Chapter 2
A Very Brief Introduction to Quantum Computing and Quantum Information Theory for Mathematicians

Joseph M. Landsberg

Abstract This is a very brief introduction to quantum computing and quantum information theory, primarily aimed at geometers. Beyond basic definitions and examples, I emphasize aspects of interest to geometers, especially connections with asymptotic representation theory. Proofs can be found in standard references such as Kitaev et al. (Classical and quantum computation, vol. 47. American Mathematical Society, Providence, 2002) and Nielson and Chuang (Quantum computation and quantum information. Cambridge University Press, Cambridge, 2000) as well as Landsberg (Quantum computation and information: Notes for fall 2017 TAMU class, 2017).

2.1 Overview

I begin, in Sect. 2.2, by presenting the postulates of quantum mechanics as a natural generalization of probability theory. In Sect. 2.3 I describe basic entanglement phenomena of "super dense coding", "teleportation", and Bell's confirmation of the "paradox" proposed by Einstein-Podolsky-Rosen. In Sect. 2.4 I outline aspects of the basic quantum algorithms, emphasizing the geometry involved. Section 2.5 is a detour into classical information theory, which is the basis of its quantum cousin briefly discussed in Sect. 2.7. Before that, in Sect. 2.6, I reformulate quantum theory in terms of density operators, which facilitates the discussion of quantum information theory. Critical to quantum information theory is *von Neumann entropy* and in Sect. 2.8 I elaborate on some of its properties. A generalization of "teleportation" is discussed in Sect. 2.9. Regarding practical computation, the exponential growth in size of $(\mathbb{C}^2)^{\otimes n}$ with n that appears in quantum information theory leads to the notion of "feasible" states discussed in Sect. 2.10, which has interesting algebraic geometry associated to it. I conclude with a discussion of representation-theoretic

J. M. Landsberg (✉)
Department of Mathematics, Texas A&M University, College Station, TX, USA
e-mail: jml@math.tamu.edu

E. Ballico et al. (eds.), *Quantum Physics and Geometry*,
Lecture Notes of the Unione Matematica Italiana 25,
https://doi.org/10.1007/978-3-030-06122-7_2

aspects of quantum information theory, including a discussion of the quantum marginal problem in Sect. 2.11. I do not discuss topological quantum computing, which utilizes the representation theory of the braid group. For those interested in more details from this perspective, see [18].

2.2 Quantum Computation as Generalized Probabilistic Computation

In this section I take the point of view advocated in [1] and other places that quantum computing should be viewed as a natural generalization of probabilistic computing, and more generally that the laws of quantum mechanics as generalizations of the laws of probability.

2.2.1 Classical and Probabilistic Computing via Linear Algebra

This section is inspired by Arora and Barak [2, Exercise 10.4].

Classical communication deals with *bits*, elements of $\{0, 1\}$, which will be convenient to think of as elements of $\mathbb{F}_2$, the field with two elements. Let $f_n : \mathbb{F}_2^n \to \mathbb{F}_2$ be a sequence of functions. Give $\mathbb{R}^2$ basis $\{|0\rangle, |1\rangle\}$ (such notation is standard in quantum mechanics) and give $(\mathbb{R}^2)^{\otimes m} = \mathbb{R}^{2^m}$ basis $\{|I\rangle \mid I \in \{0, 1\}^m\}$. In this way, we may identify $\mathbb{F}_2^m$ with the set of basis vectors of $\mathbb{R}^{2^m}$. A computation of f_n (via an arithmetic or Boolean circuit) may be phrased as a sequence of linear maps on a vector space containing $\mathbb{R}^{2^n}$, where each linear map comes from a pre-fixed set agreed upon in advance. In anticipation of what will come in quantum computation, the pre-fixed set of maps (called *gates* in the literature) will be taken from maps having the following properties:

1. Each linear map must take probability distributions to probability distributions. This implies the matrices are *stochastic*: the entries are non-negative and each column sums to 1.
2. Each linear map only alters a small number of entries. For simplicity assume it alters at most three entries, i.e., it acts on at most $\mathbb{R}^{2^3}$ and is the identity on all other factors in the tensor product.

In quantum computation, the first property will be replaced by requiring the linear maps to be completely positive and trace preserving (see Sect. 2.7). The second is the same and justified because "universal" quantum computing is possible with such maps, even requiring the three factors to be adjacent, which is essentially due to the classical Cartan-Dieudonné theorem.

To facilitate comparison with quantum computation, first restrict to reversible classical computation. The complexity class of a sequence of functions in classical reversible computation is the same as in arbitrary classical computation.

For example, if we want to effect $(x, y) \mapsto x * y$, consider the map

$$|x, y, z\rangle \mapsto |x, y, z \oplus (x * y)\rangle = |x, y, z \oplus (x \wedge y)\rangle \tag{2.1}$$

(where the second expression is for those preferring Boolean notation) and act as the identity on all other basis vectors (sometimes called *registers*). Here z will represent "workspace bits": x, y will come from the input and z will always be set to 0 in the input. In the basis $|000\rangle, |001\rangle, |010\rangle, |100\rangle, |011\rangle, |101\rangle, |110\rangle, |111\rangle$, of $\mathbb{R}^8$, the matrix is

$$\begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0 \end{pmatrix}. \tag{2.2}$$

This gate is sometimes called the *Toffoli gate* and the matrix the *Toffoli matrix*.

The swap (negation) gate $\neg$ is realized by the matrix

$$\sigma_x = \begin{pmatrix} 0&1\\ 1&0 \end{pmatrix}. \tag{2.3}$$

The swap and Toffoli matrices can perform any computation that is accomplished via a sequence of matrices drawn from some finite set of Boolean operations, each acting on a fixed number of basis vectors with at worst a polynomial in n size increase in the number of matrices needed. For those familiar with Boolean circuits, any sequence of Boolean circuits (one for each n) may be replaced by a sequence with just Toffoli and negation gates with at worst a polynomial (in n) blow up in size.

A probability distribution on $\{0, 1\}^m$ may be encoded as a vector in $\mathbb{R}^{2^m}$: If the probability distribution assigns probability p_I to $I \in \{0, 1\}^m$, assign to the distribution the vector $v = \sum_I p_I |I\rangle \in \mathbb{R}^{2^m}$.

The matrices (2.2), (2.3) realize classical computation. To add randomness to enable probabilistic computation, introduce the matrix

$$\begin{pmatrix} \frac{1}{2}&\frac{1}{2}\\ \frac{1}{2}&\frac{1}{2} \end{pmatrix}$$

which acts on a single $\mathbb{R}^2$ corresponding to a fair coin flip. Note that the coin flip matrix is not invertible, which will be one motivation for quantum computation in Sect. 2.2.2. Work in $\mathbb{R}^{2^{n+s+r}}$ where r is the number of times one needs to access a random choice and s is the number of matrices (arithmetic operations) in addition to the coin tosses needed to compute f.

A probabilistic computation, viewed this way, starts with $|x0^{r+s}\rangle$, where $x \in \mathbb{F}_2^n$ is the input. One then applies a sequence of admissible stochastic linear maps to it, and ends with a vector that encodes a probability distribution on $\{0, 1\}^{n+s+r}$. One then restricts this to $\{0, 1\}^{p(n)}$, that is, one takes the vector and throws away all but the first $p(n)$ entries. This vector encodes a probability sub-distribution, i.e., all coefficients are non-negative and they sum to a number between zero and one. One then renormalizes (dividing each entry by the sum of the entries) to obtain a vector encoding a probability distribution on $\{0, 1\}^{p(n)}$ and then outputs the answer according to this distribution. Note that even if our calculation is feasible (i.e., polynomial in size), to write out the original output vector that one truncates would be exponential in cost. A stronger variant of this phenomenon will occur with quantum computing, where the result will be obtained with a polynomial size calculation, but one does not have access to the vector created, even using an exponential amount of computation.

To further prepare for the analogy with quantum computation, define a probabilistic bit (a *pbit*) to be an element of

$$\{p_0|0\rangle + p_1|1\rangle \mid p_j \in [0, 1] \text{ and } p_0 + p_1 = 1\} \subset \mathbb{R}^2.$$

Note that the set of pbits (possible states) is a convex set, and the basis vectors are the extremal points of this convex set.

2.2.2 A Wish List

Here is a wish list for how one might want to improve upon the above set-up:

1. Allow more general kinds of linear maps to get more computing power, while keeping the maps easy to compute.
2. Have reversible computation: we saw that classical computation can be made reversible, but the coin flip was not. This property is motivated by physics, where many physical theories require time reversibility.
3. Again motivated by physics, one would like to have a continuous evolution of the probability vector, more precisely, one would like the probability vector to depend on a continuous parameter t such that if $|\psi_{t_1}\rangle = X|\psi_{t_0}\rangle$, then there exist admissible matrices Y, Z such that $|\psi_{t_0+\frac{1}{2}t_1}\rangle = Y|\psi_{t_0}\rangle$ and $|\psi_{t_1}\rangle = Z|\psi_{t_0+\frac{1}{2}t_1}\rangle$ and $X = ZY$. In particular, one wants operators to have square roots. (Physicists sometimes state this as "time evolution being described by a semi-group".)

One way to make the coin flip reversible is, instead of making the probability distribution be determined by the sum of the coefficients, one could take the sum of the squares. If one does this, there is no harm in allowing the entries of the output vectors to become negative, and one could use

$$H := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{2.4}$$

for the coin flip. The matrix H is called the *Hadamard matrix* or *Hadamard gate* in the quantum computing literature. If we make this change, we obtain our second wish, and moreover have many operations be "continuous", because the set of matrices preserving the norm-squared of a real-valued vector is the *orthogonal group* $O(n) = \{A \in Mat_{n\times n} \mid AA^T = \mathrm{Id}\}$. So for example, any rotation has a square root.

However our third property will not be completely satisfied, as the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

which represents a reflection, does not have a square root in $O(2)$.

To have the third wish satisfied, allow vectors with *complex* entries. From now on let $i = \sqrt{-1}$. For a complex number $z = x + iy$ let $\overline{z} = x - iy$ denote its complex conjugate and $|z|^2 = z\overline{z}$ the square of its norm.

So we go from pbits, $\{p|0\rangle + q|1\rangle \mid \ p, q \geq 0 \text{ and } p + q = 1\}$ to *qubits*, the set of which is

$$\{\alpha|0\rangle + \beta|1\rangle \mid \alpha, \beta \in \mathbb{C} \text{ and } |\alpha|^2 + |\beta|^2 = 1\}. \tag{2.5}$$

The set of qubits, considered in terms of real parameters, looks at first like the 3-sphere S^3 in $\mathbb{R}^4 \simeq \mathbb{C}^2$. However, the probability distributions induced by $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$ are the same so it is really S^3/S^1 (the Hopf fibration), i.e., the two-sphere S^2. In the quantum literature this is referred to as the *Bloch sphere*. Geometrically, it would be more natural (especially since we have already seen the need to re-normalize in probabilistic computation) to work with projective space $\mathbb{CP}^1 \simeq S^2$ as our space of qubits, instead of a subset of $\mathbb{C}^2$. So the set of qubits is better seen as (2.5) modulo the equivalence $|\psi\rangle \sim e^{i\theta}|\psi\rangle$.

For $v = (v_1, \ldots, v_n) \in \mathbb{C}^n$, write $|v|^2 = |v_1|^2 + \cdots + |v_n|^2$. The set of stochastic matrices is now replaced by the unitary group

$$\mathbf{U}(n) := \{A \in Mat_{n\times n}(\mathbb{C}) \mid \ |Av| = |v| \ \forall |v\rangle \in \mathbb{C}^n\}.$$

The unitary group satisfies the third wish on the list: For all $A \in \mathbf{U}(n)$, there exists a matrix $B \in \mathbf{U}(n)$ satisfying $B^2 = A$.

Consider wish 1: it is an open question! However at least our generalized probabilistic computation includes our old probabilistic computation because the matrices (2.2), (2.3), (2.4) are unitary.

An indication that generalized probability may be related to quantum mechanics is that the interference patterns observed in the famous two slit experiments is manifested in generalized probability: one obtains a "random bit" by applying H to $|0\rangle$: $H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. However, if one applies a second quantum coin flip, one loses the randomness as $H^2 = \mathrm{Id}$ so $H^2|0\rangle = |0\rangle$, which, as pointed out in [1], could be interpreted as a manifestation of interference.

2.2.3 *Postulates of Quantum Mechanics and Relevant Linear Algebra*

Here are the standard postulates of quantum mechanics and relevant definitions from linear algebra.

P1 Associated to any isolated physical system is a Hilbert space $\mathcal{H}$, called the *state space*. The system is completely described at a given moment by a unit vector $|\psi\rangle \in \mathcal{H}$, called its *state vector*, which is well defined up to a phase $e^{i\theta}$ with $\theta \in \mathbb{R}$. Alternatively one may work in projective space $\mathbb{P}\mathcal{H}$.

Explanations A *Hilbert space* $\mathcal{H}$ is a (complete) complex vector space endowed with a non-degenerate Hermitian inner-product, $h : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$, where by definition h is linear in the first factor and conjugate linear in the second, $h(|v\rangle, |w\rangle) = \overline{h(|w\rangle, |v\rangle)}$ for all v, w, and $h(|v\rangle, |v\rangle) > 0$ for all $|v\rangle \neq 0$.

The Hermitian inner-product h allows an identification of $\mathcal{H}$ with $\mathcal{H}^*$ by $|w\rangle \mapsto \langle w| := h(\cdot, |w\rangle)$. This identification will be used repeatedly. Write $h(|v\rangle, |w\rangle) = \langle w|v\rangle$ and $|v| = \sqrt{\langle v|v\rangle}$ for the *length* of $|v\rangle$.

If $\mathcal{H} = \mathbb{C}^n$ with its standard basis, where $|v\rangle = (v_1, \ldots, v_n)$, the *standard Hermitian inner-product* on $\mathbb{C}^n$ is $\langle w|v\rangle = \sum_{j=1}^n \overline{w}_j v_j$. I will always assume $\mathbb{C}^n$ is equipped with its standard Hermitian inner-product.

Remark 2.2.1 When studying quantum mechanics in general, one needs to allow infinite dimensional Hilbert spaces, but in the case of quantum computing, one restricts to finite dimensional Hilbert spaces, usually $(\mathbb{C}^2)^{\otimes N}$.

P2 The state of an isolated system evolves with time according to the *Schrödinger equation*

$$i\hbar \frac{d|\psi\rangle}{dt} = X|\psi\rangle$$

where $\hbar$ is a constant (*Planck's constant*) and X is a fixed *Hermitian operator*, called the *Hamiltonian* of the system. (Physicists, enamored of the letter H, often

also use it to denote the Hamiltonian.) Here, recall that the *adjoint* of an operator $X \in \mathrm{End}(\mathcal{H})$, is the operator $X^\dagger \in \mathrm{End}(\mathcal{H})$ such that $\langle X^\dagger v|w\rangle = \langle v|Xw\rangle$ for all $v, w \in \mathcal{H}$ and X is *Hermitian* if $X = X^\dagger$. For a general Hilbert space, the Unitary group is $\mathbf{U}(\mathcal{H}) := \{U \in \mathrm{End}(\mathcal{H}) \mid |Uv| = |v| \ \forall |v\rangle \in \mathcal{H}\}$.

How is generalized probability related to Schrödinger's equation? Let $U(t) \subset \mathbf{U}(\mathcal{H})$ be a smooth curve with $U(0) = \mathrm{Id}$. Write $U'(0) = \frac{d}{dt}|_{t=0}U(t)$. Consider

$$\begin{aligned} 0 &= \frac{d}{dt}|_{t=0}\langle v|w\rangle \\ &= \frac{d}{dt}|_{t=0}\langle U(t)v|U(t)w\rangle \\ &= \langle U'(0)v|w\rangle + \langle v|U'(0)w\rangle. \end{aligned}$$

Thus $iU'(0)$ is Hermitian. We are almost at Schrödinger's equation. Let $\mathfrak{u}(\mathcal{H}) = T_{\mathrm{Id}}\mathbf{U}(\mathcal{H})$ denote the Lie algebra of $\mathbf{U}(\mathcal{H})$ so $i\mathfrak{u}(\mathcal{H})$ is the space of Hermitian endomorphisms. For $X \in \mathrm{End}(\mathcal{H})$, write $X^k \in \mathrm{End}(\mathcal{H})$ for $X \cdots X$ applied k times. Write $e^X := \sum_{k=0}^{\infty} \frac{1}{k!}X^k$. If X is Hermitian, then $e^{iX} \in \mathbf{U}(\mathcal{H})$. Postulate 2 implies the system will evolve unitarily, by (assuming one starts at $t = 0$), $|\psi_t\rangle = U(t)|\psi_0\rangle$, where

$$U(t) = e^{\frac{-itX}{\hbar}}.$$

Measurements Our first two postulates dealt with isolated systems. In reality, no system is isolated and the whole universe is modeled by one enormous Hilbert space. In practice, parts of the system are sufficiently isolated that they can be treated as isolated systems. However, they are occasionally acted upon by the outside world, and one needs a way to describe this outside interference. For our purposes, the isolated systems will be the Hilbert space attached to the input in a quantum algorithm and the outside interference will be the measurement at the end. That is, after a sequence of unitary operations one obtains a vector $|\psi\rangle = \sum z_j|j\rangle$ (here implicitly assuming the Hilbert space is of countable dimension), and as in generalized probability:

P3 The probability of obtaining outcome j under a measurement is $|z_j|^2$.

In Sect. 2.6, motivated again by probability, **P1, P3** will be generalized to new postulates that give rise to the same theory, but are more convenient to work with in information theory.

A typical situation in quantum mechanics and quantum computing is that there are two or more isolated systems, say $\mathcal{H}_A, \mathcal{H}_B$ that are brought together (i.e., allowed to interact with each other) to form a larger isolated system $\mathcal{H}_{AB}$. The larger system is called the *composite system*. In classical probability, the composite space is $\{0, 1\}^{N_A} \times \{0, 1\}^{N_B}$. In our generalized probability, the composite space is $(\mathbb{C}^2)^{\otimes N_A} \otimes (\mathbb{C}^2)^{\otimes N_B} = (\mathbb{C}^2)^{\otimes (N_A+N_B)}$:

P4 The state of a composite system $\mathcal{H}_{AB}$ is the tensor product of the state spaces of the component physical systems $\mathcal{H}_A, \mathcal{H}_B$: $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$.

When dealing with composite systems, we will need to allow partial measurements whose outcomes are of the form $|I\rangle \otimes |\phi\rangle$ with $|\phi\rangle$ arbitrary.

This tensor product structure gives rise to the notion of *entanglement*, which accounts for phenomenon outside of our classical intuition, as discussed in the next section.

Definition 2.2.2 A state $|\psi\rangle \in \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ is called *separable* if it corresponds to a rank one tensor, i.e., $|\psi\rangle = |v_1\rangle \otimes \cdots \otimes |v_n\rangle$ with each $|v_j\rangle \in \mathcal{H}_j$. Otherwise it is *entangled*.

2.3 Entanglement Phenomena

2.3.1 Super-Dense Coding[1]

Physicists describe their experiments in terms of two characters, Alice and Bob. I generally follow this convention. Let $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 = \mathcal{H}_A \otimes \mathcal{H}_B$, and let $|epr\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$ (called the *EPR state* in the physics literature after Einstein-Podolsky-Rosen). Assume this state has been created, both Alice and Bob are aware of it, Alice is in possession of the first qubit, and Bob the second. In particular Alice can act on the first qubit by unitary matrices and Bob can act on the second. This all happens before the experiment begins.

Now say Alice wants to transmit a two classical bit message to Bob, i.e., one of the four states $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ by transmitting qubits. We will see that she can do so transmitting just one qubit. If she manipulates her qubit by acting on the first $\mathbb{C}^2$ by a unitary transformation, $|epr\rangle$ will be manipulated. She uses the following matrices depending on the message she wants to transmit:

to transmit	act by	to obtain
$\|00\rangle$	Id	$\frac{\|00\rangle + \|11\rangle}{\sqrt{2}}$
$\|01\rangle$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} =: \sigma_z$	$\frac{\|00\rangle - \|11\rangle}{\sqrt{2}}$
$\|10\rangle$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} =: \sigma_x$	$\frac{\|10\rangle + \|01\rangle}{\sqrt{2}}$
$\|11\rangle$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} =: -i\sigma_y$	$\frac{\|01\rangle - \|10\rangle}{\sqrt{2}}$

[1] Physicists use the word "super" in the same way American teenagers use the word "like".

where the names $\sigma_x, \sigma_y, \sigma_z$ are traditional in the physics literature (the *Pauli matrices*). If Alice sends Bob her qubit, so he is now in possession of the modified $|epr\rangle$ (although he does not see it), he can determine which of the four messages she sent him by measuring the state in his possession. More precisely, first Bob acts on $\mathbb{C}^2 \otimes \mathbb{C}^2$ by a unitary transformation that takes the orthonormal basis in the "to obtain" column to the standard orthonormal basis (this is a composition of two Hadamard matrices), to obtain a state vector whose probability is concentrated at one of the four classical states. He then measures, and obtains the correct classical state with probability one.

In summary, with preparation of an EPR state in advance, plus transmission of a single qubit, one can transmit two classical bits of information.

2.3.2 Quantum Teleportation

Here again, Alice and Bob share half of an EPR state, Alice is in possession of a qubit $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, and wants to "send" $|\psi\rangle$ to Bob. However Alice is only allowed to transmit classical information to Bob. We will see that she can accomplish her goal by transmitting two classical bits. Write the state of the system as

$$\frac{1}{\sqrt{2}}\left[\alpha|0\rangle\otimes(|00\rangle + |11\rangle) + \beta|1\rangle\otimes(|00\rangle + |11\rangle)\right]$$

where Alice can operate on the first two qubits. If Alice acts on the first two qubits by $H\otimes\sigma_x = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\otimes\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$, she obtains

$$\frac{1}{2}\left[|00\rangle\otimes(\alpha|0\rangle + \beta|1\rangle) + |01\rangle\otimes(\alpha|1\rangle + \beta|0\rangle) + |10\rangle\otimes(\alpha|0\rangle - \beta|1\rangle) + |11\rangle\otimes(\alpha|1\rangle - \beta|0\rangle)\right].$$

Notice that Bob's coefficient of Alice's $|00\rangle$ is the state $|\psi\rangle$ that is to be transmitted. Alice performs a measurement. If she has the good luck to obtain $|00\rangle$, then she knows Bob has $|\psi\rangle$ and she can tell him classically that he is in possession of $|\psi\rangle$. But say she obtains the state $|01\rangle$: the situation is still good, she knows Bob is in possession of a state such that, if he acts on it with $\sigma_x = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$, he will obtain the state $|\psi\rangle$, so she just needs to tell him classically to apply σ_x. Since they had communicated the algorithm in the past, all Alice really needs to tell Bob in the first case is the classical message 00 and in the second case the message 01. The cases of 10 and 11 are similar.

In summary, a shared EPR pair plus sending two classical bits of information allows transmission of one qubit.

Remark 2.3.1 In the literature this phenomenon is named *quantum teleportation*. Since information is transmitted at a speed slower than the speed of light, the use of the word "teleportation", which implies instantaneous transmission, is misleading.

2.3.3 Bell's Game

The 1935 Einstein-Podolsky-Rosen paper [10] challenged quantum mechanics with the following thought experiment that they believed implied instantaneous communication across distances, in violation of principles of relativity: Alice and Bob prepare $|epr\rangle = \frac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$, then travel far apart. Alice measures her bit. If she gets 0, then she can predict with certainty that Bob will get 0 in his measurement, even if his measurement is taken a second later and they are a light year apart.

Ironically, this thought experiment has been made into an actual experiment. One modern interpretation (see, e.g., [2]) is that there is no paradox because the system does not transmit information faster than the speed of light, but rather they are acting on information that has already been shared. What follows is a version from [7], adapted from the presentation in [2].

Charlie chooses $x, y \in \{0, 1\}$ at random and sends x to Alice and y to Bob. Based on this information, Alice and Bob, without communicating with each other, get to choose bits a, b and send them to Charlie. The game is such that Alice and Bob play on a team. They win if $a \oplus b = x \wedge y$, i.e., either $(x, y) \neq (1, 1)$ and $a = b$ or $(x, y) = (1, 1)$ and $a \neq b$.

2.3.3.1 Classical Version

Note that if Alice and Bob both always choose 0, they win with probability $\frac{3}{4}$.

Theorem 2.3.2 ([3]) *Regardless of the strategy Alice and Bob use, they never win with probability greater than* $\frac{3}{4}$.

See, e.g., [2, Thm. 10.3] for a proof.

2.3.3.2 Quantum Version

Although there is still no communication allowed between Alice and Bob, they will exploit a pre-shared $|epr\rangle$ to gain an advantage over the classical case. Alice and Bob prepare $|epr\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ in advance, and Alice takes the first qubit and Bob the second. When Alice gets x from Charlie, if $x = 1$, she applies a rotation by $\frac{\pi}{8}$ to her qubit, and if $x = 0$ she does nothing. When Bob gets y from Charlie, he applies a rotation by $-\frac{\pi}{8}$ to his qubit if $y = 1$ and if $y = 0$ he does nothing. (The order these rotations are applied does not matter because the corresponding operators on

$(\mathbb{C}^2)^{\otimes 2}$ commute.) Both of them measure their respective qubits and send the values obtained to Charlie.

Theorem 2.3.3 *With this strategy, Alice and Bob win with probability at least* $\frac{4}{5}$.

The idea behind the strategy is that when $(x, y) \neq (1, 1)$, the states of the two qubits will have an angle at most $\frac{\pi}{8}$ between them, but when $(x, y) = (1, 1)$, the angle will be $\frac{\pi}{4}$. That is, when $(x, y) \neq (1, 1)$, the manipulation makes it more likely that Alice and Bob's measurements produce the same outcomes, and less likely to produce the same outcome when $(x, y) = (1, 1)$. See [2, Thm. 10.4] for details.

2.4 Quantum Algorithms

Rather than giving a detailed description of the algorithms, I just present a few main ideas that illustrate the differences with classical and probabilistic algorithms.

2.4.1 Grover's Search Algorithm

The problem: given $F_n : \mathbb{F}_2^n \to \mathbb{F}_2$, computable by a $poly(n)$-size classical circuit, find a such that $F_n(a) = 1$ if such a exists.

Grover found a quantum circuit of size $poly(n)2^{\frac{n}{2}}$ that solves this problem with high probability. Compare this with a brute force search, which requires a circuit of size $poly(n)2^n$. No classical or probabilistic algorithm is known that does better than $poly(n)2^n$. Note that it also gives a size $poly(n)2^{\frac{n}{2}}$ probabilistic solution to the **NP**-complete problem SAT (it is stronger, as it not only determines existence of a solution, but finds it).

I present the algorithm for the following simplified version where one is promised there exists exactly one solution. All essential ideas of the general case are here.

Problem Given $F_n : \mathbb{F}_2^n \to \mathbb{F}_2$, computable by a $poly(n)$-size classical circuit, and the information that there is exactly one vector a with $F_n(a) = 1$, find a.

The idea of the algorithm is to start with a vector equidistant from all possible solutions, and then to incrementally rotate it towards a. What is strange for our classical intuition is that one is able to rotate towards the solution without knowing what it is, and similarly, we won't "see" the rotation matrix either.

Work in $(\mathbb{C}^2)^{\otimes n+s}$ where $s = s(n)$ is the size of the classical circuit needed to compute F_n. I suppress reference to the s "workspace bits" in what follows.

The following vector is the average of all the classical (observable) states:

$$|av\rangle := \frac{1}{2^{\frac{n}{2}}} \sum_{I \in \{0,1\}^n} |I\rangle. \tag{2.6}$$

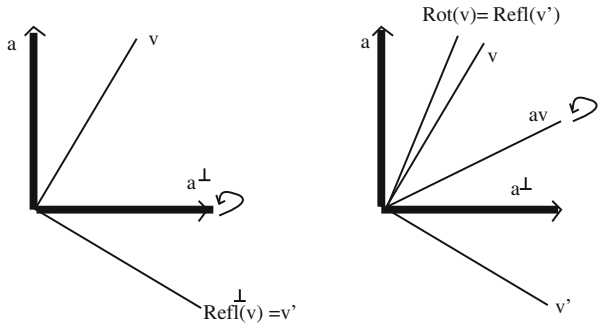


Fig. 2.1 Rotation of v to $Rot(v)$ via two reflections

To prepare $|av\rangle$, note that $H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, so applying $H^{\otimes n}$ to $|0\cdots 0\rangle$ transforms it to $|av\rangle$. The cost of this is n gates (matrices).

Since $|av\rangle$ is equidistant from all possible solution vectors, $\langle av|a\rangle = \frac{1}{2^{\frac{n}{2}}}$. We want to rotate $|av\rangle$ towards the unknown a. Recall that $\cos(\angle(|v\rangle, |w\rangle)) = \frac{\langle v|w\rangle}{|v||w|}$. Write the angle between av and a as $\frac{\pi}{2} - \theta$, so $\sin(\theta) = \frac{1}{2^{\frac{n}{2}}}$.

A rotation is a product of two reflections. In order to perform the rotation Rot that moves $|av\rangle$ towards $|a\rangle$, first reflect in the hyperplane orthogonal to $|a\rangle$, and then in the hyperplane orthogonal to $|av\rangle$, as in Fig. 2.1, which is valid for rotating any vector $|v\rangle$ towards $|a\rangle$.

Consider the map

$$|xy\rangle \mapsto |x(y \oplus F(x))\rangle \tag{2.7}$$

defined on basis vectors and extended linearly. To execute this, use the s workspace bits that are suppressed from the notation, to effect s reversible classical gates. Initially set $y = 0$ so that the image is $|x0\rangle$ for $x \neq a$, and $|x1\rangle$ when $x = a$. Next apply the quantum gate $\mathrm{Id} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ which sends $|x0\rangle \mapsto |x0\rangle$, and $|x1\rangle \mapsto -|x1\rangle$. Finally apply the map $|xy\rangle \mapsto |x(y \oplus F(x))\rangle$ again.

Thus $|a0\rangle \mapsto -|a0\rangle$ and all other vectors $|b0\rangle$ are mapped to themselves, as desired.

Next we need to reflect around $|av\rangle$. It is easy to reflect around a classical state, so first perform the map $H^{-1\otimes n} = H^{\otimes n}$ that sends $|av\rangle$ to $|0\cdots 0\rangle$, then reflect in the hyperplane perpendicular to $|0\cdots 0\rangle$ using the Boolean function $g : \mathbb{F}_2^n \to \mathbb{F}_2$ that outputs 1 if and only if its input is $(0, \ldots, 0)$, in the role of F for our previous reflection, then apply $H^{\otimes n}$ again so the resulting reflection is about $|av\rangle$. (Note that both these reflections have polynomial size cost.)

The composition of these two reflections is the desired rotation Rot. The vector $Rot|av\rangle$ is not useful as measuring it only slightly increases the probability of

obtaining $|a\rangle$, but if one composes Rot with itself $O(\frac{1}{\theta})$ times, one obtains a vector much closer to $|a\rangle$. (Note that $\theta \sim \sin(\theta)$ so $\frac{1}{\theta} \sim \sqrt{N}$.)

For more details, see, e.g., [2, Thm. 10.13] or [20, §6.1].

2.4.2 The Quantum Discrete Fourier Transform

Underlying the famous quantum algorithm of Shor for factoring integers and Simon's algorithm that led up to it, are "quantum" versions of the discrete Fourier transform on finite abelian groups.

The DFT for $\mathbb{Z}/M\mathbb{Z}$, in vector notation, for $j \in \mathbb{Z}/M\mathbb{Z}$, is

$$|j\rangle \mapsto \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} \omega^{jk} |k\rangle$$

where $\omega = e^{\frac{2\pi i}{M}}$. It is a unitary change of basis such that in the new basis, multiplication in $\mathbb{Z}/M\mathbb{Z}$ is given by a diagonal matrix, and the classical FFT writes the DFT as a product of $O(\log(M))$ sparse matrices (each with $M << M^2$ nonzero entries), for a total cost of $O(\log(M)M) < O(M^2)$ arithmetic operations.

Say $M = 2^m$. The DFT can be written as a product of $O(m^3) = O(\log(M)^3)$ controlled local unitary operators. Hence one can approximately construct the output vector by a sequence of $poly(m)$ unitary operators from our gate set with the caveat that we won't be able to "see" it.

Here is the quantum DFT: It will be convenient to express j in binary and view $\mathbb{C}^M = (\mathbb{C}^2)^{\otimes m}$, i.e., write

$$|j\rangle = |j_1\rangle \otimes \cdots \otimes |j_m\rangle$$

where $j = j_1 2^{m-1} + j_2 2^{m-2} + \cdots + j_m 2^0$ and $j_i \in \{0, 1\}$. Write the DFT as

$$\begin{aligned}
&|j_1\rangle \otimes \cdots \otimes |j_m\rangle \\
&\mapsto \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} \omega^{jk} |k\rangle \\
&= \frac{1}{\sqrt{M}} \sum_{k_i \in \{0,1\}} \omega^{j(\sum_{l=1}^{m} k_l 2^{m-l})} |k_1\rangle \otimes \cdots \otimes |k_m\rangle \\
&= \frac{1}{\sqrt{M}} \sum_{k_i \in \{0,1\}} \bigotimes_{l=1}^{m} \left[\omega^{j k_l 2^{m-l}} |k_l\rangle \right]
\end{aligned}$$

$$= \frac{1}{\sqrt{M}} \sum_{k_i \in \{0,1\}} \bigotimes_{l=1}^{m} \left[\omega^{(j_1 2^{2m-1-l} + \cdots + j_m 2^{m-l})k_l} |k_l\rangle \right]$$

$$= \frac{1}{2^{\frac{m}{2}}} (|0\rangle + \omega^{j_m 2^{-1}} |1\rangle) \otimes (|0\rangle + \omega^{j_{m-1} 2^{-1} + j_m 2^{-2}} |1\rangle) \otimes (|0\rangle + \omega^{j_{m-2} 2^{-1} + j_{m-1} 2^{-2} + j_m 2^{-3}} |1\rangle) \tag{2.8}$$

$$\otimes \cdots \otimes (|0\rangle + \omega^{\sum_{s=0}^{m-1} j_{m-s} 2^{m-(s+1)}} |1\rangle)$$

where for the last line if $2m - s - l > m$, i.e., $s + l < m$, there is no contribution with j_s because $\omega^{2^m} = 1$, and I multiplied all terms by $1 = \omega^{-2^m}$ to have negative exponents.

It will be notationally more convenient to write the quantum circuit for this vector with the order of factors reversed, so I describe a quantum circuit that produces

$$\frac{1}{\sqrt{2}} (|0\rangle + \omega^{\sum_{s=0}^{m-1} j_{m-s} 2^{m-(s+1)}} |1\rangle) \otimes \cdots \otimes \frac{1}{\sqrt{2}} (|0\rangle + \omega^{j_{m-2} 2^{-1} + j_{m-1} 2^{-2} + j_m 2^{-3}} |1\rangle) \tag{2.9}$$

$$\otimes \frac{1}{\sqrt{2}} (|0\rangle + \omega^{j_{m-1} 2^{-1} + j_m 2^{-2}} |1\rangle) \otimes \frac{1}{\sqrt{2}} (|0\rangle + \omega^{j_m 2^{-1}} |1\rangle).$$

Set

$$R_k = \begin{pmatrix} 1 & 0 \\ 0 & \omega^{2^k} \end{pmatrix}, \tag{2.10}$$

then (2.9) is obtained as follows: first apply H to $(\mathbb{C}^2)_1$ then a linear map $\Lambda^1 R_j$, defined by $|x\rangle \otimes |y\rangle \mapsto |x\rangle \otimes R_j |y\rangle$ if $|x\rangle \neq |0\rangle$ and to $|x\rangle \otimes |y\rangle$ if $|x\rangle = |0\rangle$, to $(\mathbb{C}^2)_j \otimes (\mathbb{C}^2)_1$ for $j = 2, \ldots, m$. Note that at this point only the $(\mathbb{C}^2)_1$-term has been altered. From now on leave the $(\mathbb{C}^2)_1$-slot alone. Next apply H to $(\mathbb{C}^2)_2$ then $\Lambda^1 R_{j-1}$ to $(\mathbb{C}^2)_j \otimes (\mathbb{C}^2)_2$ for $j = 3, \ldots, m$. Then apply H to $(\mathbb{C}^2)_3$ then $\Lambda^1 R_{j-2}$ to $(\mathbb{C}^2)_j \otimes (\mathbb{C}^2)_3$ for $j = 4, \ldots, m$. Continue, until finally one just applies H to $(\mathbb{C}^2)_m$. Finally to obtain the DFT, reverse the orders of the factors (a classical operation).

In practice, one has to fix a quantum gate set, i.e., a finite set of unitary operators that will be allowed in algorithms, in advance. Thus in general it will be necessary to approximate the transformations R_k from elements of our gate set, so one only obtains an approximation of the DFT.

2.4.3 The Hidden Subgroup Problem

Given a discrete group G with a specific representation of its elements in binary, a function $f : G \to \mathbb{F}_2^n$, and a device that computes f (for unit cost), and the

knowledge that there exists a subgroup $G' \subset G$ such that $f(x) = f(y)$ if and only if $xy^{-1} \in G'$, find G'.

For finitely generated abelian groups, it is sufficient to solve the problem for $G = \mathbb{Z}^{\oplus k}$ as all finitely generated abelian groups are quotients of some $\mathbb{Z}^{\oplus k}$.

Simons algorithm is for the hidden subgroup problem with $G = \mathbb{Z}_2^{\oplus m}$, see [15, §13.1]. The DFT_2 matrix is just

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

and G' is the subgroup generated by $a \in \mathbb{Z}_2^{\oplus m}$.

Shor's algorithm for factoring (after classical preparation) amounts to the case $G = \mathbb{Z}$ and F is the function $x \mapsto a^x \bmod N$. It has generated intense interest in quantum computation because no classical or probabilistic polynomial time algorithm for factoring is known. For example, most "secure" electronic communication is based on the difficulty of factoring a number into its prime factors, so the real world impact of a quantum computer would be substantial. See, e.g., http://www.math.tamu.edu/~jml/CNSA-Suite-and-Quantum-Computing-FAQ.pdf. See, e.g., [2, §10.6] for an exposition of Shor's algorithm.

2.5 Classical Information Theory

Quantum information theory is based on classical information theory, so I review the classical theory. The discovery/invention of the bit by Tukey and its development by Shannon [22] was one of the great scientific achievements of the twentieth century, as it changed the way one views information, giving it an abstract formalism that is discussed in this section. The link to quantum information is explained in Sect. 2.7.

The basic question is: given a physical channel, e.g., a telegraph wire, what is the maximum rate of transmission of messages allowing for a small amount of error? I begin with toy examples, leading up to Shannon's two fundamental theorems.

2.5.1 Data Compression: Noiseless Channels

(Following [6]) A source emits symbols x from an alphabet $\mathcal{X}$ that we want to store efficiently so we try to encode x in a small number of bits, to say $y \in \mathcal{Y}$ in a way that one can decode it later to recover x (Fig. 2.2).

The symbols from $\mathcal{X}$ do not necessarily occur with the same frequency. Let $p = P_{\mathcal{X}}$ denote the associated probability distribution. What is the minimum possible size for $\mathcal{Y}$? Since we are dealing in bits, it will be convenient to use the logarithms of cardinalities, so define the *capacity* as $\mathrm{Cap}(P_{\mathcal{X}}) := \min \log |\mathcal{Y}|$.

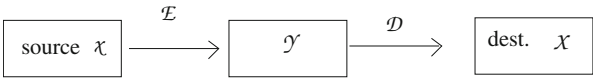


Fig. 2.2 Message from source encoded into bits then decoded

Consider the case $\mathcal{X} = \{a, b, c, d\}$ where $p(a) = 0.1$, $p(b) = 0$, $p(c) = 0.4$ and $p(d) = 0.5$. One can clearly get away with $|\mathcal{Y}| = 3$, e.g., for the encoder, send a, b to 1, c to 2 and d to 3, then for the decoder, send 1 to a, 2 to c and 3 to d. In general, one can always throw away symbols with probability zero. On the other hand, one cannot map two distinct symbols that do occur to the same symbol, as there would be no way to distinguish them when decoding. Thus $\text{Cap}(p) = \log \text{supp}(p)$, where $\text{supp}(p) = \#\{x \in \mathcal{X} \mid p(x) > 0\}$.

Now say we are willing to tolerate a small error. First rephrase what we did probabilistically: Let $p^{enc}(y|x)$ denote the conditional probability distribution of the encoder $\mathcal{E}$ and $p^{dec}(x|y)$ that of the decoder $\mathcal{D}$. Our requirement was for all x,

$$p[x = \mathcal{D} \circ \mathcal{E}(x)] = \sum_{y,x'} p^{enc}(y|x)p^{dec}(x'|y)\delta_{x,x'} = 1.$$

Now relax it to

$$\sum_{x,y,x'} p(x)p^{enc}(y|x)p^{dec}(x'|y)\delta_{x,x'} \geq 1 - \epsilon.$$

for some error ϵ that we are willing to tolerate. In addition to throwing out the symbols that do not appear, we may also discard the largest set of symbols whose total probability is smaller than ϵ. Call the corresponding quantity $\text{Cap}^{\epsilon}(p)$. In this example, if one takes $\epsilon > 0.1$, one can lower storage cost, taking $|\mathcal{Y}| = 2$.

Recall that a probability distribution $p : \mathcal{X} \to [0, 1]$ must satisfy $\sum_{x\in\mathcal{X}} p(x) = 1$. Relax this to *non-normalized* probability distributions, $q : \mathcal{X} \to [0, 1]$, where $\sum_{x\in\mathcal{X}} q(x) \leq 1$. We obtain: $\text{Cap}^{\epsilon}(p) = \min \log \text{supp}(q)$, where the min is taken over all non-normalized probability distributions q satisfying $q(x) \leq p(x)$ and $\sum_{x\in\mathcal{X}} q(x) \geq 1 - \epsilon$.

Now say we get not a single symbol, but a string of n symbols, so we seek an encoder $\mathcal{E} : \mathcal{X}^n \to \mathcal{Y}(n)$, where $\mathcal{Y}(n)$ is a set that varies with n, and decoder $\mathcal{D} : \mathcal{Y}(n) \to \mathcal{X}^n$, and we want to minimize $|\mathcal{Y}(n)|$, with a tolerance of error that goes to zero as n goes to infinity. In practice one wants to send information through a communication channel (e.g. telegraph wire). The channel can only send a limited number of bits per second, and we want to maximize the amount of information we can send per second: $\lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \text{Cap}^{\epsilon}(p^n)$.

The string $x_1 \cdots x_n =: \overline{x}^n$ is identically and independently distributed (i.i.d), that is each x_j is drawn from the same probability distribution and the draw of x_j is independent of the draws of the other x_i. Say $\mathcal{X} = \{1, \ldots, d\}$ with $p(j) = p_j$. The probability of any given string occurring depends only on the number of 1's 2's

etc. in the string and not on their order. A string with c_j j's occurs with probability $p_1^{c_1} \cdots p_d^{c_d}$. (Note that $c_1 + \cdots + c_d = n$.) The number of strings with this probability is

$$\binom{n}{c_1, \ldots, c_d} := \frac{n!}{c_1! \cdots c_d!}$$

and we need to estimate this quantity.

Stirling's formula implies $\ln(n!) = n\ln(n) - n + O(\ln(n))$. In particular, for $0 < \beta < 1$ such that $\beta n \in \mathbb{Z}$,

$$\log \binom{n}{\beta n} = n[-\beta \log(\beta) - (1-\beta)\log(1-\beta)] + O(\log(n)).$$

Let $H(\beta) = -\beta \log(\beta) - (1-\beta)\log(1-\beta)$ and more generally, for $\overline{p} = (p_1, \ldots, p_d)$, let

$$H(\overline{p}) = -\sum_{i=1}^{d} p_i \log(p_i),$$

the *Shannon entropy* of $\overline{p}$. It plays a central role in information theory.

Define a map $wt : \mathcal{X}^n \to \mathbb{R}^d$ by $\overline{x}^n \mapsto (c_1, \ldots, c_d)$, where c_j is the number of j's appearing in $\overline{x}^n$. Then the expectation is $E[wt(\overline{x}^n)] = (np_1, \ldots, np_d)$. The weak law of large numbers states that for any $\epsilon > 0$,

$$\lim_{n\to\infty} p[||\frac{1}{n} wt(\overline{x}^n) - E[wt(\overline{x}^n))]||_1 > \epsilon] = 0$$

where for $f : \mathcal{Z} \to \mathbb{R}^d$, define $||f||_1 = \sum_{z\in\mathcal{Z}} |f(z)|$. In our case, $\mathcal{Z} = \mathcal{X}^n$.

Now simply throw out all strings $\overline{x}^n$ with $||\frac{1}{n}(wt(\overline{x}^n) - E[wt(\overline{x}^n))]||_1 > \epsilon$, and take $\mathcal{Y}(n)$ of size

$$\begin{aligned} |\mathcal{Y}(n)| &= \#\{\overline{x}^n \mid ||\frac{1}{n}(wt(\overline{x}^n) - E[wt(\overline{x}^n))]||_1 < \epsilon\} \\ &= \sum_{\substack{\overline{x}^n | \\ ||\frac{1}{n}(wt(\overline{x}^n) - E[wt(\overline{x}^n))]||_1 < \epsilon}} \binom{n}{wt(\overline{x}^n)}. \end{aligned}$$

If ϵ is small, the multinomial coefficients appearing will all be very close to

$$\binom{n}{np_1, \ldots, np_d}$$

and for what follows, one can take the crude approximation

$$|\mathcal{Y}(n)| \leq poly(n)\binom{n}{np_1,\ldots,np_d} \tag{2.11}$$

(recall that d is fixed).

Taking logarithms, the right hand side of (2.11) becomes $nH(\overline{p}) + O(\log(n))$. Thus

$$\frac{1}{n}\log|\mathcal{Y}(n)| \leq H(\overline{p}) + o(1)$$

and $\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\mathrm{Cap}^{\epsilon}(p^n) \leq H(\overline{p})$.

Theorem 2.5.1 ([22]) $\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\mathrm{Cap}^{\epsilon}(p^n) = H(\overline{p})$.

The full proof uses the law of large numbers.

2.5.2 *Transmission over Noisy Channels*

Say symbols x are transmitted over a channel subject to noise, and symbols y are received so one may or may not have $y = x$. Intuitively, if the noise is small, with some redundancy it should be possible to communicate accurate messages most of the time. In a noiseless channel the maximal rate of transmission is just $H(p_{\mathcal{X}})$, but now we must subtract off something to account for the uncertainty that, upon receiving y, that it was the signal sent. This something will be the *conditional entropy*: Recall the conditional probability of i occurring given knowledge that j occurs (assuming $p(j) > 0$): $p_{\mathcal{X}|\mathcal{Y}}(i|j) = \frac{p_{\mathcal{X},\mathcal{Y}}(i,j)}{p_{\mathcal{Y}}(j)}$ (also recall $p_{\mathcal{Y}}(j) = \sum_i p_{\mathcal{X},\mathcal{Y}}(i,j)$). Define the conditional entropy

$$H(\overline{p}_{\mathcal{Y}}|\overline{p}_{\mathcal{X}}) := -\sum_{i,j} p_{\mathcal{X},\mathcal{Y}}(i,j)\log p_{\mathcal{Y}|\mathcal{X}}(j|i).$$

Note that

$$H(\overline{p}_{\mathcal{Y}}|\overline{p}_{\mathcal{X}}) = H(\overline{p}_{\mathcal{X},\mathcal{Y}}) - H(\overline{p}_{\mathcal{X}}) \tag{2.12}$$

or equivalently $H(\overline{p}_{\mathcal{X},\mathcal{Y}}) = H(\overline{p}_{\mathcal{X}}) + H(\overline{p}_{\mathcal{Y}}|\overline{p}_{\mathcal{X}})$, the uncertainty of $p_{\mathcal{X},\mathcal{Y}}$ is the uncertainty of $p_{\mathcal{X}}$ plus the uncertainty of $p_{\mathcal{Y}}$ given $p_{\mathcal{X}}$. In particular $H(\overline{p}_{\mathcal{Y}}) \geq H(\overline{p}_{\mathcal{Y}}|\overline{p}_{\mathcal{X}})$, i.e., with extra knowledge, our uncertainty about $p_{\mathcal{Y}}$ cannot increase, and decreases unless $p_{\mathcal{X}}$ and $p_{\mathcal{Y}}$ are independent.

2.5.2.1 Capacity of a Noisy Channel

Define the *capacity* of a noisy channel to be the maximum rate over all possible probability distributions on the source:

$$\text{Cap} := \max_{q_{\mathcal{X}}} \left(H(q_{\mathcal{X}}) - H(q_{\mathcal{X}}|p_{\mathcal{Y}}) \right).$$

Shannon [22] proves that Cap lives up to its name: if the entropy of a discrete channel is below Cap then there exists an encoding $\overline{p}$ of the source such that information can be transmitted over the channel with an arbitrarily small frequency of errors. The basic idea is the same as the noiseless case, however there is a novel feature that now occurs frequently in complexity theory arguments—that instead of producing an algorithm to find the efficient encoding, Shannon showed that a *random* choice of encoding will work.

After presenting the proof, Shannon remarks: "An attempt to obtain a good approximation to ideal coding by following the method of the proof is generally impractical. ... Probably this is no accident but is related to the difficulty of giving an explicit construction for a good approximation to a random sequence". To my knowledge, this is the first time that the difficulty of "finding hay in a haystack" (phrase due to Howard Karloff) is mentioned in print. This problem is central to complexity: for example, Valiant's algebraic version of $\mathbf{P} \neq \mathbf{NP}$ can be phrased as the problem of finding a sequence of explicit polynomials that are difficult to compute, while it is known that a random sequence is indeed difficult to compute. According to A. Wigderson, the difficulty of writing down random objects was also explicitly discussed by Erdös, in the context of random graphs, at least as early as 1947, in relation to his seminar paper [11]. This paper, along with [22] gave rise to the now ubiquitous probabilistic method in complexity theory.

2.6 Reformulation of Quantum Mechanics

I discuss two inconveniences about our formulation of the postulates of quantum mechanics, leading to a reformulation of the postulates in terms of density operators.

2.6.1 Partial Measurements

A measurement of a state $|\psi\rangle = \sum z_I |I\rangle$ was defined as a procedure that gives us $I = (i_1, \ldots, i_n) \in \{0,1\}^n$ with probability $|z_I|^2$. But in our algorithms, this is not what happened: we were working not in $(\mathbb{C}^2)^{\otimes n}$, but $(\mathbb{C}^2)^{\otimes n+m}$ where there were m "workspace" qubits we were not interested in measuring. So our measurement was more like the projections onto the *spaces* $|I\rangle \otimes (\mathbb{C}^2)^{\otimes m}$. I now define this generalized notion of measurement.

To make the transition, first observe that $|z_I|^2 = \langle\psi|\Pi_I|\psi\rangle$, where $\Pi_I : (\mathbb{C}^2)^{\otimes n} \to \mathbb{C}|I\rangle$ is the orthogonal projection onto the line spanned by $|I\rangle$.

Say we are only interested in the first n bits of a system of $n+m$ bits, and want to know the probability a measurement gives rise to some I represented by a vector $|I\rangle \in (\mathbb{C}^2)^{\otimes n}$, but we have $|\psi\rangle \in (\mathbb{C}^2)^{\otimes n+m}$. Adopt the notation $|\phi\rangle\langle\psi| := |\phi\rangle\otimes\langle\psi|$. Then the probability of obtaining $|I\rangle$ given $|\psi\rangle$ is

$$\begin{aligned} p(|I\rangle \mid |\psi\rangle) &= \sum_{J\in\{0,1\}^m} p(|\psi\rangle, |IJ\rangle) \\ &= \sum_J \langle\psi|IJ\rangle\langle IJ|\psi\rangle \\ &= \langle\psi|(|I\rangle\langle I|\otimes \mathrm{Id}_{(\mathbb{C}^2)^{\otimes m}})|\psi\rangle \\ &= \langle\psi|\Pi_{\mathcal{M}}|\psi\rangle \end{aligned}$$

where $\Pi_{\mathcal{M}} : (\mathbb{C}^2)^{\otimes n+m} \to |I\rangle\otimes(\mathbb{C}^2)^{\otimes m} =: \mathcal{M}$ is the orthogonal projection operator. With this definition, one can allow $\mathcal{M} \subset \mathcal{H}$ to be *any* linear subspace, which will simplify our measurements. (Earlier, if we wanted to measure the probability of a non-basis state, we had to change bases before measuring.) Write $p_\psi(\mathcal{M}) := \langle\psi|\Pi_{\mathcal{M}}|\psi\rangle$ for the probability of measuring $|\psi\rangle$ in state $\mathcal{M}$.

One may think of projection operators as representing outside interference of a quantum system, like adding a filter to beams being sent that destroy states not in $\mathcal{M}$. Recall that in classical probability, one has the identity:

$$p(M_1 \cup M_2) = p(M_1) + p(M_2) - p(M_1 \cap M_2). \tag{2.13}$$

The quantum analog is *false* in general: Let $\mathcal{H} = \mathbb{C}^2$, $\mathcal{M}_1 = \mathbb{C}|0\rangle$ and $\mathcal{M}_2 = \mathbb{C}(|0\rangle + |1\rangle)$ Let $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$. Then (and in general) $p_\psi(\mathrm{span}\{\mathcal{M}_1, \mathcal{M}_2\}) \neq p_\psi(\mathcal{M}_1) + p_\psi(\mathcal{M}_2) - p_\psi(\mathcal{M}_1 \cap \mathcal{M}_2)$.

However, one can recover (2.13) if the projection operators commute:

Proposition 2.6.1 *If* $\Pi_{\mathcal{M}_1}\Pi_{\mathcal{M}_2} = \Pi_{\mathcal{M}_2}\Pi_{\mathcal{M}_1}$ *then for all* ψ, $p_\psi(\mathrm{span}\{\mathcal{M}_1, \mathcal{M}_2\}) = p_\psi(\mathcal{M}_1) + p_\psi(\mathcal{M}_2) - p_\psi(\mathcal{M}_1 \cap \mathcal{M}_2)$.

2.6.2 Mixing Classical and Quantum Probability

A typical situation in probability is as follows: you want a cookie, but can't make up your mind which kind, so you decide to take one at random from the cookie jar to eat. However when you open the cupboard, you find there are two different cookie jars H and T, each with a different distribution of cookies, say P_H and P_T. You decide to flip a coin to decide which jar and say your coin is biased with probability

p for heads (choice H). The resulting probability distribution is

$$pP_H + (1-p)P_T.$$

Let's encode this scenario with vectors. Classically, if vectors corresponding to P_H, P_T are respectively v_H, v_T, the new vector is $pv_H + (1-p)v_T$. The probability of drawing a chocolate chip (CC) cookie is $pP_H(CC) + (1-p)P_T(CC) = pv_{H,CC} + (1-p)v_{T,CC}$.

But what should one take in generalized probability (where one uses the ℓ_2 norm instead of the ℓ_1 norm)? Given $|\psi_A\rangle = \sum z_I|I\rangle$, $|\psi_B\rangle = \sum w_J|J\rangle$, we want to make a measurement that gives us $p|z_{CC}|^2 + (1-p)|w_{CC}|^2$. Unfortunately $|pz_{CC} + (1-p)w_{CC}|^2 \neq p|z_{CC}|^2 + (1-p)|w_{CC}|^2$ in general. To fix this problem one enlarges the notion of state and further modifies the definition of measurement.

Our problem comes from having a mixture of ℓ_1 and ℓ_2 norms. The fix will be to rewrite $|\psi\rangle$ in a way that the ℓ_2 norm becomes an ℓ_1 norm. That is, construct an object that naturally contains the squares of the norms of the coefficients of $|\psi_A\rangle$. Consider the endomorphism $|\psi_A\rangle\langle\psi_A| = \sum_{I,J} z_I\overline{z}_J|I\rangle\langle J|$. It is rank one, and in the standard basis its diagonal entries are the quantities we want.

To measure them, let Π_J denote the projection onto the J-th coordinate. Then

$$\operatorname{trace}(\Pi_J|\psi_A\rangle\langle\psi_A|) = |z_{A,J}|^2$$

is the desired quantity.

Now back to our cookie jars, set

$$\rho = p|\psi_A\rangle\langle\psi_A| + (1-p)|\psi_B\rangle\langle\psi_B|$$

and observe that

$$\operatorname{trace}(\Pi_J\rho) = p|z_{A,J}|^2 + (1-p)|z_{B,J}|^2$$

as desired.

Given a finite set of states $\{|\psi_1\rangle, \ldots, |\psi_s\rangle\}$, with $p(|\psi_i\rangle) = p_i$, and $\sum_i p_i = 1$, set $\rho = \sum_k p_k|\psi_k\rangle\langle\psi_k| \in \operatorname{End}(\mathcal{H})$. Note that ρ has the properties

(1) $\rho = \rho^\dagger$, i.e., ρ is Hermitian,
(2) $\forall|\eta\rangle$, $\langle\eta|\rho|\eta\rangle \geq 0$, i.e., ρ is *positive*,
(3) $\operatorname{trace}(\rho) = 1$.

This motivates the following definition:

Definition 2.6.2 An operator $\rho \in \operatorname{End}(\mathcal{H})$ satisfying 1,2,3 above is called a *density operator*.

Note that a density operator that is diagonal in the standard basis of $\mathbb{C}^d$ corresponds to a probability distribution on $\{1, \ldots, d\}$, so the definition includes classical probability as well as our old notion of state (which are the rank one

density operators). The set of density operators is invariant under the induced action of $\mathbf{U}(\mathcal{H})$ on $\mathrm{End}(\mathcal{H})$.

Different scenarios can lead to the same density operator. However, two states with the same density operator are physically indistinguishable.

2.6.3 Reformulation of the Postulates of Quantum Mechanics

Postulate 1 Associated to any isolated physical system is a Hilbert space $\mathcal{H}$, call the *state space*. The system is described by its density operator $\rho \in \mathrm{End}(\mathcal{H})$.

Postulate 2 The evolution of an isolated system is described by the action of unitary operators on ρ.

Postulate 3 Measurements correspond to a collection of projection operators $\Pi_{\mathcal{M}_j}$ such that $\sum_k \Pi_{\mathcal{M}_k} = \mathrm{Id}_{\mathcal{H}}$. The probability that ρ is in measured in state $\mathcal{M}_j$ is $\mathrm{trace}(\Pi_{\mathcal{M}_j}\rho)$. Such measurements are called "Positive Operator-Valued Measurements", or POVM, in the literature.

Sometimes it is convenient to allow more general measurements than POVM:

Postulate 3′ Projective measurements correspond to a collection of Hermitian operators $X_j \in \mathrm{End}\,\mathcal{H}$ such that $\sum_k X_k = \mathrm{Id}_{\mathcal{H}}$. The probability that ρ is in measured in state X_j is $\mathrm{trace}(X_j\rho)$.

Postulate 4 regarding composite systems is unchanged.

Remark 2.6.3 Note that for $A \in \mathrm{End}\,\mathcal{H} = \mathcal{H}^*\otimes\mathcal{H}$, $\mathrm{trace}(A)$ is the image of A under the contraction map $\mathcal{H}^*\otimes\mathcal{H} \to \mathbb{C}$, $\langle v|\otimes|w\rangle \mapsto \langle v|w\rangle$. For $A \in \mathrm{End}(\mathcal{H}_1\otimes\mathcal{H}_2) = (\mathcal{H}_1^*\otimes\mathcal{H}_2^*)\otimes(\mathcal{H}_1\otimes\mathcal{H}_2)$, define the partial trace $\mathrm{trace}_{\mathcal{H}_1}(A)$ to be the image of A under the contraction $\mathcal{H}_1^*\otimes\mathcal{H}_2^*\otimes\mathcal{H}_1\otimes\mathcal{H}_2 \to \mathcal{H}_2^*\otimes\mathcal{H}_2$ given by $\langle\phi|\otimes\langle\psi|\otimes|v\rangle\otimes|w\rangle \mapsto \langle\phi|v\rangle\langle\psi|\otimes|w\rangle = \langle\phi|v\rangle|w\rangle\langle\psi|$.

2.6.4 Expectation and the Uncertainty Principle

Let $A \in \mathrm{End}(\mathcal{H})$ be a Hermitian operator with eigenvalues $\lambda_1,\ldots,\lambda_k$ and eigenspaces $\mathcal{M}_j$. If our system is in state ρ, one can consider A as a random variable that takes the value λ_j with probability $\mathrm{trace}(\Pi_{\mathcal{M}_j}\rho)$.

The expectation of a random variable $X : \mathcal{X} \to \mathbb{R}$ is $E[X] := \sum_{j\in\mathcal{X}} X(j)p(j)$.

If a system is in state ρ, the expectation of a Hermitian operator $A \in \mathrm{End}(\mathcal{H})$ is $\mathrm{trace}(A\rho)$ because $E[A] = \sum_{\lambda_j} \lambda_j \,\mathrm{trace}(\Pi_{\mathcal{M}_j}\rho) = \mathrm{trace}((\sum_{\lambda_j} \lambda_j \Pi_{\mathcal{M}_j})\rho) = \mathrm{trace}(A\rho)$.

One way mathematicians describe the famous Heisenberg uncertainty principle is that it is impossible to localize both a function and its Fourier transform. Another interpretation comes from probability:

First note that given a random variable, or Hermitian operator X, one can replace it with an operator of mean zero $\hat{X} := X - E(X\rho)\,\mathrm{Id}$. For notational convenience, I state the uncertainty principle for such shifted operators.

The variance $var(X)$ of a random variable is $var(X) = E[X - E(X)]^2$. The standard deviation $\sigma(X) = \sqrt{var(X)}$ of X is a measure of the failure of the corresponding probability distribution to be concentrated at a point, i.e., failure of the induced probability distribution to have a certain outcome.

Proposition 2.6.4 *Let X, Y be Hermitian operators of mean zero, corresponding to observables on a system in state ρ, let Then*

$$\sigma(X)\sigma(Y) \geq \frac{|\operatorname{trace}([X, Y]\rho)|}{2}.$$

The uncertainty principle says that the failure of two Hermitian operators to commute lower bounds the product of their uncertainties. In particular, if they do not commute, neither can give rise to a classical (certain) measurement. It is a consequence of the Cauchy-Schwarz inequality.

2.6.5 Pure and Mixed States

Definition 2.6.5 Let $\rho \in \mathrm{End}(\mathcal{H})$ be a density operator. If $\operatorname{rank}(\rho) = 1$, i.e. $\rho = |\xi\rangle\langle\xi|$, ρ is called a *pure state*, and otherwise it is called a *mixed state*.

The partial trace of a pure state can be a mixed state. For example, if $\rho = |\psi\rangle\langle\psi|$ with $\psi = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \in \mathcal{H}_1 \otimes \mathcal{H}_2$, then $\operatorname{trace}_{\mathcal{H}_2}(\rho) = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$.

The following proposition shows that one could avoid density operators altogether by working on a larger space:

Proposition 2.6.6 *An arbitrary mixed state $\rho \in \mathrm{End}(\mathcal{H})$ can be represented as the partial trace $\operatorname{trace}_{\mathcal{H}'} |\psi\rangle\langle\psi|$ of a pure state in $\mathrm{End}(\mathcal{H}\otimes\mathcal{H}')$ for some Hilbert space $\mathcal{H}'$. In fact, one can always take $\mathcal{H}' = \mathcal{H}^*$.*

Given a density operator $\rho \in \mathrm{End}(\mathcal{H})$, there is a well defined operator $\sqrt{\rho} \in \mathrm{End}(\mathcal{H})$ whose eigenvectors are the same as for ρ, and whose eigenvalues are the positive square roots of the eigenvalues of ρ. To prove the proposition, given $\rho \in \mathcal{H}\otimes\mathcal{H}^*$, consider $|\sqrt{\rho}\rangle\langle\sqrt{\rho}| \in \mathrm{End}(\mathcal{H}\otimes\mathcal{H}^*)$. Then $\rho = \operatorname{trace}_{\mathcal{H}^*}(|\sqrt{\rho}\rangle\langle\sqrt{\rho}|)$. A pure state whose partial trace is ρ is called a *purification* of ρ.

2.7 Communication Across a Quantum Channel

Now instead of having a source $\mathcal{X}^{\times n}$ our "source" is $\mathcal{H}^{\otimes n}$, where one can think of $\mathcal{H}^{\otimes n} = \mathcal{H}_A^{\otimes n}$, and Alice will "transmit" a state to Bob, and instead of a probability distribution p one has a density operator ρ.

What is a quantum channel? It should be a linear map sending $\rho \in \mathrm{End}(\mathcal{H}_A)$ to some $\Phi(\rho) \in \mathrm{End}(\mathcal{H}_B)$.

First consider the special case $\mathcal{H}_A = \mathcal{H}_B$. One should allow coupling with an auxiliary system, i.e.,

$$\rho \mapsto \rho \otimes \sigma \in \mathrm{End}(\mathcal{H}_A \otimes \mathcal{H}_C). \tag{2.14}$$

One should also allow the state $\rho\otimes\sigma$ to evolve in $\mathrm{End}(\mathcal{H}_A\otimes\mathcal{H}_C)$, i.e., be acted upon by an arbitrary $U \in \mathbf{U}(\mathcal{H}_A\otimes\mathcal{H}_C)$. Finally one should allow measurements, i.e., tracing out the $\mathcal{H}_C$ part. In summary, a quantum channel $\mathcal{H}_A \to \mathcal{H}_A$ is a map of the form $\rho \mapsto \mathrm{trace}_{\mathcal{H}_C}(U(\rho\otimes\sigma)U^{-1})$. More generally to go from $\mathcal{H}_A$ to $\mathcal{H}_B$, one needs to allow isometries as well. Such maps are the *completely positive trace preserving maps* (CPTP), where a map Λ is *completely positive* if $\Lambda\otimes \mathrm{Id}_{\mathcal{H}_E}$ is positive for all $\mathcal{H}_E$.

We seek an encoder $\mathcal{E}$ and decoder $\mathcal{D}$ and a compression space $\mathcal{H}_{0n}$:

$$\mathcal{H}^{\otimes n} \xrightarrow{\mathcal{E}} \mathcal{H}_{0n} = (\mathbb{C}^2)^{\otimes nR} \xrightarrow{\mathcal{D}} \mathcal{H}^{\otimes n}$$

with R as small as possible such that $\mathcal{E}\circ\mathcal{D}(\rho^{\otimes n})$ converges to $\rho^{\otimes n}$ as $n \to \infty$. To determine R, we need a quantum version of entropy.

Definition 2.7.1 The *von Neumann entropy* of a density operator ρ is $H(\rho) = -\mathrm{trace}(\rho\log(\rho))$.

Here $\log(\rho)$ is defined as follows: write ρ in terms of its eigenvectors and eigenvalues, $\rho = \sum_j \lambda_j |\psi_j\rangle\langle\psi_j|$, then $\log(\rho) = \sum_j \log(\lambda_j)|\psi_j\rangle\langle\psi_j|$.

If $\rho = \sum_j \lambda_j |\psi_j\rangle\langle\psi_j|$, then $H(\rho) = -\sum_j \lambda_j \log(\lambda_j)$ so if ρ is classical (i.e., diagonal), one obtains the Shannon entropy.

Proposition 2.7.2 *The von Neumann entropy has the following properties:*

(1) $H(\rho) \geq 0$ *with equality if and only if* ρ *is pure.*
(2) *Let* $\dim\mathcal{H} = d$. *Then* $H(\rho) \leq \log(d)$ *with equality if and only if* $\rho = \frac{1}{d}\mathrm{Id}_{\mathcal{H}}$.
(3) *If* $\rho = |\psi\rangle\langle\psi| \in \mathrm{End}(\mathcal{H}_A\otimes\mathcal{H}_B)$, *then* $H(\rho_A) = H(\rho_B)$, *where* $\rho_A = \mathrm{trace}_{\mathcal{H}_B}(\rho) \in \mathrm{End}(\mathcal{H}_A)$.

Notice that in particular von Neumann entropy is maximized for $|epr\rangle$. In Sects. 2.8 and 2.9 I discuss entanglement as a resource and von Neumann entropy as a measurement of that resource.

Theorem 2.7.3 ([21], The Quantum Noiseless Channel Theorem) *Let* $(\mathcal{H}, \rho)$ *be an i.i.d. quantum source. If* $R > H(\rho)$, *then there exists a reliable compression*

scheme of rate R. That is, there exists a compression space $\mathcal{H}_{0n}$, of dimension 2^{nR}, and encoder $\mathcal{E} : \mathcal{H}^{\otimes n} \to \mathcal{H}_{0n}$ and a decoder $\mathcal{D} : \mathcal{H}_{0n} \to \mathcal{H}^{\otimes n}$ such that $\mathcal{D} \circ \mathcal{E}(\rho^{\otimes n})$ converges to $\rho^{\otimes n}$ as $n \to \infty$. If $R < H(\rho)$, then any compression scheme is unreliable.

2.8 More on von Neumann Entropy and Its Variants

First for the classical case, define the relative entropy $H(\overline{p}||\overline{q}) := -\sum p_i \log \frac{q_i}{p_i} = -H(\overline{p}) - \sum_i p_i \log(q_i)$. It is zero when $\overline{p} = \overline{q}$ and is otherwise positive. Define the relative von Neumann entropy $H(\rho||\sigma) := \text{trace}(\rho \log(\rho)) - \text{trace}(\rho \log(\sigma))$. It shares the positivity property of its classical cousin: [16] $H(\rho||\sigma) \geq 0$ with equality if and only if $\rho = \sigma$.

Proposition 2.8.1 (von Neumann Entropy Is Non-decreasing Under Projective Measurements) *Let Π_i be a complete set of orthogonal projectors, set $\rho' = \sum_i \Pi_i \rho \Pi_i$. Then $H(\rho') \geq H(\rho)$ with equality if and only if $\rho' = \rho$.*

If we think of the entropy of ρ as a measurement of entanglement, i.e., a measurement of ρ as a communication resource, we see this resource decreases after a measurement.

Proof First note that $0 \leq H(\rho||\rho') = -H(\rho) - \text{trace}(\rho \log(\rho'))$. Now

$$\begin{aligned}\text{trace}(\rho \log(\rho')) &= \text{trace}\left(\sum_i \Pi_i \rho \log(\rho')\right) \\ &= \text{trace}\left(\sum_i \Pi_i \rho \log(\rho') \Pi_i\right)\end{aligned}$$

because $\Pi_i^2 = \Pi_i$ and $\text{trace}(AB) = \text{trace}(BA)$. Now Π_i commutes with ρ' and $\log(\rho')$ because $\Pi_i \Pi_j = 0$ if $i \neq j$, so

$$\begin{aligned}\text{trace}(\rho \log(\rho')) &= \text{trace}(\sum_i \Pi_i \rho \Pi_i \log(\rho')) \\ &= \text{trace}(\rho' \log(\rho')) \\ &= -H(\rho')\end{aligned}$$

Putting it all together, we obtain the result. □

Here and in what follows ρ_{AB} is a density operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ and $\rho_A = \text{trace}_{\mathcal{H}_B}(\rho_{AB})$, $\rho_B = \text{trace}_{\mathcal{H}_A}(\rho_{AB})$ are respectively the induced density operators on $\mathcal{H}_A, \mathcal{H}_B$.

von Neumann entropy is sub-additive: $H(\rho_{AB}) \leq H(\rho_A) + H(\rho_B)$ with equality if and only if $\rho_{AB} = \rho_A \otimes \rho_B$. It also satisfies a triangle inequality: $H(\rho_{AB}) \geq |H(\rho_A) - H(\rho_B)|$.

Recall the conditional Shannon entropy is defined to be $H(\overline{p}_{\mathcal{X}}|\overline{p}_{\mathcal{Y}}) = -\sum_{i,j} p_{\mathcal{X}\times\mathcal{Y}}(i,j) \log p_{\mathcal{X}|\mathcal{Y}}(i|j)$, the entropy of $p_{\mathcal{X}}$ conditioned on $y = j$, averaged over $\mathcal{Y}$. It is not clear how to "condition" one density matrix on another, so one needs a different definition. Recall that Shannon entropy satisfies $H(\overline{p}_{\mathcal{X}}|\overline{p}_{\mathcal{Y}}) = H(\overline{p}_{\mathcal{X}\times\mathcal{Y}}) - H(\overline{p}_{\mathcal{Y}})$, and the right hand side of this expression does make sense for density operators, so define, for ρ_{AB} a density operator on $\mathcal{H}_A \otimes \mathcal{H}_B$,

$$H(\rho_A|\rho_B) := H(\rho_{AB}) - H(\rho_B). \tag{2.15}$$

Note that $H(\rho_A|\rho_B)$ is a function of ρ_{AB}, as $\rho_B = \text{trace}_{\mathcal{H}_A} \rho_{AB}$.

WARNING: it is possible that the conditional von Neumann entropy is *negative* as it is possible that $H(\rho_B) > H(\rho_{AB})$. Consider the following example: Let $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \in \mathcal{H}_A \otimes \mathcal{H}_B$. Then $\rho_A = \frac{1}{2}\,\text{Id}_{\mathcal{H}_A} = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$ so $H(\rho_A) = 1$, but $H(|\psi\rangle\langle\psi|) = 0$ because $|\psi\rangle\langle\psi|$ is pure.

However, vestiges of positivity are true in the quantum case:

Theorem 2.8.2 (Strong Sub-additivity) *Let ρ_{ABC} be a density operator on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. Then*

$$H(\rho_C|\rho_A) + H(\rho_C|\rho_B) \geq 0 \tag{2.16}$$

and

$$H(\rho_{ABC}) - [H(\rho_{AB}) + H(\rho_{BC})] + H(\rho_B) \geq 0. \tag{2.17}$$

Strong sub-additivity has many consequences: entropy is non-increasing under operations such as conditioning, discarding a subsystem does not increase mutual information, and quantum operations (CPTP maps) do not increase mutual information, see, e.g. [20, §11.4.2] for a discussion.

2.9 Entanglement and LOCC

We have seen several ways that *entanglement* is a resource already for the space $\mathcal{H}_A \otimes \mathcal{H}_B = \mathbb{C}^2 \otimes \mathbb{C}^2$: given a shared $|epr\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, one can transport two bits of classical information using only one qubit ("super dense coding") and one can also transmit one qubit of quantum information from Alice to Bob by sending two classical bits ("teleportation").

Given a quantum state $\rho \in \mathrm{End}(\mathcal{H}_A \otimes \mathcal{H}_B)$, one would like to know how "entangled" it is, e.g., what quantum states could it be used to transport with the aid of classical communication (as in teleportation)? In this section I discuss measures of "quality of entanglement".

2.9.1 LOCC

Assume several different laboratories can communicate classically, have prepared some shared states in advance, and can perform unitary and projection operations on their parts of the states, as was the situation for quantum teleportation. More precisely, make the following assumptions:

- $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$, and the $\mathcal{H}_j$ share an entangled state $|\psi\rangle$. Often one will just have $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ and $|\psi\rangle = \alpha|00\rangle + \beta|11\rangle$.
- The laboratories can communicate classically.
- Each laboratory is allowed to perform unitary and measurement operations on their own spaces.

The above assumptions are called *LOCC* for "local operations and classical communication". It generalizes the set-up for teleportation Sect. 2.3.2.

Restrict to the case $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, each of dimension two. I will use $|epr\rangle$ as a benchmark for measuring the quality of entanglement.

We will not be concerned with a single state $|\psi\rangle$, but the tensor product of many copies of it, $|\psi\rangle^{\otimes n} \in (\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes n}$. "How much" entanglement does $|\psi\rangle^{\otimes n}$ have? An answer is given in Sect. 2.9.4.

To gain insight as to which states can be produced via LOCC from a given density operator, return to the classical case. For the classical cousin of LOCC, by considering diagonal density operators, we see we should allow alteration of a probability distribution by permuting the p_j (permutation matrices are unitary), and more generally averaging our probability measure under some probability measure on elements of $\mathfrak{S}_d$ (the classical cousin of a projective measurement), i.e., we should allow

$$\overline{p} \mapsto \sum_{\sigma \in \mathfrak{S}_d} q_\sigma \mu(\sigma)\overline{p} \tag{2.18}$$

where $\mu : \mathfrak{S}_d \to GL_d$ is map sending a permutation to a $d \times d$ permutation matrix, and q is a probability distribution on $\mathfrak{S}_d$.

This is because the unitary and projection local operators allowed amount to

$$\rho \mapsto \sum_{j=1}^{k} p_j U_j \rho {U_j}^{-1}$$

where the U_j are unitary and p is a probability distribution on $\{1,\dots,k\}$ for some finite k.

2.9.2 A Partial Order on Probability Distributions Compatible with Entropy

Shannon entropy is non-increasing under an action of the form (2.18). The partial order on probability distributions determined by (2.18) is the *dominance order*:

Definition 2.9.1 Let $x, y \in \mathbb{R}^d$, write $x^{\downarrow}$ for x re-ordered such that $x_1 \geq x_2 \geq \cdots \geq x_d$. Write $x \prec y$ if for all $k \leq d$, $\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow}$.

Note that if p is a probability distribution concentrated at a point, then $\overline{q} \prec \overline{p}$ for all probability distributions q, and if p is such that $p_j = \frac{1}{d}$ for all j, then $\overline{p} \prec \overline{q}$ for all q, and more generally the dominance order is compatible with the entropy in the sense that $\overline{p} \prec \overline{q}$ implies $H(\overline{p}) \geq H(\overline{q})$.

Recall that a matrix $D \in Mat_{d\times d}$ is doubly stochastic if $D_{ij} \geq 0$ and all column and row sums equal one. Let $\mathcal{DS}_d \subset Mat_{d\times d}$ denote the set of doubly stochastic matrices. Birkhoff [5] showed $\mathcal{DS}_d$ is the convex hull of $\mu(\mathfrak{S}_d)$, and Hardy-Littlewood-Polya [13] showed $\{x \mid x \prec y\} = \mathcal{DS}_d \cdot y$.

2.9.3 A Reduction Theorem

The study of LOCC is potentially unwieldy because there can be numerous rounds of local operations and classical communication, making it hard to model. The following result eliminates this problem:

Proposition 2.9.2 *If $|\psi\rangle \in \mathcal{H}_A\otimes\mathcal{H}_B$ can be transformed into $|\phi\rangle$ by LOCC, then it can be transformed to $|\phi\rangle$ by the following sequence of operations:*

(1) *Alice performs a single measurement with operators Π_{M_j}.*
(2) *She sends the result of her measurement (some j) to Bob classically.*
(3) *Bob performs a unitary operation on his system.*

The key point is that for any vector spaces V, W, an element $f \in V\otimes W$, may be considered as a linear map $W^* \to V$. In our case, $\mathcal{H}_B^* \simeq \mathcal{H}_B$ so $|\psi\rangle$ induces a linear map $\mathcal{H}_B \to \mathcal{H}_A$ which gives us the mechanism to transfer Bob's measurements to Alice.

For $X \in \mathcal{H}_A\otimes\mathcal{H}_B$, let singvals$(X)$ denote the set of its singular values Now I can state the main theorem on LOCC:

Theorem 2.9.3 ([19]) *For states $|\psi\rangle, |\phi\rangle \in \mathcal{H}_A\otimes\mathcal{H}_B$, $|\psi\rangle \rightsquigarrow |\phi\rangle$ by LOCC if and only if* singvals$(|\psi\rangle) \prec$ singvals$(|\phi\rangle)$.

2.9.4 Entanglement Distillation (Concentration) and Dilution

To compare the entanglement resources of two states $|\phi\rangle$ and $|\psi\rangle$, consider $|\phi\rangle^{\otimes m}$ for large m with the goal of determining the largest $n = n(m)$ such that $|\phi\rangle^{\otimes m}$ may be degenerated to $|\psi\rangle^{\otimes n}$ via LOCC. Due to the approximate and probabilistic nature of quantum computing, relax this to degenerating $|\phi\rangle^{\otimes m}$ to a state that is close to $|\psi\rangle^{\otimes n}$.

There is a subtlety for this question worth pointing out. Teleportation was defined in such a way that Alice did not need to know the state she was teleporting, but for distillation and dilution, she will need to know that its right singular vectors are standard basis vectors. More precisely, if she is in possession of $|\psi\rangle = \sqrt{p_1}|v_1\rangle\otimes|1\rangle + \sqrt{p_2}|v_2\rangle\otimes|2\rangle$, she can teleport the second half of it to Bob if they share $|epr\rangle \in \mathcal{H}_A\otimes\mathcal{H}_B$. More generally, if she is in possession of $|\psi\rangle = \sum_{j=1}^{d}\sqrt{p_j}|v_j\rangle\otimes|j\rangle \in \mathcal{H}_{A'}\otimes\mathcal{H}_{A''}$, she can teleport it to Bob if they share enough EPR states. In most textbooks, Alice is assumed to possess states whose singular vectors are $|jj\rangle$'s and I will follow that convention here. Similarly, if $|\psi\rangle = \sum_{j=1}^{d}\sqrt{p_j}|jj\rangle \in \mathcal{H}_A\otimes\mathcal{H}_B$, I discuss how many shared EPR states they can construct from a shared $|\psi\rangle^{\otimes m}$.

Define the *entanglement cost* $E_C(\psi)$ to be $\inf_m \frac{n(m)}{m}$ where $n(m)$ copies of ψ can be constructed from $|epr\rangle^{\otimes m}$ by LOCC with error going to zero as $m \to \infty$. Similarly, define the *entanglement value*, or *distillable entanglement* $E_V(\psi)$ to be $\sup_m \frac{n(m)}{m}$ where $n(m)$ copies of $|epr\rangle$ can be constructed with diminishing error from $|\psi\rangle^{\otimes m}$ by LOCC. One has $E_V(\psi) = E_C(\psi) = H(|\psi\rangle\langle\psi|)$.

Remark 2.9.4 In classical computation one can reproduce information, but this cannot be done with quantum information in general. This is because the map $|\psi\rangle \mapsto |\psi\rangle\otimes|\psi\rangle$, called the *Veronese map* in algebraic geometry, is not a linear map. This observation is called the *no cloning theorem* in the quantum literature. However, one can define a linear map, e.g., $\mathbb{C}^2 \to \mathbb{C}^2\otimes\mathbb{C}^2$ that duplicates basis vectors, i.e., $|0\rangle \mapsto |0\rangle\otimes|0\rangle$ and $|1\rangle \mapsto |1\rangle\otimes|1\rangle$. But then of course $\alpha|0\rangle + \beta|1\rangle \mapsto \alpha|0\rangle\otimes|0\rangle + \beta|1\rangle\otimes|1\rangle \neq (a|0\rangle + \beta|1\rangle)^{\otimes 2}$.

For mixed states ρ on $\mathcal{H}_A\otimes\mathcal{H}_B$, one can still define $E_C(\rho)$ and $E_V(\rho)$, but there exist examples where they differ, so there is not a canonical measure of entanglement. A wish list of what one might want from an entanglement measure E:

- Non-increasing under LOCC.
- If ρ is a product state, i.e., $\rho = |\phi_A\rangle\langle\phi_A|\otimes|\psi_B\rangle\langle\psi_B|$, then $E(\rho) = 0$.

The two conditions together imply any state constructible from a product state by LOCC should also have zero entanglement. Hence the following definition:

Definition 2.9.5 A density operator $\rho \in \mathrm{End}(\mathcal{H}_1\otimes\cdots\otimes\mathcal{H}_n)$ is *separable* if $\rho = \sum_i p_i\rho_{i,1}\otimes\cdots\otimes\rho_{i,n}$, where $\rho_{i,\alpha} \in \mathrm{End}(\mathcal{H}_\alpha)$ are density operators, $p_i \geq 0$, and $\sum_i p_i = n$. If ρ is not separable, ρ is *entangled*.

Definition 2.9.6 An *entanglement monotone* E is a function on density operators on $\mathcal{H}_A \otimes \mathcal{H}_B$ that is non-increasing under LOCC.

An example of an entanglement monotone different from E_V, E_C useful for general density operators is the *squashed entanglement* [9]

$$E_{sq}(\rho_{AB}) := \inf_C \{\frac{1}{2}[H(\rho_A \mid \rho_C) + H(\rho_B \mid \rho_C) - H(\rho_{AB} \mid \rho_C)] \mid \rho_{AB} = \text{trace}_{\mathcal{H}_C}(\rho_{ABC})\}.$$

For bipartite states, all entanglement measures are compatible with the order of states from most to least entangled. This breaks down already for tripartite states.

Remark 2.9.7 An entanglement measure appealing to geometers is SLOCC (stochastic local operations and classical communication) defined originally in [4], which asks if $|\psi\rangle \in \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_d$ is in the same $SL(\mathcal{H}_1) \times \cdots \times SL(\mathcal{H}_d)$ orbit as $|\phi\rangle \in \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_d$. If one relaxes this to orbit closure, then it amounts to being able to convert $|\psi\rangle$ to $|\phi\rangle$ with positive probability. While appealing, and while there is literature on SLOCC, given the probabilistic nature of quantum computing, its use appears to be limited to very special cases, where the orbit structure is understood (e.g., $d \leq 4$, $\dim \mathcal{H}_j = 2$).

2.10 Tensor Network States

Assuming interactions between particles should be short-ranged enough (which is satisfied in most physically relevant set-ups), if we have an arrangement of electrons, say on a circle, as in Fig. 2.3.

It is highly improbable that the electrons will share entanglement with any but their nearest neighbors. This is fortuitous, because if one is dealing with thousands of electrons and would like to describe their joint state, a priori one would have to work with a vector space of dimension 2^n, with n in the thousands, which is not

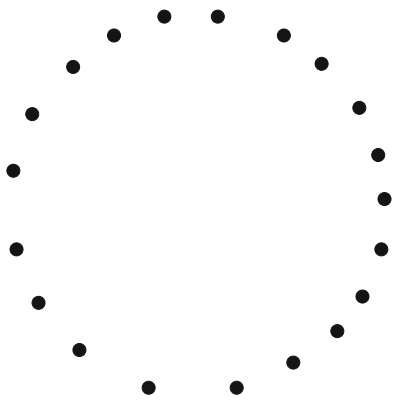


Fig. 2.3 Electrons arranged on a circle

feasible. The practical solution to this problem is to define a subset of $(\mathbb{C}^2)^{\otimes n}$ of reasonable dimension (e.g. $O(n)$) consisting of the probable states.

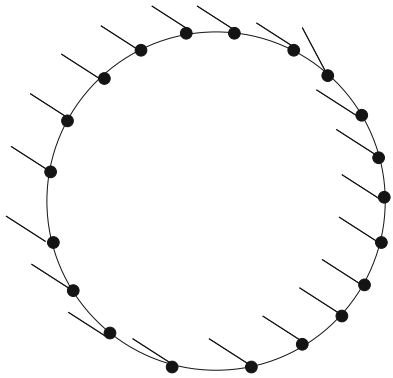
For another example, say the isolated system consists of electrons arranged along a line as below.

and we only want to allow electrons to be entangled with their nearest neighbors. This leads to the notion of *Matrix Product States (MPS)*: draw a graph reflecting this geometry, with a vertex for each electron. To each vertex, attach edges going from the electron's vertex to those of its nearest neighbors, and add an additional edge not attached to anything else (these will be called physical edges). If our space is $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$, then, assuming vertex j has two neighbors, attach two auxiliary vector spaces, E_{j-1}, E_j^*, and a tensor $T_j \in \mathcal{H}_j \otimes E_{j-1} \otimes E_j^*$. If we are on a line, to vertex one, we just attach $T_1 \in \mathcal{H}_1 \otimes E_1^*$, and similarly, to vertex n we attach $T_n \in \mathcal{H}_n \otimes E_{n-1}$. Now consider the tensor product of all the tensors

$$T_1 \otimes \cdots \otimes T_n \in (\mathcal{H}_1 \otimes E_1^*) \otimes (\mathcal{H}_2 \otimes E_1 \otimes E_2^*) \otimes \cdots \otimes (\mathcal{H}_{n-1} \otimes E_{n-2} \otimes E_{n-1}^*) \otimes (\mathcal{H}_n \otimes E_{n-1})$$

Assume each E_j has dimension k. We can contract these to obtain a tensor $T \in \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$. If $k = 1$, we just obtain the product states. As we increase k, we obtain a steadily larger subset of $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$, that fills the entire space for sufficiently large (exponential size) k. The claim is that the tensors obtainable in this fashion (for some k determined by the physical setup) are exactly those locally entangled states that we seek. (The first and last tensors in this set-up may be interpreted as boundary values, related to interaction with the outside world.)

For the circle, the only difference in the construction is to make the construction periodic, so $T_1 \in \mathcal{H}_1 \otimes E_n \otimes E_1^*$ and $T_n \in \mathcal{H}_n \otimes E_{n-1} \otimes E_n^*$. Such states are called *Matrix product states* or *MPS* in the physics literature.



Sometimes for applications (e.g. translation invariant systems on the circle) one requires the same tensor be placed at each vertex. If the tensor is $\sum_{i,j,\alpha} T_{i,j,\alpha}\langle i|\otimes|j\rangle\otimes v_\alpha$, the resulting tensor is $\sum T_{i_1,i_2,\alpha_1}T_{i_2,i_3,\alpha_2}\cdots T_{i_n,i_1,\alpha_n}v_{\alpha_1}\otimes\cdots\otimes v_{\alpha_n}$.

For a second example, consider electrons arranged in a rectangular array (or on a grid on a torus), where each vertex is allowed to interact with its four nearest neighbors. Such states are called *projected entangled pair states* or *PEPS* in the physics literature.

Assume we place the same tensor at each vertex. If our grid is $n\times n$ and periodic, we obtain a map $(\mathbb{C}^k)^{\otimes 4}\otimes\mathbb{C}^d\to(\mathbb{C}^d)^{\otimes n^2}$.

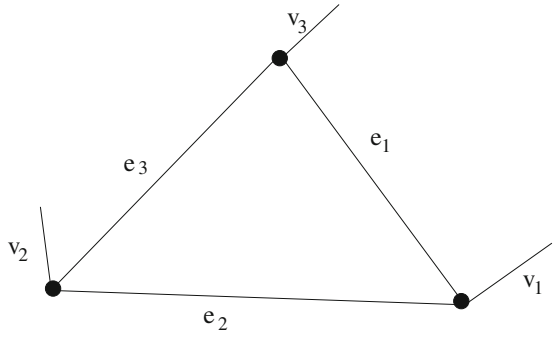
Definition 2.10.1 Let Γ be a directed graph with vertices v_α and two kinds of edges: "physical" edges e_i, that are attached to a single vertex, and "auxiliary" (or *entanglement*) edges e_s between two vertices. Associate to each physical edge a vector space V_i (in the quantum case, $V_i=\mathcal{H}_i$ is a Hilbert space), and to each auxiliary edge a vector space E_s, of dimension $\mathbf{e}_s$. Let $\overline{\mathbf{e}}=(\mathbf{e}_1,\ldots,\mathbf{e}_f)$ denote the vector of these dimensions. A *tensor network state* associated to $(\Gamma,\{V_i\},\overline{\mathbf{e}})$ is a tensor $T\in V_1\otimes\cdots\otimes V_n$ obtained as follows: To each vertex v_α, associate a tensor

$$T_\alpha\in\otimes_{i\in\alpha}V_i\otimes_{s\in in(\alpha)}E_s^*\otimes_{t\in out(\alpha)}E_t.$$

Here $in(\alpha)$ are the edges going into vertex α and $out(\alpha)$ are the edges going out of the vertex. The *tensor network state* associated to this configuration is $T:=contr(T_1\otimes\cdots\otimes T_g)\in V_1\otimes\cdots\otimes V_n$. Let $TNS(\Gamma,V_1\otimes\cdots\otimes V_n,\mathbf{e})\subset V_1\otimes\cdots\otimes V_n$ denote the set of tensor network states.

Other graphs that occur are trees, which also appear in the numerical analysis literature, see [12].

Example 2.10.2 Let Γ be:



Then

$$\begin{aligned} TNS(\Gamma, V_1 \otimes V_2 \otimes V_3, \overline{\mathbf{e}}) &= TNS(\Gamma, (E_1^* \otimes E_2) \otimes (E_2^* \otimes E_3) \otimes (E_3^* \otimes E_1), \overline{\mathbf{e}}) \\ &= \mathrm{End}(V_1) \times \mathrm{End}(V_2) \times \mathrm{End}(V_3) \cdot M_{\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle}. \end{aligned}$$

Here $M_{\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle}$ is the *matrix multiplication tensor*, for $A \in Mat_{\mathbf{e}_1 \times \mathbf{e}_2}$, $B \in Mat_{\mathbf{e}_2 \times \mathbf{e}_3}$, $C \in Mat_{\mathbf{e}_3 \times \mathbf{e}_1}$, $(A, B, C) \mapsto \mathrm{trace}(ABC)$. Let $e_1, \ldots, e_{\mathbf{e}_1}$ be a basis of E_1, $f_1, \ldots, f_{\mathbf{e}_2}$ be a basis of E_2, and $g_1, \ldots, g_{\mathbf{e}_3}$ be a basis of E_3.

There are many open questions about tensor network states: only in very few cases is there a satisfactory description of the states producible from a given graph and parameters. Regarding algebraic geometry, one can ask for a description of the ideal of the Zariski closure of the set of states producible from a given graph and parameters. Such information would be extremely useful for applications.

2.11 Representation Theory in Quantum Information Theory

Say we have a state $\rho \in \mathcal{H}_A \otimes \mathcal{H}_B$ or in $\mathcal{H}_{A_1} \otimes \cdots \otimes \mathcal{H}_{A_d}$ create-able by a device or experiment and we perform the experiment numerous times to get a state $\rho^{\otimes n} \in \mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n}$. What is the "value" of such states for information theory? What are measurements of these states likely to produce? It turns out (partial) answers to these questions can be gained by exploiting representation theory, see Theorem 2.11.5. I review the relevant representation theory and then apply it to describe the solution to the quantum marginal problem, which discusses which pairs of states on $\mathcal{H}_A$, $\mathcal{H}_B$ may arise as partial traces of some $\rho_{AB} \in \mathrm{End}(\mathcal{H}_A \otimes \mathcal{H}_B)$.

2.11.1 Review of Relevant Representation Theory

(Isomorphism classes of) irreducible representations of the permutation group $\mathfrak{S}_d$ are indexed by partitions of d, write $[\pi]$ for the $\mathfrak{S}_d$-module corresponding to the partition π. The irreducible polynomial representations of $GL(V)$ are indexed by partitions $\pi = (p_1, \ldots, p_{\ell(\pi)})$ with $\ell(\pi) \leq \dim V$. Write $S_\pi V$ for the corresponding $GL(V)$-module.

Theorem 2.11.1 (Schur-Weyl Duality) *As a $GL(V) \times \mathfrak{S}_d$-module,*

$$V^{\otimes d} = \bigoplus_{|\pi|=d} S_\pi V \otimes [\pi].$$

Let $P_\pi : V^{\otimes d} \to S_\pi V \otimes [\pi]$ denote the $GL(V) \times \mathfrak{S}_d$-module projection operator.

One is often interested in decompositions of a module under the action of a subgroup. For example $S^d(V \otimes W)$ is an irreducible $GL(V \otimes W)$-module, but as a $GL(V) \times GL(W)$-module it has the decomposition, called the *Cauchy formula*,

$$S^d(V \otimes W) = \oplus_{|\pi|=d} S_\pi V \otimes S_\pi W. \tag{2.19}$$

We will be particularly interested in the decomposition of $S^d(U \otimes V \otimes W)$ as a $GL(U) \times GL(V) \times GL(W)$-module. An explicit formula for this decomposition is *not known*. Write

$$S^d(U \otimes V \otimes W) = \bigoplus_{|\pi|,|\mu|,|\nu|=d} (S_\pi U \otimes S_\mu V \otimes S_\nu W)^{\oplus k_{\pi,\mu,\nu}}.$$

The numbers $k_{\pi,\nu,\mu}$ that record the multiplicities are called *Kronecker coefficients*. They have several additional descriptions. For example, $S_\pi(V \otimes W) = \bigoplus_{|\mu|,|\nu|=d} (S_\mu V \otimes S_\nu W)^{\oplus k_{\pi,\mu,\nu}}$, and $k_{\pi,\mu,\nu} = \dim([\pi] \otimes [\mu] \otimes [\mu])^{\mathfrak{S}_d} = \text{mult}([d], [\pi] \otimes [\mu] \otimes [\nu]) = \text{mult}([\pi], [\mu] \otimes [\nu])$.

2.11.2 Quantum Marginals and Projections onto Isotypic Subspaces of $\mathcal{H}^{\otimes d}$

In this section I address the question: what are compatibility conditions on density operators ρ on $\mathcal{H}_A \otimes \mathcal{H}_B$, ρ' on $\mathcal{H}_A$ and ρ'' on $\mathcal{H}_B$ such that $\rho' = \text{trace}_{\mathcal{H}_B}(\rho)$, $\rho'' = \text{trace}_{\mathcal{H}_A}(\rho)$? As you might expect by now, compatibility will depend only on the spectra of the operators.

Above I discussed representations of the general linear group $GL(V)$ where V is a complex vector space. In quantum theory, one is interested in representations on the unitary group $\mathbf{U}(\mathcal{H})$ on a Hilbert space $\mathcal{H}$. The unitary group is a real Lie

group, not a complex Lie group, because complex conjugation is not a complex linear map. It is a special case of a general fact about representations of a maximal compact subgroups of complex Lie groups have the same representation theory as the original group, so in particular the decomposition of $\mathcal{H}^{\otimes d}$ as a $\mathbf{U}(\mathcal{H})$-module coincides with its decomposition as a $GL(\mathcal{H})$-module.

For a partition $\pi = (p_1, \ldots, p_d)$ of d, introduce the notation $\overline{\pi} = (\frac{p_1}{d}, \ldots, \frac{p_d}{d})$ which is a probability distribution on $\{1, \ldots, d\}$.

Theorem 2.11.2 ([8]) *Let ρ_{AB} be a density operator on $\mathcal{H}_A \otimes \mathcal{H}_B$. Then there exists a sequence (π_j, μ_j, ν_j) of triples of partitions such that $k_{\pi_j,\mu_j,\nu_j} \neq 0$ for all j and*

$$\lim_{j\to\infty} \overline{\pi}_j = \operatorname{spec}(\rho_{AB})$$

$$\lim_{j\to\infty} \overline{\mu}_j = \operatorname{spec}(\rho_A)$$

$$\lim_{j\to\infty} \overline{\nu}_j = \operatorname{spec}(\rho_B).$$

Theorem 2.11.3 ([17]) *Let ρ_{AB} be a density operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ such that* $\operatorname{spec}(\rho_{AB})$, $\operatorname{spec}(\rho_A)$ *and* $\operatorname{spec}(\rho_B)$ *are all rational vectors. Then there exists an integer $M > 0$ such that*

$$k_{M \operatorname{spec}(\rho_A), M \operatorname{spec}(\rho_B), M \operatorname{spec}(\rho_C)} \neq 0.$$

Theorem 2.11.4 ([17]) *Let π, μ, ν be partitions of d with $k_{\pi,\mu,\nu} \neq 0$ and satisfying $\ell(\pi) \leq mn$, $\ell(\mu) \leq m$, and $\ell(\nu) \leq n$. Then there exists a density operator ρ_{AB} on $\mathbb{C}^n \otimes \mathbb{C}^m = \mathcal{H}_A \otimes \mathcal{H}_B$ with* $\operatorname{spec}(\rho_{AB}) = \overline{\pi}$, $\operatorname{spec}(\rho_A) = \overline{\mu}$, *and* $\operatorname{spec}(\rho_B) = \overline{\nu}$.

Klyatchko's proofs are via co-adjoint orbits and vector bundles on flag varieties, while the proof of Christandl-Mitchison is information-theoretic in flavor.

Recall the relative entropy $H(\overline{p}||\overline{q}) = -\sum_i p_i \log \frac{q_i}{p_i}$, which may be thought of as measuring how close p, q are because it is non-negative, and zero if and only if $p = q$. A key step in the Christandl-Mitchison proof is the following theorem:

Theorem 2.11.5 ([14]) *Let $\rho \in \operatorname{End}(\mathcal{H})$ be a density operator, where* $\dim \mathcal{H} = n$. *Let $|\pi| = d$ and let $P_\pi : \mathcal{H}^{\otimes d} \to S_\pi \mathcal{H} \otimes [\pi]$ be the projection operator. Then*

$$\operatorname{trace}(P_\pi \rho^{\otimes d}) \leq (d+1)^{\binom{n}{2}} e^{-dH(\overline{\pi}||\operatorname{spec}(\rho))}.$$

A key step of the proof is that the projection of e_I to $S_\pi V \otimes [\pi]$ is nonzero if and only if $wt(e_I) \prec \pi$.

Let $Spec_{m,n,mn}$ denote the set of admissible triples $(\operatorname{spec}(\rho_A), \operatorname{spec}(\rho_B), \operatorname{spec}(\rho_{AB}))$ and $KRON_{m,n,mn}$ the triples $(\overline{\mu}, \overline{\nu}, \overline{\pi})$ of normalized partitions (μ, ν, π) with $\ell(\mu) \leq m$, $\ell(\nu) \leq n$, $\ell(\pi) \leq mn$ and $k_{\pi,\mu,\nu} \neq 0$.

The theorems above imply:

$$Spec_{m,n,mn} = \overline{KRON_{m,n,mn}}.$$

In particular, $Spec_{m,n,mn}$ is a convex polytope.

Acknowledgements I thank the organizers of the International workshop on Quantum Physics and Geometry, especially Alessandra Bernardi, who also co-organized an intensive summer class on Quantum computing and quantum information theory that I gave June–July 2017. I also thank L. Chiantini, F. Gesmundo, F. Holweck, and G. Ottaviani for useful comments on a draft of this article. I am especially grateful to the anonymous referee for a very careful reading of the draft and numerous useful suggestions.

The author Landsberg supported by NSF grant DMS-1405348.

References

1. S. Aaronson, *Quantum Computing Since Democritus* (Cambridge University Press, Cambridge, 2013). MR 3058839
2. S. Arora, B. Barak, *Computational Complexity: A Modern Approach* (Cambridge University Press, Cambridge, 2009). MR 2500087 (2010i:68001)
3. J.S. Bell, On the Einstein-Podolsky-Rosen paradox. Physics **1**, 195–200 (1964)
4. C.H. Bennett, S. Popescu, D. Rohrlich, J.A. Smolin, A.V. Thapliyal, Exact and asymptotic measures of multipartite pure-state entanglement. Phys. Rev. A **63**, 012307 (2000)
5. G. Birkhoff, Three observations on linear algebra. Univ. Nac. Tucumán Rev. A. **5**, 147–151 (1946). MR 0020547
6. F.G.S.L. Brandao, M. Christandl, A.W. Harrow, M. Walter, The Mathematics of Entanglement, ArXiv e-prints (2016)
7. J.F. Clauser, M.A. Horne, A. Shimony, R.A. Holt, Proposed experiment to test local hidden-variable theories. Phys. Rev. Lett. **23**, 880–884 (1969)
8. M. Christandl, G. Mitchison, The spectra of quantum states and the Kronecker coefficients of the symmetric group. Commun. Math. Phys. **261**(3), 789–797 (2006). MR 2197548
9. M. Christandl, A. Winter, "Squashed entanglement": an additive entanglement measure. J. Math. Phys. **45**(3), 829–840 (2004). MR 2036165
10. A. Einstein, B. Podolsky, N. Rosen, Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777–780 (1935)
11. P. Erdös, Some remarks on the theory of graphs. Bull. Am. Math. Soc. **53**, 292–294 (1947). MR 0019911
12. W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*. Springer Series in Computational Mathematics, vol. 42 (Springer, Heidelberg, 2012). MR 3236394
13. G.H. Hardy, J.E. Littlewood, G. Pólya, *Inequalities*, 2nd edn. (Cambridge University Press, Cambridge, 1952). MR 0046395
14. M. Keyl, R.F. Werner, Estimating the spectrum of a density operator. Phys. Rev. A (3) **64**(5), 052311 (2001). MR 1878924
15. A.Yu. Kitaev, A.H. Shen, M.N. Vyalyi, *Classical and Quantum Computation*. Graduate Studies in Mathematics, vol. 47 (American Mathematical Society, Providence, 2002). Translated from the 1999 Russian original by Lester J. Senechal. MR 1907291
16. O. Klein, Quantum coding. Z. Phys. **72**, 767–775 (1931)
17. A. Klyachko, Quantum marginal problem and representations of the symmetric group (2004). Preprint, arXiv:quant-ph/0409113v1

18. J.M. Landsberg, Quantum computation and information: Notes for fall 2017 TAMU class (2017). Available at http://www.math.tamu.edu/~jml/quantumnotes.pdf
19. M.A. Nielsen, Conditions for a class of entanglement transformations. Phys. Rev. Lett. **83**, 436–439 (1999)
20. M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000). MR MR1796805 (2003j:81038)
21. B. Schumacher, Quantum coding. Phys. Rev. A (3) **51**(4), 2738–2747 (1995). MR 1328824
22. C.E. Shannon, A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948). MR 0026286