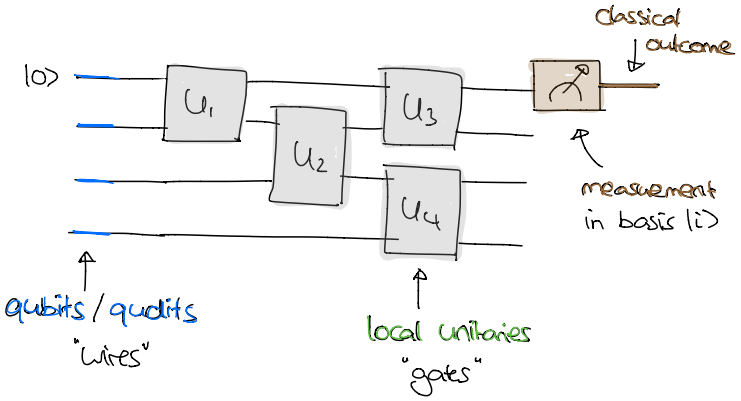


Figure 15.2: Illustration of a quantum circuit, composed of four unitary quantum gates and a single measurement. The first qubit is initialized in state $|0\rangle$ and the other three wires are inputs to the circuit.

output of some quantum gates by "wires" with the inputs of others. We will allow both gates that apply *unitaries* as well *measurements* of individual qubits in the standard basis $\{|i\rangle\}$. In addition, we will allow ourselves to add qubits that are *initialized* in a basis state $|i\rangle$ (such qubits are often called "ancillas"). For example, the circuit in Fig. 15.2 first adds a qubit in state $|0\rangle$, then performs the unitary

$$(U_3 \otimes U_4)\,(I_{\mathbb{C}^2} \otimes U_2 \otimes I_{\mathbb{C}^2})\,(U_1 \otimes I_{\mathbb{C}^2} \otimes I_{\mathbb{C}^2})$$

and then measures one of the qubits. In the absence of measurements and initializations, a quantum circuit performs a unitary transformation from the input qubits to the output qubits. In the absence of measurements alone, but allow initializations, the quantum circuit implements an *isometry* from the input qubits to the outputs qubits.

The number of gates in a quantum circuit is known as the *(gate) complexity* of that circuit. Intuitively, the higher the complexity the longer it would take a quantum computer to run this circuit. This is because we expect that a quantum computer, in completely analogy to a classical computer, will be able to implement each gate and measurement in a small, fixed amount of time. Much of the field of *quantum computation* is concerned with finding quantum circuits and algorithms of minimal complexity – with a particular emphasis on finding quantum algorithms that outperform all known classical algorithms. For example, Peter Shor's famous factoring algorithm outperforms all known classical factoring algorithms. Just like quantum information theory, this is a very rich subject on its own.

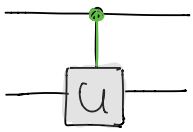
In this course, we only have time for a glance, but I encourage you to look at (or attend!) Ronald de Wolf's lecture notes (see [DW19]) or at the textbook [NC10] for further detail if you are interested in this subject.

To practice, let us consider some interesting gates. For any single-qubit unitary U, there is a corresponding *single-qubit gate*. For example, the Pauli X-operator $X = \left(\begin{smallmatrix} & 1 \\ 1 & \end{smallmatrix}\right)$ gives rise to the so-called X-gate or *NOT-gate*

which maps $X|0\rangle = |1\rangle$, $X|1\rangle = |0\rangle$. Another example is the so-called *Hadamard gate*

which maps $H|0\rangle = |+\rangle$, $H|1\rangle = |-\rangle$. Written as a unitary matrix, $H = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$.

Single-qubit gates are not enough – for example, they do not allow us to create an entangled state starting from product states. A powerful class of gates can be obtained by performing a unitary transformation U depending on the value of a *control qubit*. This standard terminology might be slightly confusing – we do not actually want to measure the value of the control qubit. Instead, we define the *controlled unitary gate*



by

$$\begin{aligned} \mathrm{C}\,U(|0\rangle \otimes |\psi\rangle) &= |0\rangle \otimes |\psi\rangle\,, \\ \mathrm{C}\,U(|1\rangle \otimes |\psi\rangle) &= |0\rangle \otimes (U\,|\psi\rangle) \end{aligned} \tag{15.4}$$

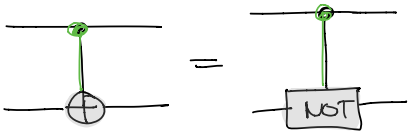
(and extend by linearity). It is easy to see that $\mathrm{C}\,U$ is indeed a unitary (indeed, $\mathrm{C}(U^\dagger)$ is its inverse). For example, if U is the NOT-gate then the *controlled not (CNOT) gate* maps

$$\begin{aligned} \mathrm{CNOT}\,|0,0\rangle &= |0,0\rangle\,, \\ \mathrm{CNOT}\,|0,1\rangle &= |0,1\rangle\,, \\ \mathrm{CNOT}\,|1,0\rangle &= |1,1\rangle\,, \\ \mathrm{CNOT}\,|1,1\rangle &= |1,0\rangle\,, \end{aligned}$$

i.e.,

$$\mathrm{CNOT}\,|x,y\rangle = |x, x \oplus y\rangle\,,$$

where $\oplus$ denotes addition modulo 2. This explains why the CNOT gate is often denoted by



Remark 15.1. More generally, if $U(0)$, $U(1)$ are two unitaries then we can define a controlled unitary that selects one or the other based on the control qubit, i.e.,

$$|x\rangle \otimes |\psi\rangle \mapsto |x\rangle \otimes U(x)\,|\psi\rangle\,.$$

Another possible generalization is to use more than one qubit as the control. For example, the *doubly-controlled unitary* $\mathrm{C}\,\mathrm{C}\,U$ applies U if and only if both control qubits are in the $|1\rangle$ state:

$$\mathrm{C}\,\mathrm{C}\,U(|x\rangle \otimes |y\rangle \otimes |\psi\rangle) = \begin{cases} |x\rangle \otimes |y\rangle \otimes |\psi\rangle\,, & \text{if } x = 0 \text{ or } y = 0, \\ |1\rangle \otimes |1\rangle \otimes U\,|\psi\rangle\,, & \text{if } x = y = 1. \end{cases}$$

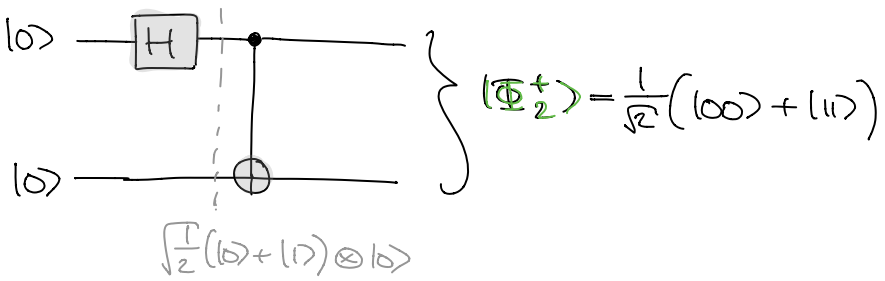
We can also combine these two ideas and use, e.g., two controls to select a unitary from a family $\{U(x,y)\}$. We will use this generalization below when constructing a quantum circuit for the Clebsch-Gordan transformation.

Using these ingredients, we can already build a number of interesting circuits.

Remark 15.2. In fact, any N-qubit unitary can be to arbitrarily high fidelity approximated by quantum circuits composed only of CNOT-gates and single qubit gates. We say, that the CNOT gate together with the single qubit gates form a *universal gate set*. (One can show that, in fact, CNOT together with a finite number of single qubit gates suffices.)

Entanglement and teleportation

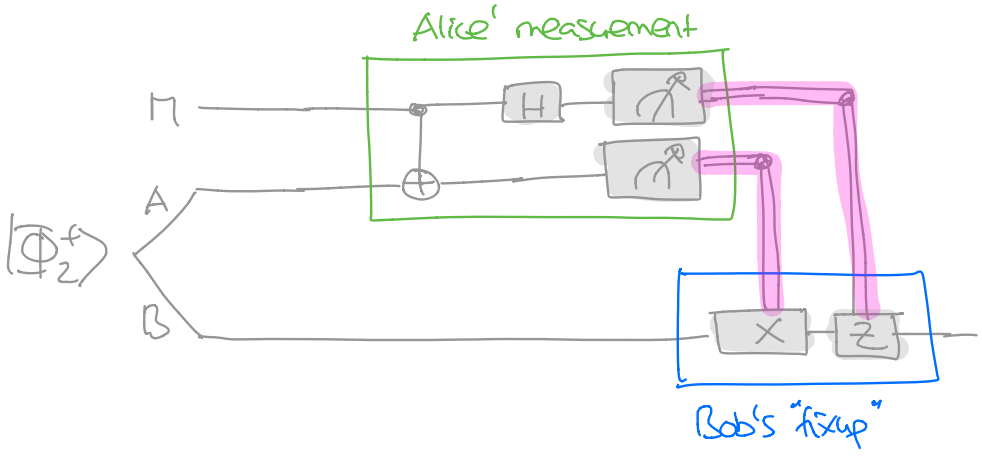
For example, consider the following circuit:



It is plain that this creates an ebit starting from the product state $|00\rangle$. More generally, for each product basis state $|xy\rangle$ the circuit produces one of the four maximally entangled basis vectors $|\phi_k\rangle$ from Eq. (2.2) that we used in superdense coding and teleportation. Indeed, the circuit maps

$$|x, y\rangle \mapsto \frac{1}{\sqrt{2}} \left(|0\rangle + (-1)^x |1\rangle\right) \otimes |y\rangle = \frac{1}{\sqrt{2}} \left(|0, y\rangle + (-1)^x |1, y\rangle\right).$$

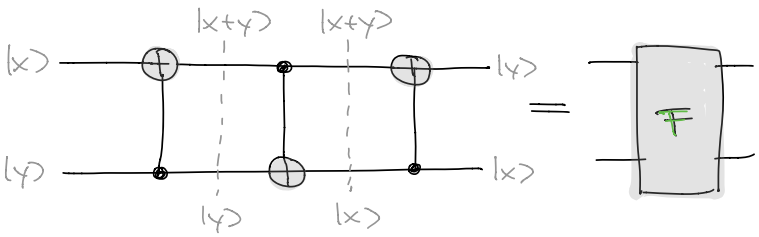
As a consequence, this allows us to write down a more detailed version of the teleportation circuit from Chapter 2:



The doubled wires (pink) denote the classical measurement outcomes (two bits x and y, corresponding to the single integer $k \in \{0, 1, 2, 3\}$ from last time). It is a fun exercise to verify that this circuit works as desired, i.e., that it implements an identity map from the input qubit M to the output qubit B.

15.2 The swap test

We can implement the swap unitary $F\colon |xy\rangle \mapsto |yx\rangle$ by a quantum circuit composed of three CNOTs:



This is called the *swap gate*.

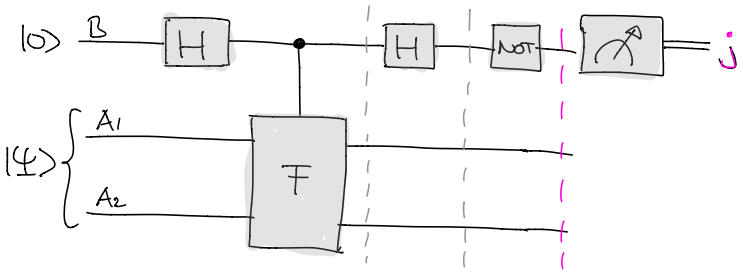
We can also write down a corresponding *controlled swap gate*, defined as in Eq. (15.4) for $U = F$. Note that this is a *three-qubit* gate! The decomposition of the swap gate into three CNOTs immediately yields a decomposition of the controlled swap gate into three CCNOTs – i.e., doubly controlled NOTs, also called *Toffoli gates*. It is not completely straightforward to decompose the Toffoli gate into a quantum circuit that involves only single-qubit and two-qubit gates (Exercise 15.2).

When we started studying the spectrum estimation problem in Chapter 12, we first considered the case that we were given $n = 2$ two copies of our state as a "warmup" (Section 12.2). The idea was that the two-qubit Hilbert space decomposes into the symmetric (triplet) and antisymmetric (singlet) subspaces,

$$\mathbb{C}^2 \otimes \mathbb{C}^2 = \mathrm{Sym}^2(\mathbb{C}^2) \oplus \mathbb{C}\,|\Psi^-\rangle\,.$$

This is of course a special case of Eq. (15.2)! In Section 12.2, we also saw that the corresponding measurement $\{P_{2,2}, P_{2,0}\} = \{\Pi_2, I - \Pi_2\}$ already gave useful information about the spectrum. But how can we implement this measurement by a quantum circuit?

Consider the following circuit, which uses the *controlled* swap gate discussed above:



(15.5)

Why does this circuit perform the desired measurement? Suppose that we initialize the B-wire in state $|0\rangle$ and the A-qubits in some arbitrary two-qubit state $|\Psi\rangle_A = |\Psi\rangle_{A_1A_2}$. The Hadamard gate sends $|0\rangle \mapsto |+\rangle$ and so the quantum state right after the controlled swap gate (first dashed line) is equal to

$$\frac{1}{\sqrt{2}}\left(|0\rangle_B \otimes |\Psi\rangle_A + |1\rangle_B \otimes F\,|\Psi\rangle_A\right)$$

After the second Hadamard gate (second dashed line), we obtain

$$\begin{aligned}
&\frac{1}{2}\left[(|0\rangle_B + |1\rangle_B) \otimes |\Psi\rangle_A + (|0\rangle_B - |1\rangle_B) \otimes F\,|\Psi\rangle_A\right] \\
&= |0\rangle_B \otimes \frac{I+F}{2}\,|\Psi\rangle_A + |1\rangle_B \otimes \frac{I-F}{2}\,|\Psi\rangle_A \\
&= |0\rangle_B \otimes \Pi_2\,|\Psi\rangle_A + |1\rangle_B \otimes (I - \Pi_2)\,|\Psi\rangle_A \\
&= |0\rangle_B \otimes P_{2,2}\,|\Psi\rangle_A + |1\rangle_B \otimes P_{2,0}\,|\Psi\rangle_A\,,
\end{aligned}$$

where $\Pi_2 = P_{2,2}$ is the projector onto symmetric subspace! The NOT gate now simply relabels $|0\rangle_B \leftrightarrow |1\rangle_B$, leading to

$$|1\rangle_B \otimes P_{2,2} |\Psi\rangle_A + |0\rangle_B \otimes P_{2,0} |\Psi\rangle_A .$$

Thus, right up to before the measurement of the B-qubit (last, pink dashed line) the quantum circuit achieves the following isometry:

$$|\Psi\rangle_A \mapsto \sum_{j=0,1} |j\rangle_B \otimes P_{2,2j} |\Psi\rangle_A .$$

For general density operators Γ_A, this means that

$$\Gamma_A \mapsto \Gamma'_{BA} := \sum_{j,j'} |j\rangle \langle j'|_B \otimes P_{2,2j} \Gamma_A P_{2,2j'}.$$

since there were no measurements involved up to this point. As a consequence,

$$\mathbf{Pr}_\Gamma(\text{outcome } j = \frac{k}{2}) = \operatorname{tr}\left[\Gamma'_{BA} \left(|j\rangle \langle j|_B \otimes I_{A_1} \otimes I_{A_2}\right)\right] = \operatorname{tr}\left[\Gamma_A P_{2,2j}\right] = \operatorname{tr}\left[\Gamma_A P_{2,k}\right]$$

and the post-measurement state on the A-qubits is proportional to $P_{2,k}\Gamma_A P_{2,k}$. Thus, we have successfully implemented the projective measurement $\{P_{2,2}, P_{2,0}\}$! The quantum circuit (15.5) is known as the *swap test*.

Applications

The swap test has many applications:

- If we choose $\Gamma = \rho^{\otimes 2}$ as input state for the A-qubits, then

  $$\mathbf{Pr}(\text{outcome } 1) = \operatorname{tr}\left[P_{2,2}\rho^{\otimes 2}\right] = \frac{1}{2}\left(1 + \operatorname{tr}\rho^2\right).$$

  Thus we can estimate the *purity* $\operatorname{tr}\rho^2$ which gives us information about the spectrum of the unknown quantum state ρ. This was our original motivation for implementing the swap test (cf. Section 12.2).

- If we choose the tensor product of two pure states $|\psi\rangle \otimes |\phi\rangle$ as input state,

  $$\mathbf{Pr}(\text{outcome } 1) = \frac{1}{2}\left(1 + |\langle\psi|\phi\rangle|^2\right), \tag{15.6}$$

  which allows us to estimate the fidelity $|\langle\psi|\phi\rangle|$. Thus, the swap test can be used to test two unknown pure states for equality.

The swap test can be readily generalized to qu*d*its.

Remark 15.3. There is a fun application of the swap test known as *quantum fingerprinting*, which we might discuss in class if there is enough time [BCWDW01]: The rough idea goes as follows: We can find 2^n many pure states $|\psi(\boldsymbol{x})\rangle \in \mathbb{C}^{cn}$, indexed by classical bit strings $\boldsymbol{x}$ of length n, with pairwise overlaps

$$|\langle\psi(\boldsymbol{x})|\psi(\boldsymbol{y})\rangle| \leq \frac{1}{2}.$$

Here $c > 0$ is some constant. Thus the quantum states live in a space of only order $\log n$ many qubits! (How can we justify the existence of such vectors? One way is to just choose them at random and estimate probabilities using a more refined version of our calculations

for the symmetric subspace, see [Har13] for more detail.) If we perform k swap tests on $|\psi(\boldsymbol{x})\rangle^{\otimes k} \otimes |\psi(\boldsymbol{y})\rangle^{\otimes k}$ then we obtain

$$\boldsymbol{x} \neq \boldsymbol{y} \quad \Rightarrow \quad \mathbf{Pr}(\text{outcome 1 for all } k \text{ swap tests}) = \left(\frac{3}{4}\right)^k \approx 0$$

Thus the probability of outcome 1 is arbitrarily small, controlled only by the parameter k (but not n). In this sense, we can use the states $|\psi(\boldsymbol{x})\rangle$ as short "fingerprints" for the classical bit strings $\boldsymbol{x}$. The latter are require n bits to specify, while the fingerprints only need order $k \log n$ many qubits (this is not even optimal, but sufficient for our purposes).

Remarkably, while this allows us to test the fingerprints pairwise for equality with high certainty, it is *not* possible to determine the original bitstring $|\boldsymbol{x}\rangle$ from its fingerprint $|\psi(\boldsymbol{x})\rangle$ to good fidelity. This is ensured by the *Holevo bound*, mentioned briefly in Chapter 2, which ensures that we cannot communicate at a rate higher than one classical bit per qubit sent (in the absence of ebits).

15.3 The quantum Schur transform

Now that we have acquired some familiarity with quantum circuitry, we will turn towards solving our actual goal for today – finding a quantum circuit for the Schur transform (15.3),

$$V_{\text{Schur}} : (\mathbb{C}^2)^{\otimes n} \cong \bigoplus_k \text{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^{m(n,k)} \longrightarrow \mathbb{C}^{n+1} \otimes \mathbb{C}^{n+1} \otimes (\mathbb{C}^2)^{\otimes n}$$

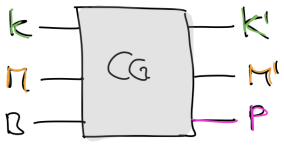
(cf. Fig. 15.1). We will follow the exposition in [Chr10]. A general solution is given in [BCH07, BCH06].

How could we go about finding such a quantum circuit? Remember how we proved Eq. (15.2) in Chapter 12. There we used the Clebsch-Gordan rule (12.9), which asserted that there exists a unitary intertwiner

$$J_k : \text{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^2 \longrightarrow \begin{cases} \bigoplus_{p=\pm 1} \text{Sym}^{k+p}(\mathbb{C}^2) & \text{if } k > 0, \\ \text{Sym}^1(\mathbb{C}^2) = \mathbb{C}^2 & \text{if } k = 0. \end{cases} \tag{15.7}$$

We started with $k = 0$ (zero qubits) and applied the rule in an inductive fashion – after n steps, we managed to decompose the n-qubit Hilbert space into SU(2)-irreps. We can easily lift this procedure from a mere counting scheme to the construction of an actual intertwiner:

(a) Construct a circuit for the Clebsch-Gordan transformation:



This circuit is supposed to implement the following functionality: For every $k \geq 0$, $m \in \{0, 1, \ldots, k\}$, and $b \in \{0, 1\}$,

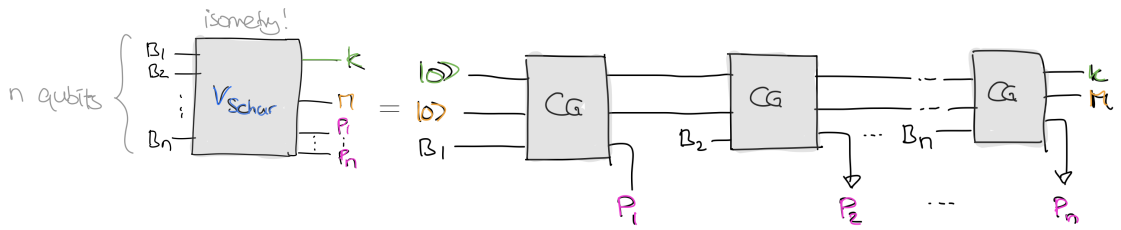
$$|k\rangle_K \otimes |m\rangle_M \otimes |b\rangle_B \mapsto \sum_{p=\pm 1} \sum_{m'} \underbrace{\langle \omega_{m',(k+p)-m'}| \, J_k \, (|\omega_{m,k-m}\rangle \otimes |b\rangle)}_{} \, |k+p\rangle_{K'} \otimes |m'\rangle_{M'} \otimes |p\rangle_P \, . \tag{15.8}$$

For fixed k, the underbraced term is simply an arbitrary matrix element of the Clebsch-Gordan transformation (15.7). Thus, (15.8) applies the Clebsch-Gordan transformation – with k is controlled by the K input and the other two inputs corresponding to $\mathrm{Sym}^k(\mathbb{C}^2)$ and in $\mathbb{C}^2$, respectively. The output subsystem K' contains the label $k' = k \pm p$ of the symmetric subspace that we ended up in, the output M' corresponds to the symmetric subspace $\mathrm{Sym}^{k'}(\mathbb{C}^2)$ itself, and P' contains the path information ($p = \pm 1$).
To obtain a finite transformation, we should restrict the possible values of k that we allow to not exceed some $k_{\max}$. Then the output can be as large as $k_{\max} + 1$, so Eq. (15.8) (partially) defines an isometry, which we will call a *Clebsch-Gordan isometry*

$$\mathrm{CG}\colon\ \mathbb{C}^{k_{\max}+1} \otimes \mathbb{C}^{k_{\max}+1} \otimes \mathbb{C}^2 \longrightarrow \mathbb{C}^{k_{\max}+2} \otimes \mathbb{C}^{k_{\max}+2} \otimes \mathbb{C}^2 \tag{15.9}$$

(On all other basis vectors we can define this isometry in an arbitrary way.) We know from Fig. 12.2 that $k_{\max} := \ell$ is a good choice for the ℓ-th step ($\ell = 0, 1, \ldots, n-1$).

(b) Then the quantum Schur transform can be obtained in the following inductive fashion:



Each Clebsch-Gordan isometry is an isometry between Hilbert spaces of size at most $2n^2$ and we need to apply n such maps to implement the quantum Schur transform. This already implies (using general principles which we have not learned in this course) that the quantum Schur transform can be efficiently implemented!

The Clebsch-Gordan isometry

We will sketch how the Clebsch-Gordan isometries can be implemented in more detail. It is clear that a crucial role is played by the underbraced matrix elements in Eq. (15.8). In the physics literature, these are often called the Clebsch-Gordan *coefficients*.

To understand the situation better, we proceed as in Chapters 6 and 11. If $\mathcal{H}$ is a representation of SU(2) with operators $\{R_U\}$, we previously associated with any operator M on $\mathbb{C}^2$ an operator

$$r_M := -i\partial_{s=0}\left[R_{e^{isM}}\right]$$

on $\mathcal{H}$. We used these operators to analyze representations of SU(2) – in particular, to prove that the symmetric subspaces are irreducible and to establish the Clebsch-Gordan rule! In particular, if $J\colon \mathcal{H} \to \mathcal{H}'$ is an intertwiner then the r_M are likewise intertwined, i.e.,

$$J r_M = r'_M J, \tag{15.10}$$

which in particular implied that J maps eigenvectors of r_Z to eigenvectors of r'_Z with the same eigenvalue. We used this in Chapter 11 to decompose a given representation simply by studying the multiset of eigenvalues of r_Z.

Indeed, recall that for symmetric subspace $\mathcal{H} = \mathrm{Sym}^k(\mathbb{C}^2)$, $R_U = T_U^{(k)}$ is the restriction of $U^{\otimes k}$ and we computed previously that $r_Z = t_Z^{(k)}$ is simply the restriction of $\widetilde{Z} = Z \otimes I \otimes \ldots \otimes I + \cdots + I \otimes \ldots \otimes I \otimes Z$ to the symmetric subspace. The eigenvectors are precisely our favorite basis vectors $|\omega_{m,k-m}\rangle$ for $m = 0, 1, \ldots, k$, with corresponding eigenvalue $m - (k-m) = 2m - k \in$

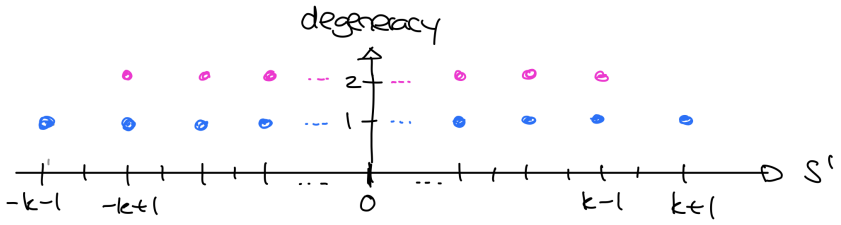


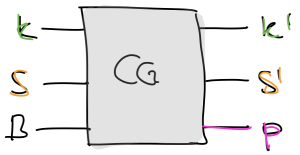
Figure 15.3: Multiplicites of the eigenvalues of r_Z in $\mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^2$. The color coding indicates the decomposition $\mathrm{Sym}^{k+1}(\mathbb{C}^2) \oplus \mathrm{Sym}^{k-1}(\mathbb{C}^2)$.

$\{k, k-2, \ldots, -k\}$ (each nondegenerate). What this means is that we can decompose an arbitrary other representation $\mathcal{H}$ simply by decomposing the multiset of eigenvalues of its corresponding r_Z into sets of the form $\{k', k'-2, \ldots, -k'\}$. In other words, the eigenvalue spectrum of the r_Z operator *uniquely* characterizes the decomposition into irreducible SU(2)-representations!

At this point it will be useful to change notation one last time, since this makes the below arguments much more transparent (and also closer to the literature). Specifically, let us label the basis vectors by the eigenvalue $s = 2m - k$, i.e., define

$$|k; s\rangle := |\omega_{(k+s)/2,(k-s)/2}\rangle \in \mathrm{Sym}^k(\mathbb{C}^2), \quad s \in \{k, k-2, \ldots, -k\},$$

so that $t_Z^{(k)} |k; s\rangle = s\, |k; s\rangle$. In the situation at hand, this means that we would like to think of the Clebsch-Gordan isometry as a quantum circuit of the format



mapping

$$|k\rangle_K \otimes |s\rangle_S \otimes |b\rangle_B \mapsto \sum_{p=\pm 1} \sum_{s'} \langle k+p; s'|\, J_k\big(|k; s\rangle \otimes |b\rangle\big)\ |k+p\rangle_{K'} \otimes |s'\rangle_{S'} \otimes |p\rangle_P\,. \tag{15.11}$$

(This amounts to a simple relabeling $m \mapsto 2m - k$. If you prefer the old labeling, you can conjugating with the controlled unitary $|k\rangle_K \otimes |m\rangle_M \mapsto |k\rangle_K \otimes |2m-k\rangle_S$!)

Now consider the left-hand side and the right-hand side representations that appear in the intertwiner

$$J_k\colon\ \mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^2 \longrightarrow \bigoplus_{p=\pm 1} \mathrm{Sym}^{k+p}(\mathbb{C}^2). \tag{15.12}$$

We shall focus on the interesting case that $k > 0$, since for $k = 0$ we can just use the identity map.

- For $\mathcal{H} = \mathrm{Sym}^k(\mathbb{C}^2) \otimes \mathbb{C}^2$, the group action is $R_U = T_U^{(k)} \otimes U$ and so $r_Z = t_Z^{(k)} \otimes I + I \otimes Z$. This means that the vectors $|k; s\rangle \otimes |b\rangle$ form an eigenbasis, with eigenvalues

$$s + (-1)^b \in \{k+1, k-1, \ldots, -(k+1)\}.$$

  (Note that $|b\rangle \cong |1; (-1)^b\rangle$ if we identify $\mathbb{C}^2 \cong \mathrm{Sym}^1(\mathbb{C}^2)$ and use our new notation.)

- For $\mathcal{H}' = \bigoplus_{p=\pm 1} \mathrm{Sym}^{k+p}(\mathbb{C}^2)$, the action is $R'_U = T_U^{(k+1)} \oplus T_U^{(k-1)}$, so $r'_Z = t_Z^{(k+1)} \oplus t_Z^{(k-1)}$. Hence the vectors $|k'; s'\rangle$ form an eigenbasis, where $k' = k \pm p$, with eigenvalues

$$s' \in \{k', k'-2, \ldots, -k'\} \subseteq \{k+1, k-1, \ldots, -(k+1)\}.$$

Note that, in both cases, the eigenvalues are $\{k+1, k-1, \ldots, -(k+1)\}$ and that each eigenvalue appears twice, except for $\pm(k+1)$, which implies that the representations must be equivalent! See Fig. 15.3 for an illustration. This was precisely argument that we used in Chapter 11 to establish the Clebsch-Gordan rule. Thus, we reproved the fact that there must exist a unitary intertwiner J_k as in Eq. (15.12). Let us now go further and construct such an intertwiner precisely.

Since J_k preserves the eigenspaces, it must necessarily map the eigenvectors of eigenvalue $s' = k+1$ onto each other, up to possibly a phase. Since any scalar multiple of an intertwiner is again an intertwiner, we may in fact assume that

$$J_k \left(|k; k\rangle \otimes |0\rangle \right) = |k+1, k+1\rangle \,. \tag{15.13}$$

For $s' = k-1$, we likewise know that

$$J_k \left(|k; -k\rangle \otimes |1\rangle \right) \propto |k-1; k-1\rangle \,. \tag{15.14}$$

For all other eigenvalues, $s' \in \{k-1, k-3, \ldots, -k+1\}$, the eigenspaces are two-dimensional, so there must exist unitary 2×2-matrices $U(k, s')$ such that

$$J_k \left(|k; s' - (-1)^b\rangle \otimes |b\rangle \right) = \sum_{p=\pm 1} U(k, s')_{p,b} \, |k+p; s'\rangle \tag{15.15}$$

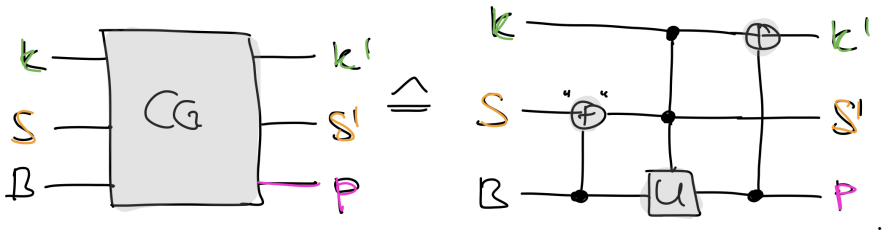
for $b = 0, 1$. Substituting $s = s' - (-1)^b$, we can write this as

$$J_k \left(|k; s\rangle \otimes |b\rangle \right) = \sum_{p=\pm 1} U(k, s + (-1)^b)_{p,b} \, |k+p; s + (-1)^b\rangle \,.$$

We can also bring Eq. (15.13) in this form by defining $U(k, k+1)_{+,0} = 1$, and similarly for Eq. (15.14). Thus, the Clebsch-Gordan isometry (15.11) takes the following simple form:

$$|k\rangle_K \otimes |s\rangle_S \otimes |b\rangle_B \mapsto \sum_{p=\pm 1} U(k, s + (-1)^b)_{p,b} \, |k+p\rangle_{K'} \otimes |s + (-1)^b\rangle_{S'} \otimes |p\rangle_P \,.$$

In other words, the Clebsch-Gordan isometry in essence takes the form of a controlled unitary (with input the B qubit and output the P qubit), controlled by the various inputs! This means that it can be implemented by a circuit of the following form:



The notation on the right-hand side needs some explanation: In the first step, we apply a controlled "addition" that maps $|s\rangle_S \otimes |b\rangle_B$ to $|s + (-1)^b\rangle'_S \otimes |b\rangle_B$. The middle part uses the slightly more general notion of a controlled unitary described in Theorem 15.1, mapping $|k\rangle_K \otimes |s'\rangle_{S'} \otimes |b\rangle_B$ to $|k\rangle_K \otimes |s'\rangle_{S'} \otimes U(k, s') \, |b\rangle_B$. And in the last step we again apply a controlled addition, this time mapping $|k\rangle_K \otimes |p\rangle_P$ to $|k+p\rangle_{K'} \otimes |p\rangle_P$.

Computing the matrix elements

We still need to give a prescription for computing the matrices $U(k, s')$. As mentioned before, $U(k, k+1)_{+,0} = 1$ is the only relevant matrix element for $s' = k+1$, corresponding to Eq. (15.13), which we restate for convenience:

$$J_k \left(|k; k\rangle \otimes |0\rangle \right) = |k+1, k+1\rangle \,. \tag{15.16}$$

To determine the other coefficients, we consider $M_- = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$. If we apply r'_{M_-} to Eq. (15.16) and use Eq. (15.10), we obtain

$$J_k r_{M_-} \left(|k; k\rangle \otimes |0\rangle \right) = r'_{M_-} J_k \left(|k; k\rangle \otimes |0\rangle \right) = r'_{M_-} \, |k+1, k+1\rangle \,.$$

Recall that $r_{M_-} = t^{(k)}_{M_-} \otimes I + I \otimes M_-$ and $r'_{M_-} = t^{(k+1)}_{M_-} \oplus t^{(k-1)}_{M_-}$. Since $t^{(k)} \, |k, s\rangle \propto |k, s-2\rangle$ etc. (Eq. (6.5)), it follows that

$$J_k \left(\alpha \, |k; k-2\rangle \otimes |0\rangle + \beta \, |k; k\rangle \otimes |1\rangle \right) = |k+1, k-1\rangle \tag{15.17}$$

for certain coefficients α and β that we can calculate explicitly. By unitarity, $|\alpha|^2 + |\beta|^2 = 1$. But we know from above that J_k preserves the two-dimensional eigenspace corresponding to $s' = k-1$ (Eq. (15.15)), so it follows that

$$J_k \left(\gamma \, |k; k-2\rangle \otimes |0\rangle + \delta \, |k; k\rangle \otimes |1\rangle \right) = |k-1, k-1\rangle \tag{15.18}$$

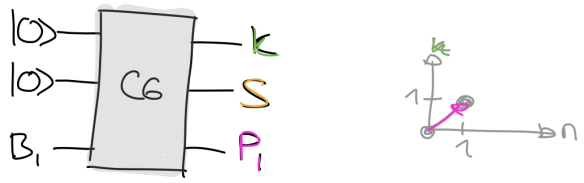
for some coefficients γ and δ. By unitarity, $|\gamma|^2 + |\delta|^2 = 1$ and $\gamma\bar{\alpha} + \delta\bar{\beta} = 0$, which determines these coefficients up to phase. Any choice of phase will lead to a valid intertwiner, since this is exactly the freedom that we have from Theorem 13.2. If we define $U(k, k-1) := \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)^{-1}$, then Eq. (15.15) is satisfied for $s' = k-1$.

We can now simply keep applying r'_{M_-} to Eqs. (15.17) and (15.18) to obtain the matrices $U(k, s')$ for all other values of s'.

Examples

At last, let us discuss some concrete examples to make sure that we fully understand what is going on:

Example 15.4 (n=1)**.** For a single qubit, the quantum Schur transform is completely trivial:

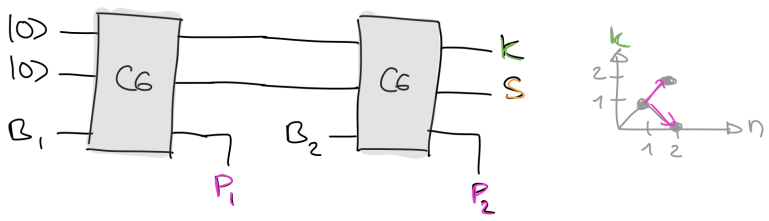


It maps

$$|0\rangle_{B_1} \mapsto |1\rangle_K \otimes |1\rangle_S \otimes |+\rangle_{P_1}$$
$$|1\rangle_{B_1} \mapsto |1\rangle_K \otimes |-1\rangle_S \otimes |+\rangle_{P_1}$$

Note that the K-system is always in state $|1\rangle_K$ and the P_1-system always in state $|+\rangle_{P_1}$, corresponding to the unique $(0,0) \to (1,1)$.

Example 15.5 (n=2)**.** For two qubits, the quantum Schur transform



maps

$$|00\rangle_B \mapsto |2\rangle_K \otimes |2\rangle_S \otimes |{+}{+}\rangle_P$$
$$|11\rangle_B \mapsto |2\rangle_K \otimes |{-}2\rangle_S \otimes |{+}{+}\rangle_P\,,$$

while

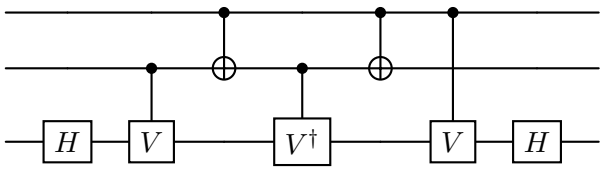
$$|01\rangle_B = \frac{1}{\sqrt{2}}\underbrace{\frac{|01\rangle+|10\rangle}{\sqrt{2}}}_{\in \mathrm{Sym}^2(\mathbb{C}^2)} + \frac{1}{\sqrt{2}}\underbrace{\frac{|01\rangle-|10\rangle}{\sqrt{2}}}_{\in \mathbb{C}|\Psi^-\rangle} \mapsto \frac{1}{\sqrt{2}}|2\rangle_K \otimes |0\rangle_S \otimes |{+}{+}\rangle_P + \frac{1}{\sqrt{2}}|0\rangle_K \otimes |0\rangle_S \otimes |{+}{-}\rangle_P\,,$$
$$|10\rangle_B = \frac{1}{\sqrt{2}}\underbrace{\frac{|01\rangle+|10\rangle}{\sqrt{2}}}_{\in \mathrm{Sym}^2(\mathbb{C}^2)} - \frac{1}{\sqrt{2}}\underbrace{\frac{|01\rangle-|10\rangle}{\sqrt{2}}}_{\in \mathbb{C}|\Psi^-\rangle} \mapsto \frac{1}{\sqrt{2}}|2\rangle_K \otimes |0\rangle_S \otimes |{+}{+}\rangle_P - \frac{1}{\sqrt{2}}|0\rangle_K \otimes |0\rangle_S \otimes |{+}{-}\rangle_P\,.$$

It is instructive to verify this explicitly by following the algorithm outlined above.

Exercises

15.1 **Schur transform for $n = 3$:** Can you write down the Schur transform (concretely) for $n = 3$? Compare the result with your solution to Exercise 13.2.

15.2 **Toffoli gate:** Verify that the following three-qubit circuit computes the Toffoli (CCNOT) gate:



Here, $V = \left(\begin{smallmatrix} 1 & \\ & i \end{smallmatrix}\right)$ is a square root of the Z-gate.

Chapter 16

Quantum entropy and mutual information

Today we will study the von Neumann entropy more generally and discuss its mathematical properties. We will also introduce a new correlation measure – the mutual information. Finally, we will introduce a quantum information processing task called (coherent) quantum state merging. This is a very general task that encompasses several others that we previously studied in this course, and we will explain how to solve it tomorrow.

16.1 Shannon and von Neumann Entropy

Let us first revisit the classical case. For a probability distribution $\{p, 1-p\}$ with two outcomes, we previously defined the binary Shannon entropy as $h(p) = -p \log p - (1-p) \log(1-p)$ (Chapter 9). We will now define the *Shannon entropy* of general probability distribution $\{p_i\}_{i=1}^d$ with d many outcomes by

$$H(\{p_i\}_{i=1}^d) := -\sum_{i=1}^d p_i \log p_i.$$

As before, we set $0 \log 0 := 0$. It is clear that $H(\{p, 1-p\}) = h(p)$, so this is a proper generalization. Everything that we discussed in Chapter 9 generalizes to probability distributions with d outcomes. Note that

$$0 \leq H(\{p_i\}) \leq \log d. \tag{16.1}$$

The lower bound is attained for deterministic distributions and the upper bound for a uniform distribution. How to see this? For the lower bound, note that $p_i \log p_i \geq 0$ for every $p_i \in [0, 1]$, with equality if and only if each $p_i \in \{0, 1\}$. For the upper bound we use Jensen's inequality for the concave log function, which shows that $\sum_{i=1}^d p_i \log \frac{1}{p_i} \leq \log(\sum_{i=1}^d p_i \frac{1}{p_i}) = \log d$. Since the logarithm is strictly concave, we have equality if and only if all the $1/p_i$ are equal.

Now consider a density operator ρ on $\mathbb{C}^d$. We define its *von Neumann entropy* by

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

Clearly, $S(\rho) = H(\{p_i\})$ for $\{p_i\}_{i=1}^d$ the eigenvalues of ρ (repeated according to their multiplicity). This generalizes the definition given previously in Chapter 10 for qubits. Note that

$$0 \leq S(\rho) \leq \log d,$$

The lower bound is attained precisely for pure states and the upper bound if and only if ρ is a maximally mixed state, i.e., $\rho = I/d$. This follows directly from the discussion below Eq. (16.1).

The von Neumann entropy is the optimal asymptotic rate for compression and quantum state transfer (Chapters 9 and 10). The basic reason is that the following *asymptotic equipartition property (AEP)*: For every $\varepsilon > 0$ there exist typical projectors P_n on $(\mathbb{C}^d)^{\otimes n}$, $n = 1, 2, \ldots$, such that

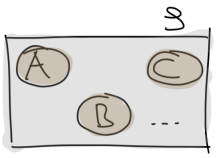
(a) $\mathrm{tr}[P_n \rho^{\otimes n}] \to 1$ (typicality),
(b) $\mathrm{rk}[P_n] \leq 2^{n(S(\rho)+\varepsilon)}$, and
(c) the eigenvalues of $P_n \rho^{\otimes n} P_n$ are within $2^{-n(S(\rho)\pm\varepsilon)}$.

For qubits, we proved the first two property in class. In fact, we gave two constructions – one using the eigendecomposition in Chapter 10 and a universal one using Schur-Weyl duality in Chapter 14). The third property is also useful as we will see tomorrow. For construction in Chapter 10, it follows readily using the continuity of the binary entropy function, and for the other you can proceed as in the derivation of Eq. (14.14).

The first property implies that $\rho^{\otimes n} \approx P_n \rho^{\otimes n} P_n$ for large n (this follows directly from the gentle measurement lemma, Exercise 7.8). The second and then third property show that $P_n \rho^{\otimes n} P_N$ in turn looks – roughly speaking – like a uniform probability distribution on a space of approximately $nS(\rho)$ qubits. This explains the term "asymptotic equipartition property".

16.2 Entropies of subsystems and mutual information

Supose that $\rho_{ABC\ldots}$ is a density operator on a tensor product Hilbert space. We can then not only compute the entropy of the overall state but also the reduced density operators such as ρ_A describing the subsystems, as visualized below.



In order to emphasize the subsystem, let us define the following useful notation:

$$S(A)_\rho := S(\rho_A)$$

We will often omit the subscript ρ and write $S(A)$ when the state is clear from the context. Let us discuss some examples for a density operator on a bipartite system:

- If ρ_{AB} is pure then

$$S(AB) = 0, \quad S(A) = S(B). \tag{16.2}$$

  Note that $S(A) = S(B)$ is nothing but $S_E(\rho)$, the entanglement entropy of the pure state.

- If $\rho_{AB} = \rho_A \otimes \rho_B$ is a tensor product of two density operators, then $S(AB) = S(A) + S(B)$. Indeed, if $\{p_i\}$ and $\{q_j\}$ are the eigenvalues of ρ_A and ρ_B, respectively, then $\{p_i q_j\}$ are the eigenvalues of ρ_{AB} and so

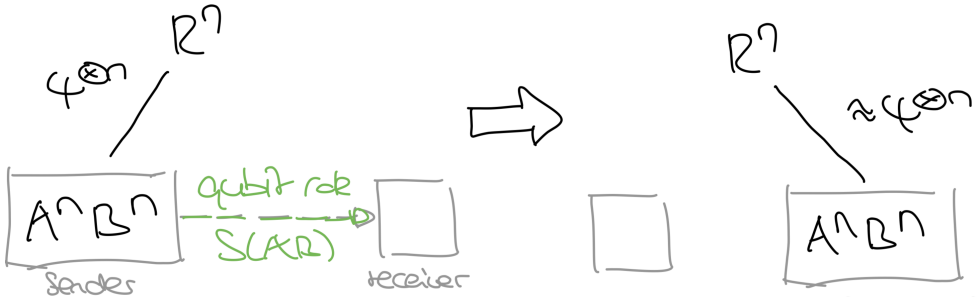
$$\begin{aligned} S(AB) &= -\sum_{i,j} p_i q_j \log(p_i q_j) = -\sum_{i,j} p_i q_j \log p_i - \sum_{i,j} p_i q_j \log q_j \\ &= -\sum_i p_i \log p_i - \sum_j q_j \log q_j = S(A) + S(B). \end{aligned}$$

The second example shows that the von Neumann entropy is additive under tensor products (we can also write it as $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$ to emphasize this aspect).

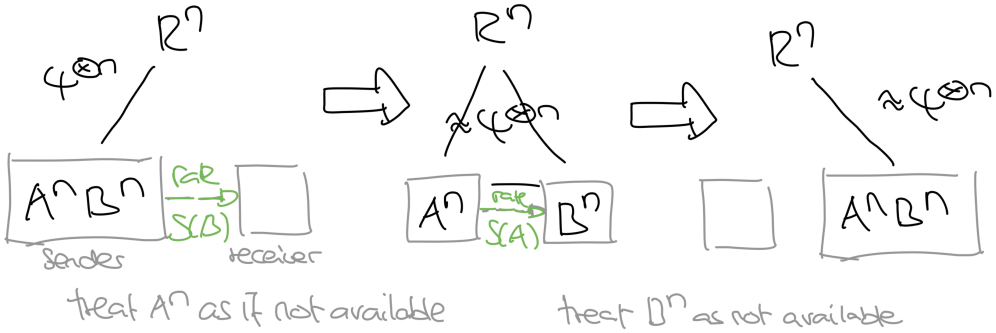
When ρ_{AB} is a general density operator, it is still true that the entropy is *subadditive*:

$$S(AB) \leq S(A) + S(B) \tag{16.3}$$

MW: Give a short recent proof. This is very important result follows, e.g., from a result called Klein's inequality (see [NC10] for all details). In class, we instead gave a plausibility argument based on the operational interpretation of the von Neumann entropy as the *optimal* rate for the quantum state transfer task. Indeed, consider $|\psi\rangle_{ABR}^{\otimes n}$, where $|\psi\rangle_{ABR}$ is a purification of ρ_{AB}. On the one hand, we know that Alice can (approximately) transfer her AB-systems to Bob at (a rate arbitrarily close to) the optimal rate $S(AB)$:



On the other hand, she can certainly first send the B-systems and then the A-systems, at a rate $S(A) + S(B)$.



By optimality of the former, it follows that $S(AB) \leq S(A) + S(B)$. This argument would be a completely rigorous mathematical proof – except that we did not quite prove optimality! (Can you see why Exercise 10.1 is not quite enough?)

Equation (16.3) is an example of an entropy inequality. Another example is *Araki-Lieb inequality*:

$$|S(A) - S(B)| \leq S(AB). \tag{16.4}$$

We can prove it by a convenient trick that allows us to produce new entropy inequalities from old ones. Choose a purification $|\psi\rangle_{ABR}$ of ρ_{AB}. Then, using that the entropies of complementary subsystems are the same (Eq. (16.2)),

$$aS(A) - S(B) = S(BR) - S(B) \leq S(R) = S(AB),$$

and similarly for $S(B) - S(A)$.

Remark 16.1. There is also a *strong* subadditivity inequality which asserts that $S(AC) + S(BC) \leq S(ABC) + S(C)$. It is not so easy to prove but enormously useful in quantum information theory.

Mutual Information

The preceding suggests that the *mutual information*, defined for any density operator ρ_{AB} on $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ by

$$I(A:B)_\rho := S(A)_\rho + S(B)_\rho - S(AB)_\rho,$$

might be an interesting property to consider. The state transfer argument given above indicates that this quantity to be related to the information that we lose by treating A and B as independent. Let us discuss some of its mathematical properties:

- $I(A:B) \geq 0$ by the subadditivity inequality 16.3. One can show (but we will not) that $I(A:B) = 0$ if and only if $\rho_{AB} = \rho_A \otimes \rho_B$.
- If ρ_{AB} is pure then $I(A:B) = 2S(A) = 2S(B)$.
- More generally, $I(A:B) \leq 2\min\{S(A), S(B)\} \leq 2\min\{\log d_A, \log d_B\}$. The former is a consequence of the Araki-Lieb inequality 16.4.
- For separable states, $I(A:B) \leq \min\{S(A), S(B)\}$. It follows that if $I(A:B) > S(A)$ or $S(B)$ then the state ρ_{AB} is necessarily entangled!

For an example of the latter, contrast:

- For $|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, we have $I(A:B) = 1 + 1 - 0 = 2$.
- For $\rho_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)$, we have $I(A:B) = 1 + 1 - 1 = 1$.

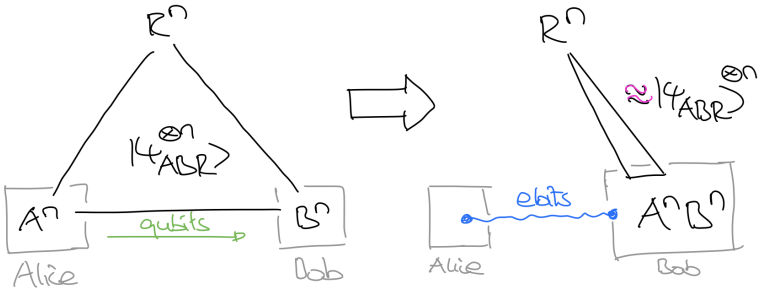
Tomorrow, we will prove that in this case we can even extract ebits at a positive rate given many copies of the state ρ_{AB}.

Remark 16.2. There exist further measures than the ones we have discussed here. For example, the binary relative entropy, which we so far only defined for classical probability distributions with two outcomes each, can be defined for general probability distributions and even for quantum states, by $S(\rho\|\sigma) := \mathrm{tr}[\rho \log \rho] - \mathrm{tr}[\rho \log \sigma]$.

Moreover, there are other linear combinations of the von Neumann entropy that are meaningful. For example, the conditional entropy $S(A|B) = S(AB) - S(B)$ and its negative, the coherent information $S(A > B) := S(B) - S(AB)$. We will see the meaning of the latter tomorrow.

16.3 A glance at quantum state merging

We will close today's lecture with a review of tomorrow's topic – a task called *(coherent) quantum state merging*. Here, we imagine that Alice, Bob, and an unspecified reference system share n copies of a pure state $|\psi\rangle_{ABR}$. Alice's and Bob's goal is transfer the A systems from Alice to Bob by sending as few qubits as possible, as illustrated in the below figure:



Note that we already know how to solve this problem by sending $S(A)$ qubits – simply use our usual state transfer protocol (as we did above when discussing subadditivity). However, this ignores that Bob already has part of the quantum state. Thus, this strategy will in general not be optimal (unless there is no B system, in which we are back in the state transfer scenario).

How about if there is not R system? In this case, Alice and Bob share many copies of a pure state $|\psi\rangle_{AB}$. Here, no quantum communication is required at all, since Bob can simply re-create the state in his laboratory. Instead, Alice and Bob can use $|\psi\rangle_{AB}$ "for free" for other purposes, such as for distilling perfect ebits $|\Phi^+\rangle$ at some rate (as indicated in the figure).

Tomorrow we will see that this is indeed possible and prove the following result: There exists a quantum protocol (sometimes called the *mother protocol* or the *fully quantum Slepian-Wolf protocol*) that, given $|\psi\rangle_{ABR}^{\otimes n}$,

- achieves the state merging task by sending qubit at an asymptotic rate $\frac{1}{2}I(A:R)$,
- distills ebits at an asymptotic rate $\frac{1}{2}I(A:B)$.

Since $\frac{1}{2}I(A:R) \leq S(A)$, this indeed improves over the qubit rate over the naive protocol. But it will also teach us how to distill ebits (even when ρ_{AB} is mixed), which is something that we only alluded to briefly in Section 10.3!

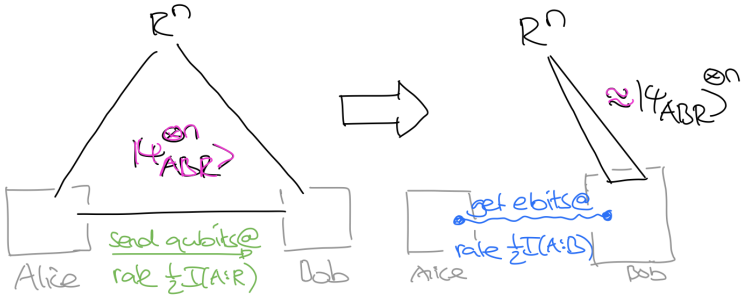
Chapter 17

Quantum state merging via the decoupling approach

Today we will study the (coherent) quantum state merging task in more detail and discuss its many applications. We will discuss a protocol based on the *decoupling approach*, which is a beautiful technique for solving quantum communication tasks. We will close with an outlook on some of the topics that we did not manage to cover in this course.

17.1 Quantum state merging

In yesterday's Chapter 16, we discussed the (coherent) quantum state merging task: Here, Alice, Bob, and an unspecified reference system share n copies of a pure state $|\psi\rangle_{ABR}$. They would like to transfer the A systems from Alice to Bob by sending as few qubits as possible and, in addition, obtain as many ebits as possible. The situation is illustrated in the following figure, which also already states the main result:



That is, we will see that it suffices to send qubits at an asymptotic rate arbitrarily close to $\frac{1}{2}I(A:R)$ and that we will obtain ebits at an asymptotic rate arbitrarily close to $\frac{1}{2}I(A:B)$.

For comparison, naively applying the quantum state transfer protocol from 10 requires a qubit rate of $S(A) \geq \frac{1}{2}I(A:R)$ and yields no ebits at all!

Remark 17.1. There are other possible variants that can be analyzed similarly. In *quantum state splitting*, the "dual" scenario, we imagine that Bob starts out with the AB systems and he wants to send the A systems over to Alice, while holding on to the B systems. *Quantum state redistribution* is the generalization of both scenarios, where we start with many copies of a four-party state $|\psi\rangle_{ABCR}$; initially, the AC systems belong to Alice, Bob has the B systems, and after the termination of the protocol we would like for Alice to keep A while Bob is in possession of BC.

Special cases and applications

- If there is no B system (which you can formally model by taking $\mathcal{H}_B = \mathbb{C}$) then everything reduces to quantum state transfer. Indeed, $\frac{1}{2}I(A:R) = S(A)$ and $\frac{1}{2}I(A:B) = 0$.

- *Entanglement distillation:* Suppose that Alice and Bob share many copies of a quantum state ρ_{AB} and that they would like to obtain as many ebits as possible *by sending (classical) bits* only. This task is known as entanglement distillation (cf. Section 10.3, where we discussed this briefly). Note that here we do not seem to care about the R systems at all. Yet, the quantum state merging protocol can be usefully applied (simply choose any purification $|\psi\rangle_{ABR}$)! Simply use teleportation (Chapter 2) to replace the quantum communication (at rate $\frac{1}{2}I(A:R)$) by classical communication (at rate $I(A:R)$) and consuming ebits (at rate $\frac{1}{2}I(A:R)$). In this way, we can distill ebits at a *net rate*

$$\begin{aligned}\frac{1}{2}I(A:B) - \frac{1}{2}I(A:R) &= \frac{1}{2}\big(S(A) + S(B) - S(AB) - S(A) - S(AB) + S(B)\big) \\ &= S(B) - S(AB)\end{aligned}$$

  by sending bits at rate $I(A:R)$. The right-hand side quantity is called the coherent information and often denoted by $I(A>B)$. It can have either sign – but if it is positive then this procedure allows us to distill entanglement at a positive rate!

  For example, if there is no R system then ρ_{AB} is pure and so $S(B) - S(AB) = S(B)$, which means that we can distill ebits at rate $S(A) = S(B)$! This was a result that we had announced in Chapter 2.

- *Noisy teleportation:* Once we have obtained ebits using the entanglement distillation procedure sketched above, we can use it as a resource for other tasks, such as teleportation. This means that using "noisy" density operators ρ_{AB} we can teleport qubits at rate $S(B) - S(AB)$ (provided this rate is nonnegative) by sending bits at rate

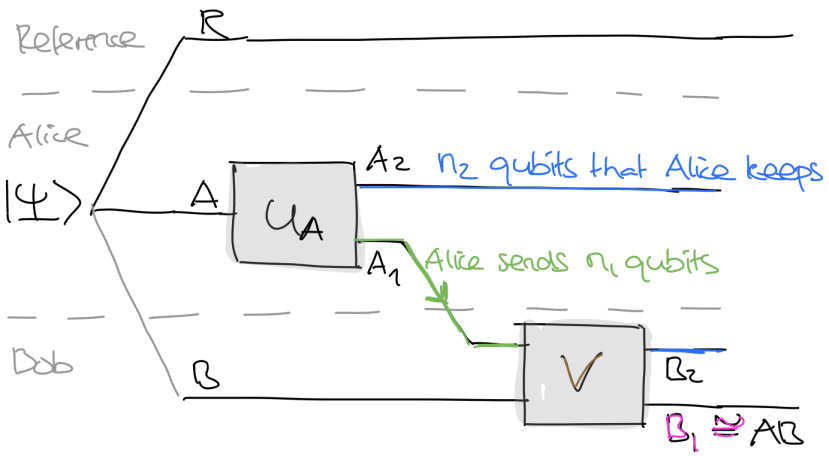
$$I(A:R) + 2\left(S(B) - S(AB)\right) = I(A:B).$$

- *Noisy superdense coding:* Similarly, we can do superdense coding by using general density operators ρ_{AB}. Here we take the quantum state merging protocol and do ordinary superdense coding with the ebits obtained. This allows us to communicate classical bits at the "superdense rate" $I(A:B)$ by sending qubits at rate $\frac{1}{2}I(A:R) + \frac{1}{2}I(A:B) = S(A)$. Note that this is only interesting if $I(A:B) > S(A)$ (or $S(B) > S(AB)$), which is precisely the threshold which implied that ρ_{AB} had to be entangled.

For $\rho_{AB} = |\Phi^+\rangle$, the above reduce to ordinary teleportation and superdense coding, respectively.

17.2 The decoupling approach

How should we go about solving the state merging problem? Here is a natural template for what such a protocol could look like:



Here we assume that the initial state is some arbitrary state $|\Psi\rangle_{ABR}$ (not necessary a tensor power state $|\psi\rangle^{\otimes n}$)! First, Alice applies a unitary U_A. Next, she considers her Hilbert space as a tensor product $\mathcal{H}_A = \mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2}$, with n_1 qubits in the first and n_2 qubits in the second tensor factor, and sends over n_1 of the qubits to Bob. Lastly, Bob applies an isometry $V_{A_1B\to B_1B_2}$, where $\mathcal{H}_{B_1} \cong \mathcal{H}_A \otimes \mathcal{H}_B$ and $\mathcal{H}_{B_2} \cong \mathcal{H}_{A_2}$. This protocol would be successful if it leads to a state that is close to

$$\underbrace{|\Phi^+\rangle^{\otimes n_2}}_{\text{on } A_2B_2} \otimes \underbrace{|\Psi\rangle_{ABR}}_{\text{on } B_1R}. \tag{17.1}$$

Hopefully we can achieve this by choosing n_1 not too large (and hence n_2 not too small). How should we define the objects in the protocol so that this procedure is successful?

The crucial observation is that we can analyze the situation purely by considering the state

$$|\Gamma\rangle_{ABR} := (U_A \otimes I_{BR})\,|\Psi\rangle_{ABR}\,.$$

Indeed, if the state at the end of the protocol is close to the desired state Eq. (17.1) then this implies that

$$\Gamma_{A_2R} \approx \frac{I_{A_2}}{2^{n_2}} \otimes \Psi_R. \tag{17.2}$$

Indeed, note that the isometry acts only on A_1B and hence does not change the state of the A_2R systems, so we can simply trace out B_1B_2 in Eq. (17.1). In fact, Eq. (17.2) is not only necessary, but also *sufficient* in the following sense: Since $|\Gamma\rangle_{ABR}$ is a purification of Γ_{A_2R} and Eq. (17.1) is a purification of $\frac{I_{A_2}}{2^{n_2}} \otimes \Psi_R$, Eq. (17.2) implies that there must exist an isometry $V_{A_1B\to B_1B_2}$ that maps one purification to another. If Eq. (17.2) held with equality then this would be precisely what you proved in Exercise 7.5! In the approximate case, you can use the fidelity from Section 14.1 to prove this assertion – can you fill in the details?

The upshot of the preceding discussion is the following: Remarkably, we do not need to cleverly construct the isometry V at all – we rather get it for free provided that we manage to find a unitary U_A such that the system A_2 that remain with Alice *decouple* from the reference system R in the sense of Eq. (17.2). This is the essence of the *decoupling argument*.

How can we obtain the unitary U_A? The following theorem shows that, on average, a randomly chosen unitary does a good job provided that we choose A_2 not too large.

Theorem 17.2 (Decoupling theorem)**.** *Let Ψ_{AR} be a positive semidefinite operator on $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_R}$, where $d_A = d_{A_1}d_{A_2}$. Then:*

$$\int dU_A \left\| \mathrm{tr}_{A_1}\left[(U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R)\right] - \frac{I_{A_2}}{d_{A_2}} \otimes \Psi_R \right\|_1^2 \leq \frac{d_A d_R}{d_{A_1}^2}\,\mathrm{tr}\left[\Psi_{AR}^2\right].$$

Asymptotics

We will prove Theorem 17.2 momentarily, but let us first see why it allows us to solve the quantum state merging problem. For this, we will use the asymptotic equipartition property (see Chapter 16)! Let $|\psi\rangle_{ABR}$ denote an arbitrary pure state and $P_{A,n}$, $P_{B,n}$, $P_{R,n}$ typical projectors for ψ_A, ψ_B, ψ_R and some fixed $\varepsilon > 0$, respectively, and define

$$|\Psi\rangle_{A^n B^n R^n} := (P_{A,n} \otimes P_{B,n} \otimes P_{R,n}) \, |\psi\rangle_{ABR}^{\otimes n} .$$

Then, by typicality and the gentle measurement lemma (applied three times),

$$|\Psi\rangle_{A^n B^n R^n} \approx |\psi\rangle_{ABR}^{\otimes n} ,$$

so we may safely construct a protocol for the state $|\Psi\rangle$ instead of for $|\psi\rangle^{\otimes n}$.

We will follow the decoupling approach. Let us regard $|\Psi\rangle$ as a vector in $\mathbb{C}^{d_{A'}} \otimes \mathbb{C}^{d_{B'}} \otimes \mathbb{C}^{d_{R'}}$, where $d_{A'}$, $d_{B'}$, $d_{C'}$ denote the ranks of those projectors. We will correspondingly write $|\Psi\rangle_{A'B'R'}$. Then, by the asymptotic equipartition property,

$$\begin{aligned} &d_{A'} \leq 2^{n(S(A)+\varepsilon)}, \\ &d_{R'} \leq 2^{n(S(R)+\varepsilon)}, \\ &\operatorname{tr}\left[\Psi_{A'R'}^2\right] = \operatorname{tr}\left[\Psi_{B'}^2\right] \leq 2^{n(S(B)+\varepsilon)} 2^{-2n(S(B)-\varepsilon)} = 2^{n(-S(B)+3\varepsilon)} = 2^{n(-S(AR)+3\varepsilon)}. \end{aligned}$$

(The last inequality requires some thought!) Together,

$$d_{A'} d_{R'} \operatorname{tr}\left[\Psi_{A'R'}^2\right] \leq 2^{n(I(A:R)+5\varepsilon)}.$$

Thus, Theorem 17.2 ensures the existence of a decoupling unitary U_A provided that we choose

$$d_{A'_1} \gg 2^{n(\frac{1}{2} I(A:R)+\frac{5}{2}\varepsilon)}$$

and n large enough. In other words, we need to send over qubits at a rate arbitrarily close to $\frac{1}{2} I(A:R)$. This is exactly the desired asymptotic qubit rate!

As a consequence, it is also true that we will obtain ebits at a rate arbirarily close to $\frac{1}{2} I(A:B)$> Indeed, we have $\frac{1}{2} I(A:R) + \frac{1}{2} I(A:B) = S(A)$, $d_{A'_1} d_{A'_2} = d_{A'}$, and you proved in Exercise 10.1 that any typical subspace for ψ_A has to grow faster than $2^{n(S(A)-\delta)}$ for any $\delta > 0$.

17.3 Proof of the decoupling theorem

In order to prove Theorem 17.2, we first need to understand how to compute averages with respect to the Haar measure.

Haar averages

First, suppose that M is an arbitrary operator on $\mathbb{C}^d$. Then:

$$\int dU \; UMU^\dagger = \frac{\operatorname{tr}[M]}{d} I. \tag{17.3}$$

Indeed, $\mathbb{C}^d$ is an irreducible representation of $U(d)$ and the invariance property (14.4) of the Haar measure guarantees that the left-hand side of the equation is an intertwiner; thus, Schur's lemma implies that it is proportional to the identity operator. Since the traces agree, Eq. (17.3) follows.

Now consider an arbitrary operator M on $\mathbb{C}^d \otimes \mathbb{C}^d$. Here one can similarly show that

$$\int dU \; (U \otimes U) M (U \otimes U)^\dagger = \begin{cases} \gamma \Pi_2 + \delta(I - \Pi_2), \\ \alpha I + \beta F, \end{cases} \tag{17.4}$$

where F denotes the swap operator and α, β, γ, δ are suitable constants that depend linearly on M. Why is this true? Let us first observe that, since $\Pi_2 = \frac{1}{2}(I + F)$, we necessarily have that $\alpha = (\gamma + \delta)/2$, $\beta = (\gamma - \delta)/2$, so it suffices to prove either expression. We will still give a justification for each expression individually. Since the left-hand side operator

- As a representation of $U(d)$, $\mathbb{C}^d \otimes \mathbb{C}^d$ decomposes into the symmetric and the anti-symmetric subspace, which are both irreducible. (The proof that the latter is irreducible is very similar to the proof for the former, see Chapter 6.) By Schur's lemma, it follows that any operator that commutes with every $U^{\otimes 2}$ can necessarily be written as a linear combination of Π_2 and $I - \Pi_2$. See Exercise 7.4 where you proved a very closely related statement in the case of qubits ($d = 2$)!
- On the other hand, one can prove directly that any operator that commutes with every $U^{\otimes n}$ can necessarily be written as a linear combination of the permutation operators $\{R_\pi\}_{\pi \in S_n}$ – see Theorem 13.9. The above is the special case $n = 2$ of this general result.

We still need to determine the coefficients. Since there are two coefficients, two equations suffice to determine both. For example, we can compare the trace of the left and the right-hand side operators, as well as the trace after multiplying the equation by F (which amounts to replacing M by FM and interchanging α and β). Using that $\mathrm{tr}[I] = d^2$ and $\mathrm{tr}[F] = d$, this leads to

$$\begin{aligned} \alpha &= \frac{d}{d^3 - d} \mathrm{tr}[M] - \frac{1}{d^3 - d} \mathrm{tr}[FM] \\ \beta &= \frac{d}{d^3 - d} \mathrm{tr}[FM] - \frac{1}{d^3 - d} \mathrm{tr}[M]. \end{aligned} \tag{17.5}$$

Sanity check

Let us first compute the average of the operator $\mathrm{tr}_{A_1}[(U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R)]$ to get some intuition why Theorem 17.2 should be true. Using Eq. (17.3), it is not hard to see that

$$\int dU_A \; \mathrm{tr}_{A_1} \left[(U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R) \right] = \mathrm{tr}_{A_1} \left[\frac{I_A}{d_A} \otimes \Psi_R \right] = \frac{I_{A_2}}{d_{A_2}} \otimes \Psi_R \tag{17.6}$$

which is exactly the decoupled operator that we would like to obtain. (In case this calculation is not clear: This follows simply by applying Eq. (17.3) to each "block" obtained by applying $\langle r|_R$ on the left and $|r'\rangle_R$ on the right.)

Note tracing out the A_1 system was not important at all. However, this is only an average statement – if we would like to show that there exist single unitaries U_A that decouple then we need to control the fluctuations! The content of Theorem 17.2 is that the fluctuations are indeed arbitrarily small provided we choose A_1 to be sufficiently large.

Proof of the theorem

We will now prove the decoupling theorem. First, it will be useful to introduce a new norm – the *Frobenius norm* (or *Hilbert-Schmidt norm*) of an operator M, which is often denoted by

$$\|M\|_2 := \sqrt{\mathrm{tr}\,[M^\dagger M]}. \tag{17.7}$$

Note that $\|M\|_2$ is nothing but the ℓ^2-norm of the singular values of M. Thus it can be related to the trace norm in the following way:

$$\|M\|_2 \leq \|M\|_1 \leq \sqrt{\mathrm{rk}(M)}\|M\|_2$$

(the second inequality is the Cauchy-Schwarz inequality). Let's start calculating using the Frobenius norm:

$$\begin{aligned}
&\int dU_A \, \|\mathrm{tr}_{A_1}\left[(U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R)\right] - \frac{I_{A_2}}{d_{A_2}} \otimes \Psi_R\|_2^2 \\
&= \int dU_A \, \mathrm{tr}\left[\left(\mathrm{tr}_{A_1}\left[(U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R)\right] - \frac{I_{A_2}}{d_{A_2}} \otimes \Psi_R\right)^2\right] \\
&= \int dU_A \, \mathrm{tr}\left[\mathrm{tr}_{A_1}^2\left[(U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R)\right]\right] - \frac{1}{d_{A_2}}\mathrm{tr}\left[\Psi_R^2\right], \qquad (17.8)
\end{aligned}$$

where the second equality follows from Eq. (17.6). Note that only the first term depends on the unitary U_A! We can compute its average by using the swap trick – this is the main advantage of using the Frobenius norm:

$$\begin{aligned}
&\int dU_A \, \mathrm{tr}\left[\mathrm{tr}_{A_1}^2\left[(U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R)\right]\right] \\
&= \int dU_A \, \mathrm{tr}\left[\left((U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R)\right)^{\otimes 2}\left(I_{A_1A_1'} \otimes F_{A_2A_2'} \otimes F_{RR'}\right)\right] \\
&= \mathrm{tr}\Big[\Psi_{AR}^{\otimes 2}\big(\underbrace{\int dU_A \, U_A^{\dagger,\otimes 2}\left(I_{A_1A_1'} \otimes F_{A_2A_2'}\right)U_A^{\otimes 2}}_{\alpha I_{AA'} + \beta F_{AA'}} \otimes F_{RR'}\big)\Big] \\
&= \alpha \, \mathrm{tr}\left[\Psi_R^2\right] + \beta \, \mathrm{tr}\left[\Psi_{AR}^2\right].
\end{aligned}$$

In the underbraced expressen we used Eq. (17.4). The coefficients can be calculated using Eq. (17.5):

$$\begin{aligned}
\alpha &= \frac{d_A}{d_A^3 - d_A} d_{A_1}^2 d_{A_2} - \frac{1}{d_A^3 - d_A} d_{A_1} d_{A_2}^2 = \frac{d_A d_{A_1} - d_{A_2}}{d_A^2 - 1} \leq \frac{1}{d_{A_2}} \\
\beta &= \text{ roles of } A_1 \text{ and } A_2 \text{ reversed } = \frac{d_A d_{A_2} - d_{A_1}}{d_A^2 - 1} \leq \frac{1}{d_{A_1}}.
\end{aligned}$$

If we plug this back into Eq. (17.8) and take the average, we see that the α term cancels! Thus we obtain

$$\int dU_A \, \|\mathrm{tr}_{A_1}\left[(U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R)\right] - \frac{I_{A_2}}{d_{A_2}} \otimes \Psi_R\|_2^2 \leq \frac{1}{d_{A_1}}\mathrm{tr}\left[\Psi_{AR}^2\right].$$

Finally, we use the upper bound on the trace norm in terms of the Frobenius norm in Eq. (17.7):

$$\int dU_A \, \|\mathrm{tr}_{A_1}\left[(U_A \otimes I_R)\Psi_{AR}(U_A^\dagger \otimes I_R)\right] - \frac{I_{A_2}}{d_{A_2}} \otimes \Psi_R\|_1^2 \leq \frac{d_{A_2}d_R}{d_{A_1}}\mathrm{tr}\left[\Psi_{AR}^2\right] = \frac{d_A d_R}{d_{A_1}^2}\mathrm{tr}\left[\Psi_{AR}^2\right].$$

This is the desired result. □

17.4 Outlook

Now that we have reached the end of this course, we will close with a brief discussion of two important topics that we did not have time to cover this term:

- *Converses:* Over the past weeks, we constructed many useful information processing protocols, but only rarely proved optimality. To do so in a systematic way requires extending the formalism of quantum information theory to include so-called *quantum channels*, which provide a natural model for arbitrary sequences of operations composed of unitaries, measurements, adding and removing auxiliary systems, etc. On a mathematical level, they are described by completely positive, trace-preserving maps.

- *Noisy communication channels and their capacities:* Throughout these lectures, we always assumed that we could transmit bits, qubits, etc. in a perfect way from Alice and Bob. (In contrast, our quantum data sources were noisy and we often considered arbitrary quantum states shared between Alice and Bob quantum states, not just idealized resource states such as ebits.) An important part of quantum information research is to determine the ultimate capacities of noisy communication channels to transmit bits, qubits, etc.

See, e.g., [NC10, Wil17] for much more material than what we had time to discuss this term.

Appendix A

Handout: The formalism of quantum information theory

This handout summarizes the formalism of quantum information theory that is developed in this course.

(A) **Systems**: To every quantum mechanical system, we associate a *Hilbert space* $\mathcal{H}$. For a joint system composed of two subsystems A and B, with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the Hilbert space is the tensor product $\mathcal{H}_{AB} := \mathcal{H}_A \otimes \mathcal{H}_B$.

(B) **States**: A *density operator* ρ is an operator on $\mathcal{H}$ that satisfies (i) $\rho \geq 0$ and (ii) $\mathrm{tr}[\rho] = 1$. Any density operator describes the state of a quantum mechanical system. If the rank of ρ is one (i.e., of the form $\rho = \psi := |\psi\rangle\langle\psi|$ for some unit vector $|\psi\rangle \in \mathcal{H}$) then we say that ρ is a *pure state*. Otherwise, ρ is called a *mixed state*. An *ensemble* $\{p_i, \rho_i\}$ of quantum states can be described by the density operator $\rho = \sum_i p_i \rho_i$.
If ρ_{AB} is the state of a joint system, the state of its subsystems can be described by the *reduced density matrices* $\rho_A = \mathrm{tr}_B[\rho_{AB}]$ and $\rho_B = \mathrm{tr}_A[\rho_{AB}]$. The latter states can be mixed even if ρ_{AB} is pure. Conversely, any density operator ρ_A has a *purification* $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ (see Chapter 7).

(C) **Unitary dynamics**: Given a *unitary* operator U on $\mathcal{H}$, the transformation $\rho \mapsto U\rho U^\dagger$ is in principle physical. In other words, the laws of quantum mechanics allow a way of evolving the quantum system for some finite time such that, when we start in an arbitrary initial state ρ, the final state is $U\rho U^\dagger$. If $\rho = |\psi\rangle\langle\psi|$ is a pure state, then this corresponds to $|\psi\rangle \mapsto U|\psi\rangle$.

(D) **Measurements**: A *POVM measurement* $\{Q_x\}_{x\in\Omega}$ with outcomes in some finite set Ω is a collection of operators on $\mathcal{H}$ that satisfies (i) $Q_x \geq 0$ and (ii) $\sum_{x\in\Omega} Q_x = I$. Born's rule asserts that the probability of outcome x in state ρ is given by the *Born rule*:

$$\mathbf{Pr}_\rho(\text{outcome } x) = \mathrm{tr}\left[\rho Q_x\right].$$

If $\rho = |\psi\rangle\langle\psi|$ is a pure state, then this can also be written as $\langle\psi|Q_x|\psi\rangle$. A POVM measurement that has precisely two outcomes is called a *binary POVM measurement*, and it has the form $\{Q, I-Q\}$, hence is specified by a single POVM element $0 \leq Q \leq I$. We can also consider POVMs with a continuum of possible outcomes (see Chapter 4). We say that $\{P_x\}$ is a *projective measurement* if $\{P_x\}_{x\in\Omega}$ is a POVM where the P_x are projections that are pairwise orthogonal (i.e., $Q_xQ_y = \delta_{x,y}Q_x$). If $\Omega \subseteq \mathbb{R}$, then the data $\{P_x\}_{x\in\Omega}$ is equivalent to specifying a Hermitian operator with spectral decomposition

$O = \sum_x xP_x$, called an *observable*. If the outcome of a projective measurement is x then the state of the system "collapses" into the *post-measurement state*

$$\rho' = \frac{P_x \rho P_x}{\mathrm{tr}[P_x \rho]}$$

If $\rho = |\psi\rangle\langle\psi|$ is a pure state, then $\rho' = |\psi'\rangle\langle\psi'|$, where $|\psi'\rangle = P_x|\psi\rangle / \|P_x|\psi\rangle\|$.
Any POVM can be implemented using projective measurements on a larger system (see Chapter 2).

(E) **Operations on subsystems:** Consider a joint system with Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. If we want to perform a unitary U_A on the subsystem modeled by $\mathcal{H}_A$, then the appropriate unitary on the joint system is $U_A \otimes I_B$. Similarly, if $\{Q_{A,x}\}_{x\in\Omega}$ is a POVM measurement on $\mathcal{H}_A$ then the appropriate POVM measurement on the joint system is $\{Q_{A,x} \otimes I_B\}_{x\in\Omega}$.

The standard formalism of quantum information theory includes two further notions that we did not discuss in this course: *Quantum channels* model general evolutions that can be obtained by composing unitary dynamics, adding ancillas, and taking partial traces. *Quantum instruments* can be thought of as implementations of POVM measurements that not only describe the statistics of outcomes but also model the post-measurement state.

Index

Bibliography

[AM16] Antonio Acín and Lluis Masanes. Certified randomness in quantum physics. *Nature*, 540:213–219, 2016. arXiv:1708.00265. Cited on p. 27.

[BBC+93] Charles H Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895, 1993. Cited on p. 16.

[BCH06] Dave Bacon, Isaac L Chuang, and Aram W Harrow. Efficient quantum circuits for Schur and Clebsch-Gordan transforms. *Physical Review Letters*, 97(17):170502, 2006. arXiv:quant-ph/0407082. Cited on p. 120.

[BCH07] Dave Bacon, Isaac L Chuang, and Aram W Harrow. The quantum schur transform: I. efficient qudit circuits. In *Proceedings of the 18th annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2007)*, pages 1235–1244. SIAM, 2007. arXiv:quant-ph/0601001. Cited on p. 120.

[BCHW16] Fernando GSL Brandao, Matthias Christandl, Aram W Harrow, and Michael Walter. The Mathematics of Entanglement. 2016. arXiv:1604.01790. Cited on p. 36, 65, and 68.

[BCWDW01] Harry Buhrman, Richard Cleve, John Watrous, and Ronald De Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16):167902, 2001. arXiv:quant-ph/0102001. Cited on p. 119.

[BW92] Charles H Bennett and Stephen J Wiesner. Communication via one-and two-particle operators on einstein-podolsky-rosen states. *Physical Review Letters*, 69(20):2881, 1992. Cited on p. 15.

[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):190403, 2007. Cited on p. 33.

[Chi10] Giulio Chiribella. On quantum estimation, quantum cloning and finite quantum de finetti theorems. In *Proceedings of Theory of Quantum Computation, Communication, and Cryptography (TQC 2011)*, volume 6519/2011 of *Lecture Notes in Computer Science*, pages 9–25. Springer, 2010. arXiv:1010.1875. Cited on p. 36.

[Chr06] Matthias Christandl. *The structure of bipartite quantum states-insights from group theory and cryptography*. PhD thesis, 2006. arXiv:quant-ph/0604183. Cited on p. 112.

[Chr10] Matthias Christandl. Symmetries in Quantum Information Theory. 2010. URL: http://edu.itp.phys.ethz.ch/hs10/sqit/. Cited on p. 120.

[CK11] Roger Colbeck and Adrian Kent. Private randomness expansion with untrusted devices. *Journal of Physics A: Mathematical and Theoretical*, 44(9):095305, 2011. arXiv:1011.4474. Cited on p. 27.

[CM06] 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. Cited on p. 89.

[Col09] Roger Colbeck. *Quantum and relativistic protocols for secure multi-party computation*. PhD thesis, 2009. arXiv:0911.3814. Cited on p. 26.

[DPS02] Andrew C Doherty, Pablo A Parrilo, and Federico M Spedalieri. Distinguishing separable and entangled states. *Physical Review Letters*, 88(18):187904, 2002. arXiv:quant-ph/0112007. Cited on p. 65.

[DPS04] Andrew C Doherty, Pablo A Parrilo, and Federico M Spedalieri. Complete family of separability criteria. *Physical Review A*, 69(2):022308, 2004. arXiv:quant-ph/0308032. Cited on p. 65.

[DW19] Ronald De Wolf. Quantum Computing: Lecture Notes. 2019. arXiv:1907.09415. Cited on p. 115.

[EGH+11] Pavel I Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina. *Introduction to Representation Theory*, volume 59 of *Student Mathematical Library*. American Mathematical Society, 2011. arXiv:0901.0827. Cited on p. 112.

[FH13] William Fulton and Joe Harris. *Representation Theory: A First Course*, volume 129 of *Graduate Texts in Mathematics*. Springer, 2013. Cited on p. 112.

[GHSZ90] Daniel M. Greenberger, Michael A. Horne, Abner Shimony, and Anton Zeilinger. Bell's theorem without inequalities. *American Journal of Physics*, 58(12):1131–1143, 1990. Cited on p. 23.

[Gur03] Leonid Gurvits. Classical deterministic complexity of edmonds' problem and quantum entanglement. In *Proceedings of the 35th annual ACM Symposium on Theory of Computing (STOC 2003)*, pages 10–19. ACM, 2003. arXiv:quant-ph/0303055. Cited on p. 63.

[Har05] Aram W Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. PhD thesis, 2005. arXiv:quant-ph/0512255. Cited on p. 112.

[Har13] Aram W Harrow. The church of the symmetric subspace. 2013. arXiv:1308.6595. Cited on p. 36, 39, and 120.

[HHJ+16] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the 48th annual ACM Symposium on Theory of Computing (STOC 2016)*, pages 913–925, 2016. arXiv:1508.01797. Cited on p. 109.

[HNW17] Aram W Harrow, Anand Natarajan, and Xiaodi Wu. An improved semidefinite programming hierarchy for testing entanglement. *Communications in Mathematical Physics*, 352(3):881–904, 2017. arXiv:1506.08834. Cited on p. 65.

[Key06] Michael Keyl. Quantum state estimation and large deviations. *Reviews in Mathematical Physics*, 18(01):19–60, 2006. arXiv:quant-ph/0412053. Cited on p. 109.

[KR05] Robert König and Renato Renner. A de finetti representation for finite symmetric quantum states. *Journal of Mathematical physics*, 46(12):122108, 2005. arXiv:quant-ph/0410229. Cited on p. 65.

[KW01] Michael Keyl and Reinhard F Werner. Estimating the spectrum of a density operator. *Physical Review A*, 64(5):052311, 2001. arXiv:quant-ph/0102027. Cited on p. 89.

[Mer90] N. David Mermin. Quantum mysteries revisited. *American Journal of Physics*, 58(8):731–734, 1990. Cited on p. 23.

[MY98] Dominic Mayers and Andrew Yao. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS 1998)*, pages 503–509. IEEE, 1998. arXiv:quant-ph/9809039. Cited on p. 27.

[NC10] Michael A Nielsen and Isaac L Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010. Cited on p. ii, 20, 115, 129, and 139.

[NOP09] Miguel Navascues, Masaki Owari, and Martin B Plenio. Power of symmetric extensions for entanglement detection. *Physical Review A*, 80(5):052306, 2009. arXiv:0906.2731. Cited on p. 65.

[OW16] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the 48th annual ACM Symposium on Theory of Computing (STOC 2016)*, pages 899–912, 2016. arXiv:1508.01907. Cited on p. 109.

[OW17a] Ryan O'Donnell and John Wright. Efficient quantum tomography ii. In *Proceedings of the 49th annual ACM Symposium on Theory of Computing (STOC 2017)*, pages 962–974, 2017. arXiv:1612.00034. Cited on p. 109.

[OW17b] Ryan O'Donnell and John Wright. Guest column: A primer on the statistics of longest increasing subsequences and quantum states. *SIGACT News*, 48:37–59, 2017. URL: https://www.cs.cmu.edu/~odonnell/papers/tomography-survey.pdf. Cited on p. 109.

[Pre22] John Preskill. Lecture notes for physics 229: Quantum computation, 2022. URL: http://theory.caltech.edu/~preskill/ph219/. Cited on p. ii.

[Ser77] Jean-Pierre Serre. *Linear representations of finite groups*, volume 42 of *Graduate Texts in Mathematics*. Springer, 1977. Cited on p. 39.

[Vid20] Thomas Vidick. Course fsmp, fall'20: Interactions with quantum devices, 2020. Cited on p. 29.

[Wal14] Michael Walter. *Multipartite quantum states and their marginals*. PhD thesis, 2014. arXiv:1410.6820. Cited on p. 112.

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. URL: https://cs.uwaterloo.ca/~watrous/TQI/. Cited on p. ii.

[Wil17] Mark M Wilde. *Quantum Information Theory*. Cambridge University Press, second edition, 2017. arXiv:1106.1445. Cited on p. ii and 139.