A concise introduction to quantum probability, quantum mechanics, and quantum computation

Greg Kuperberg*
UC Davis, visiting Cornell University
(Dated: 2005)

Quantum mechanics is one of the most interesting and surprising pillars of modern physics. Its basic precepts require only undergraduate or early graduate mathematics; but because quantum mechanics is surprising, it is more difficult than these prerequisites suggest. Moreover, the rigorous and clear rules of quantum mechanics are sometimes confused with the more difficult and less rigorous rules of quantum field theory.

Many working mathematicians have an excellent intuitive grasp of two parent theories of quantum mechanics, namely classical mechanics and probability theory. The empirical interpretations of each of these theories — above and beyond their mathematical formalism — have been a great source of ideas for mathematics proper. I believe that more mathematicians could and should learn quantum mechanics and borrow its interpretation for mathematical problems. Two subdisciplines of mathematics that have assimilated the precepts of quantum mechanics are mathematical physics and operator algebras. However, the prevailing intention of mathematical physics is the converse, to apply mathematics to problems in physics. The theory of operator algebras is closer to the spirit of this article; in this theory the precepts of quantum mechanics are sometimes called "non-commutative probability".

Recently quantum computation has entered as a new reason for both mathematicians and computer scientists to learn the precepts of quantum mechanics. Just as randomized algorithms can be moderately faster than deterministic algorithms for some computational problems (such as testing primality), some problems admit quantum algorithms that are faster (sometimes much faster) than their classical and randomized alternatives. These quantum algorithms can only run on a new kind of computer called a quantum computer. As of this writing, convincing quantum computers do not exist. Nonetheless, theoretical results suggest that quantum computers are possible rather than impossible. Entirely apart from its potential as a technology, quantum computation is a beautiful subject that combines mathematics, physics, and computer science.

This article is a concise introduction to quantum probability theory, quantum mechanics, and quantum computation for the mathematically prepared reader. Chapters 2 and 3 depend on Section 1 but not on each other, so the reader who is interested in quantum computation can go directly from Chapter 1 to Chapter 3.

This article owes a great debt to the textbook on quantum computation by Nielsen and Chuang [20], and to the Feynman Lectures, Vol. III [12]. Another good textbook written for physics students is by Sakurai [21].

Exercises

These exercises are meant to illustrate how empirical interpretations can lead to solutions of mathematical problems.

1. The probabilistic method: The Ramsey number $R(n)$ is defined as the least R such that if a simple graph Γ has R vertices, then either it or its complement must have a complete subgraph with n vertices. By considering random graphs, show that
$$R(n) \geq \frac{2^{(n-1)/2}}{(2(n!))^{1/n}}.$$
2. Angular momentum: Let S be a smooth surface of revolution about the z-axis in $\mathbb{R}^3$, and let $\vec{p}(t)$ be a geodesic arc on S, parameterized by length, that begins at the point $(1, 0, 0)$ at $t = 0$. Show that $\vec{p}(t)$ never reaches any point within $1/|p_y'(0)|$ of the vertical axis.
3. Kirchoff's laws: Suppose that a unit square is tiled by finitely many smaller squares. Show that the edge lengths are uniquely determined by the combinatorial structure of the tiling, and that they are rational. (Hint: Build the unit square out of material with unit resistivity with a battery connected to the top and bottom edges. Cut slits along the vertical edges of the tiles and affix zero-resistance wires to the horizontal edges. Each square becomes a unit resistor in an electrical network.)

*Electronic address: greg@math.ucdavis.eduandgreg@math.cornell.edu

1. QUANTUM PROBABILITY

The precepts of quantum mechanics are neither a set of physical forces nor a geometric model for physical objects. Rather, they are a variant, and ultimately a generalization, of classical probability theory. (This is following the standard Copenhagen interpretation; see Section 1.6.) Quantum probability is usually defined using the matrix mechanics model, which describes vector states (or pure states) and offers a probabilistic interpretation of final measurement. We will present this model together with an important extension to mixed states. In physics, wave mechanics is sometimes presented as an alternate definition of quantum mechanics; we will describe it as a special case of pure-state matrix mechanics.

Since classical probability is a major analogy for us, it is reviewed in Section 1.10. In short, we can think of classical probability as a category **Prob** whose objects are measure spaces (or in the finite case, finite sets) and whose morphisms are stochastic maps. (For readers who are not comfortable with this terminology, Section 1.11 is a cursory review.) Even though category theory can be very abstract [18], our interpretation of this category is very empirical: A measure space is the natural model for a physical (or otherwise empirical) object that can be in a random state, and stochastic maps are the actions on such objects that are empirically allowed in classical probability. Stochastic maps also subsume the notions of events and random variables. Finally (and crucially) the probability category **Prob** is a tensor category: A Cartesian product of measure spaces, which is in spirit a tensor product, carries the joint states of two (or more) separate probabilistic objects.

We will define a category **Quant** for quantum probability which is analogous to the category **Prob**. The ultimate generalization, discussed in Section 1.8, is a category **vN** that contains both **Quant** and **Prob**. Its objects are von Neumann algebras, which are sometimes called "noncommutative measure spaces". The objects of **Quant** are, famously, Hilbert spaces. Until Section 1.7, we will consider only finite-dimensional vector spaces. These are enough to learn from, just as the finite case is enough to learn most of the empirical interpretation of classical probability.

1.1. Vector states and unitary maps

Although it lacks some crucial empirical structure, most of quantum mechanics and much of quantum computation relies only on a simpler category (than **Quant**) which we will call **U**. The objects of **U** are complex Hilbert spaces and the morphisms are unitary maps. We also add subunitary maps to **U** to make a moderately larger category **U**'. We will also mostly restrict our attention to the subcategory $\mathbf{U}_{<\infty}$ of finite-dimensional Hilbert spaces.

Recall that a *Hilbert space* is a complex vector space $\mathcal{H}$ with a positive-definite *Hermitian inner product* $\langle\cdot|\cdot\rangle$. This means that $\langle\cdot|\cdot\rangle$ is a function from $\mathcal{H} \to \mathcal{H}$ to $\mathbb{C}$ that satisfies these axioms:

$$\begin{aligned}\langle\psi_1 + \psi_2|\psi_3\rangle &= \langle\psi_1|\psi_3\rangle + \langle\psi_2|\psi_3\rangle\\ \langle\psi_1|\psi_2\rangle &= \overline{\langle\psi_2|\psi_1\rangle}\\ \langle\psi_1|\alpha\psi_2\rangle &= \alpha\langle\psi_1|\psi_2\rangle\\ \text{for } \alpha &\in \mathbb{C}\\ \langle\psi|\psi\rangle &> 0 \text{ for } \psi \neq 0.\end{aligned}$$

(In the infinite case, $\mathcal{H}$ must also be complete relative to the norm

$$||\psi|| = \sqrt{\langle\psi|\psi\rangle}.)$$

In quantum theory, the traditional notation is $|\psi\rangle$ (a "ket") for ψ and $\langle\psi|$ (a "bra") for the dual vector

$$\langle\psi| = \psi^* = \langle\psi|\cdot\rangle.$$

If X is an operator on $\mathcal{H}$, then

$$\langle\psi_1|X|\psi_2\rangle$$

is an expression for "the inner product of ψ_1 with $X(\psi_2)$". If

$$X = |\psi_1\rangle \otimes \langle\psi_2|$$

has rank 1, then we can omit the "$\otimes$" and just write

$$X = |\psi_1\rangle\langle\psi_2|.$$

This notation is due to Dirac [10] and is called "bra-ket" notation. A linear map $U : \mathcal{H}_1 \to H_2$ is *unitary* if it preserves the inner product $\langle\cdot|\cdot\rangle$; it is *subunitary* if it preserves or decreases the attendant norm $||\cdot||$. Recall also that a linear map from a Hilbert space to itself is called an *operator*.

The standard finite example of a Hilbert space is the standard complex vector space $\mathbb{C}^n$ with the inner product

$$\langle\vec{x}|\vec{y}\rangle = \overline{x_1}y_1 + \overline{x_2}y_2 + \cdots + \overline{x_n}y_n.$$

We can generalize this to say that for any finite set A, the vector space $\mathbb{C}^A$ is a Hilbert space with standard orthonormal basis A. Every finite-dimensional Hilbert space is isomorphic to $\mathbb{C}^n$ for some n, and therefore $\mathbb{C}^A$ for any A with $|A| = n$.

In finite quantum mechanics, as in classical probability, we can define a physical object by specifying a finite set A of independent configurations. In information theory (both quantum and classical), the object is often called "Alice". In the classical case, the set of all normalized states of Alice is the simplex Δ_A spanned by A in the vector space $\mathbb{R}^A$ (see Section 1.10). *I.e.*, a general state has the form

$$\mu = \sum_{a \in A} p_a [a]$$

for probabilities $p_a \geq 0$ that sum to 1. (For unnormalized states, the sum need not be 1.) The number p_a is interpreted as the probability that Alice is in state a. Quantumly, Alice's set of *vector states* is the vector space $\mathbb{C}^A$. In formulas, a state of this type is a vector

$$|\psi\rangle = \sum_{a \in A} \alpha_a |a\rangle.$$

The state $|\psi\rangle$ is *normalized* if

$$\langle \psi | \psi \rangle = \sum_{a \in A} |\alpha_a|^2 = 1$$

and *subnormalized* if the left side is at most 1. The coefficient α_a is called the *amplitude* of the quantum state $|a\rangle$ and the square norm $|\alpha_a|^2$ is interpreted as the probability that Alice is in state $|a\rangle$. The phase of α_a (*i.e.*, its argument or angle as a complex number) has no direct probabilistic interpretation, but it will be immediately relevant when we consider operations on $|\psi\rangle$. More precisely, the relative phase of two coordinates α_a and $\alpha_{a'}$ is indirectly measurable. It will turn out that the global phase of $|\psi\rangle$ is not empirical; Section 1.4 discusses a change in formalism that eliminates it.

The state $|\psi\rangle$ is also called a *quantum superposition*, an *amplitude function*, or a *wave function*. This last name, perhaps the most common term in physics, is motivated by the fact that $|\psi\rangle$ typically satisfies a wave equation in infinite quantum mechanics (Example 1.7.1 and Section 2.1). It also predates the Copenhagen interpretation and arguably distracts from it.

If A and B are the configuration sets of two quantum systems ("Alice" and "Bob"), then, as we said, an empirical transition from Alice's state to Bob's state is a unitary (or subunitary) map

$$U : \mathbb{C}^A \to \mathbb{C}^B.$$

The requirement that U be linear is the *quantum superposition principle*. It contradicts the similar-looking classical superposition principle: if amplitudes add, then probabilities usually do not. (They will eventually be reconciled.) The entries of U are also called amplitudes, just as the entries of a stochastic map are themselves probabilities. The unitary condition is interpreted as conservation of probability. Since we have posited that $|\alpha_a|^2$ is a probability, U conserves total probability if and only if

$$||U\psi|| = ||\psi||$$

for all $\psi \in \mathbb{C}^A$. If U is allowed to extinguish the state ψ, then in general

$$||U\psi|| \leq ||\psi||$$

for all $\psi \in \mathbb{C}^A$, *i.e.*, U is subunitary.

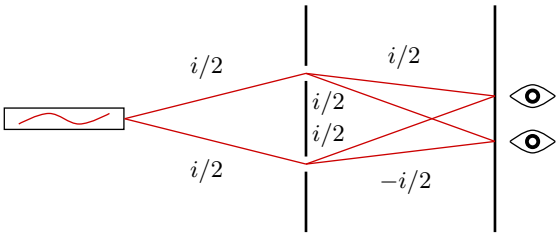


Figure 1: An idealized two-slit experiment.

It is traditional to illustrate the quantum superposition principle in an idealized setting called the "two-slit experiment" (or a more general diffraction experiment). Figure 1 shows the basic idea: A laser emits photons that can travel through either of two slits in a grating and then may (or may not) reach a detector. The source has a single state (the state set A has one element), while the grating has two states and there are two detectors (B and C each have two elements). The transitions for each photon, as it passes from A to B to C, are described by two subunitary matrices

$$U : \mathbb{C}^A \to \mathbb{C}^B \qquad V : \mathbb{C}^B \to \mathbb{C}^C.$$

The matrices are

$$U = \begin{pmatrix} \frac{i}{2} \\ \frac{i}{2} \end{pmatrix} \qquad V = \begin{pmatrix} \frac{i}{2} & \frac{i}{2} \\ \frac{i}{2} & -\frac{i}{2} \end{pmatrix},$$

and

$$VU = \begin{pmatrix} -\frac{1}{2} \\ 0 \end{pmatrix}.$$

The total amplitude of the photon reaching the top detector is $-\frac{1}{2}$ and the probability is $\frac{1}{4}$; this case is called *constructive interference*. The total amplitude reaching the bottom detector is 0, so the photon never reaches it; this case is called *destructive interference*. On the other hand, if one of the slits of

blocked, then we can discard one of the states in $|B|$, with the result that each detector is reached with probability $\frac{1}{16}$. The classical superposition principle would dictate a probability of $\frac{1}{8}$ for each detector with both slits open; thus it is violated.

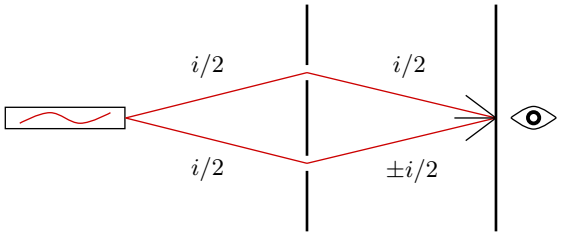


Figure 2: An angle-dependent detector in the two-slit experiment.

A natural reaction to the violation of classical superposition is to try to determine which slit the photon went through. One way to do so is to use a detector which is sensitive to the angle that the photon comes in, as in Figure 2. But then this detector represents two distinct states rather than one. Thus the final state vector is

$$|\psi\rangle = \begin{pmatrix} -\frac{1}{4} \\ \pm\frac{1}{4} \end{pmatrix}$$

and its total probability is

$$\langle\psi|\psi\rangle = ||\psi||^2 = \frac{1}{8},$$

regardless of the phases of path segments to and from the slits. The broader lesson is that amplitudes of different trajectories of an object only add when there is no evidence of which trajectory it took; otherwise the probabilities add. If we want to see quantum superposition, it is not enough to wittingly or unwittingly ignores such evidence. Rather, if the two trajectories induce different states of the universe, so that some observer could in principle distinguish them, then they obey classical superposition. Moreover, the effect is not the result of interaction between photons; photons do not interact with each other[1]. Indeed, the laser could be tuned to shoot only one photon at a time. Of course, our two-slit "experiment" is only an idealization of a real experiment; but see Sections 1.3 and 1.6.

Examples 1.1.1. A *qubit* is a two-state quantum object with configuration set $\{0, 1\}$. Two of their quantum superpositions are:

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \qquad |-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}$$

Both of these states have probability $\frac{1}{2}$ of being in either configuration $|0\rangle$ or $|1\rangle$, but they are different states. This is demonstrated by the effect of a unitary operator H called the *Hadamard gate*:

$$H = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

It exchanges $|0\rangle$ with $|+\rangle$ and $|1\rangle$ with $|-\rangle$.

The spin state of a spin-$\frac{1}{2}$ particle is a two-state system which is important in physics. (Electrons, protons, and neutrons are all spin-$\frac{1}{2}$ particles.) The conventional orthonormal basis is $|\uparrow\rangle$ ("spin up") and $|\downarrow\rangle$ ("spin down"). The names of the states refer to the property of the electron spinning (according to the right-hand rule) about a vertical axis in these two states. Even though a rotated electron is still an electron, this configuration set for it does not rotate to itself; neither does any other. The resolution of this paradox is that rotated states appear as superpositions. For example, the states "spin left" and "spin right" are analogous to $|+\rangle$ and $|-\rangle$:

$$|\rightarrow\rangle = \frac{|\uparrow\rangle + |\downarrow\rangle}{\sqrt{2}} \qquad |\leftarrow\rangle = \frac{|\uparrow\rangle - |\downarrow\rangle}{\sqrt{2}}.$$

Exercises

1. Suppose that the lengths of the entries of a complex matrix U are all fixed, but the phases are all chosen uniformly randomly. (If you like, you can also suppose that for any choice of the amplitudes, U is subunitary.) Show that on average, each entry of $U|\psi\rangle$ satisfies the classical superposition principle.

2. If U is a matrix, then the matrix

   $$M_{ab} = |U_{ab}|^2$$

   can be called *dephasing* of U. A dephasing of a unitary matrix is always doubly stochastic, meaning that the entries are non-negative and the rows and columns add to 1. Find a 3×3 doubly stochastic matrix which is not the dephasing of any unitary matrix.

3. Show that every $n \times k$ subunitary matrix U can be extended to an $(n+k) \times (n+k)$ unitary matrix V:

   $$V = \left(\begin{array}{c|c} U & * \\ \hline * & * \end{array}\right).$$

[1] More precisely, detecting photon-photon interactions requires enormous particle accelerators.

Show that V cannot usually have order less than $n+k$.

4. If $U_1, U_2, \ldots, U_n$ are unitary operators, then each entry of their product

$$U = U_n \ldots U_2 U_1$$

can be expressed as a sum of products of entries of the factors:

$$\langle a_n|U_n \ldots U_2 U_1|a_0\rangle = \sum_{a_0,a_1,\ldots,a_n} \langle a_n|U_n|a_{n-1}\rangle \ldots \langle a_2|U_1|a_1\rangle\langle a_1|U_1|a_0\rangle.$$

Such an expansion is interpreted as *path summation*; it is the same idea as a sum over histories in classical probability.

For example, let $n = 4$ and let each

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

Find the amplitudes of the 16 paths and group them according to how they sum.

5. In general for a spin-$\frac{1}{2}$ particle, the state

$$|\vec{v}\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle$$

spins in the direction

$$\vec{v} = (\text{Re } \alpha\overline{\beta}, \text{Im } \alpha\overline{\beta}, |\alpha|^2 - |\beta|^2).$$

Check that this is a unit vector when $|\vec{v}\rangle$ is normalized, and that every unit vector in $\mathbb{R}^3$ is achieved. This formula is therefore a surjective function from the unit 3-sphere $S^3 \subset \mathbb{C}^2$ to the 2-sphere $S^2 \subset \mathbb{R}^3$. What is its usual name in mathematics?

1.2. Measurements and basis independence

Suppose that $\mathcal{H}$ (or $\mathcal{H} = \mathbb{C}^A$) is the Hilbert space of a quantum object, and that the object is in the state $|\psi\rangle \in \mathcal{H}$. A *measurement* or *real-valued quantum random variable* is a Hermitian operator X on $\mathcal{H}$. The eigenvalues of X are interpreted as its range as a random variable. (Since we are assuming that $\mathcal{H}$ is finite-dimensional, X admits a complete set of orthogonal eigenvectors. For the infinite case see Section 1.7.) The assertion that $X = \lambda$ as a random variable is interpreted as the condition that $|\psi\rangle$ is an eigenvector of X with eigenvalue λ. More generally, for any $|\psi\rangle$, the probability that $X = \lambda$ is given by the formula

$$P[X = \lambda] = \langle\psi|P_\lambda|\psi\rangle,$$

where P_λ is the orthogonal projection onto the eigenspace of λ. (Note that this probability does not depend on the global phase of $|\psi\rangle$.) Moreover, if the value λ is measured, the conditional state afterward is

$$|\psi'\rangle = \frac{P_\lambda|\psi\rangle}{\sqrt{\langle\psi|P_\lambda|\psi\rangle}}.$$

Conditioning on a measurement is also called "state collapse" or "wave function collapse".

This abstract definition of a measurement, and the references to abstract Hilbert spaces, can be motivated by the more concrete discussion in Section 1.1, and they lead to a better presentation of unitary quantum probability. In Section 1.1, we tacitly accepted that if $\mathcal{H} = \mathbb{C}^A$ is Alice's state space, then one kind of a valid measurement is whether Alice is in the configuration $a \in A$, and we said that its probability of this is the square amplitude $|\alpha_a|^2$. More generally, if $D : A \to S$ is some function, then

$$P[D = s] = \sum_{D(a)=s} |\alpha_a|^2;$$

this was implied by the discussion about distinct and identical states. But, taking $S = \mathbb{R}$, the function D uniquely extends to a Hermitian operator on $\mathcal{H}$ which is diagonal in the basis A. At the same time, we posited that unitary operators represent the empirical operations on Alice. Since every Hermitian operator X is diagonalized by a unitary operator,

$$X = U^{-1}DU,$$

we can think of a general measurement X as a measurement of Alice's configuration $a \in A$ after Alice is prepared by the transition map U.

Example 1.2.1. Consider a spin-$\frac{1}{2}$ particle and let

$$J_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad J_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

be two Hermitian operators, given as matrices in the standard basis $\{\uparrow, \downarrow\}$. These operators measure the particle's spin in horizontal and vertical directions. If J_z is definite, then the spin state is either $|\uparrow\rangle$ or $|\downarrow\rangle$. Both of these states are superpositions of $|\leftarrow\rangle$ and $|\rightarrow\rangle$, so if J_z is definite, J_x is not; rather, it has a $\frac{1}{2}$ chance of being either 1 or -1. If J_x is measured, then the particle's state becomes one of the two conditional states $|\leftarrow\rangle$ or $|\rightarrow\rangle$, after which J_z is no longer definite; its old value is forgotten.

This example illustrates that every state of a quantum system is a source of randomness; every state is indefinite. The popular paraphrase of Einstein, "God does not play dice with the universe," refers to this principle.

By the same token, if $\mathcal{H}$ is the Hilbert space of a quantum object, we can think of any orthonormal basis A of $\mathcal{H}$ as its configuration set. Two completely different orthonormal bases can be equally empirical; a very important part of empirical thinking in quantum theory is to be able to change from one orthonormal basis to another. In physics such a change of description is often called a "duality". For example, one form of particle-wave duality (namely, second quantization of bosons) is very similar to an orthonormal change of basis (Section 2.6).

Example 1.2.2. We can now have a second understanding of a qubit as a quantum object with a two-dimensional Hilbert space $\mathcal{H}$. We can label any orthonormal basis $|0\rangle$ and $|1\rangle$, or we can choose not to distinguish any particular basis. For example, one person's $|0\rangle$ and $|1\rangle$ may be another person's $|+\rangle$ and $|-\rangle$. One important quantum algorithm, the Grover search algorithm (Section **??**) alternates between (dilated) classical computations in the two bases.

A spin-$\frac{1}{2}$ particle illustrates the same point more geometrically. As it happens, every orthonormal basis of its spin state space consists of the positive and negative spin states in some direction. But the model of a qubit as a spin-$\frac{1}{2}$ particle is ultimately misleading. Particle spin has been successfully employed as a qubit, but some other qubit devices have much more complicated states. Figure 3 shows one example.

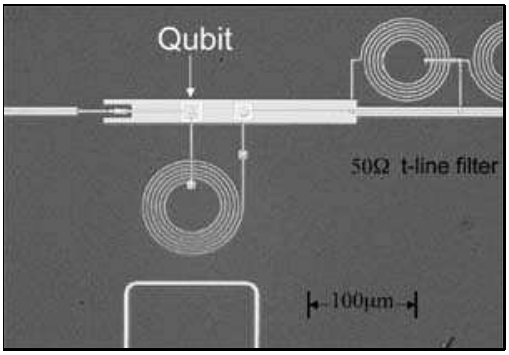


Figure 3: A Josephson junction qubit: superconducting aluminum on a silicon chip [17].

A Boolean measurement or quantum random variable can be represented as a Hermitian operator P whose eigenvalues are 0 (for "no") and 1 (for "yes"). *I.e.*, P is an orthogonal projection on $\mathcal{H}$. More generally, a random variable X that takes (discrete) values in a set S can be represented by an orthogonal decomposition

$$\mathcal{H} = \bigoplus_{s \in S} \mathcal{H}_s.$$

The outcome $X = s$ corresponds to the orthogonal projection P_s onto the summand $\mathcal{H}_s$. Its probability of occurrence in the state $|\psi\rangle$ is

$$\langle \psi | P_s | \psi \rangle,$$

which is also the squared length of the projected vector $P_s|\psi\rangle$. The corresponding conditional state is

$$|\psi_s\rangle = \frac{P_s|\psi\rangle}{\langle \psi | P_s | \psi \rangle}.$$

One common case is that of several random variables $X_1, \ldots, X_n$. If they commute, then they have a common diagonalization, and they induce an orthogonal decomposition of $\mathcal{H}$ with $S = \mathbb{R}^n$. If two measurement operators X_1 and X_2 do not commute, then the set of states for which they are both definite does not span $\mathcal{H}$. As in Example 1.2.1, there is often no state for which X_1 and X_2 are both definite; they do not share an eigenvector. In words, two such variables are *mutually uncertain*; they are not *simultaneously measurable*.

Exercises

1. Verify that if X and Y are commuting Hermitian operators, then $X + Y$ and XY correspond, as measurements, to adding and multiplying the outcomes of the measurements X and Y.

2. Let $\mathcal{H} = \mathbb{C}^{\mathbb{Z}/n}$ be a state space whose basis is the cyclic group $\mathbb{Z}/n$. Define operators X and Z by

   $$X|k\rangle = |k+1\rangle \qquad Z|k\rangle = e^{2\pi i/n}|k\rangle$$

   Confirm that X has the same eigenvalues as Z. Find the eigenvalues of $X + Z$.

3. Show that if X is an anti-Hermitian operator, it represents an imaginary random variable; that if X is unitary, it represents a random variable with values in the unit circle $S^1 \subset \mathbb{C}$; and that if X commutes with its adjoint, it represents a complex random variable. In the last case, X is called a *normal* operator.

4. Suppose that $|\psi\rangle$ and $|\phi\rangle$ are two states in the same Hilbert space $\mathcal{H}$, and suppose that a physical object is in state $|\psi\rangle$. Show that the probability that it is in state $|\phi\rangle$ is

   $$|\langle \phi | \psi \rangle|^2.$$

5. Following Exercise 1.1.5, show that every orthonormal basis of the spin-$\frac{1}{2}$ Hilbert space consists of two spin states that point in opposite directions.

6. Show that a state $|\psi\rangle$ which is simultaneously definite for two Hermitian operators X and Y lies in the kernel of the commutator

$$[X, Y] = XY - YX.$$

Show that these states span $\ker[X, Y]$ when X and Y commute with $[X, Y]$, but not in general.

7. Show that if a measurement X is performed on a state ψ, its expectation (or average value) is given by:

$$E[X] = \langle\psi|X|\psi\rangle.$$

8. Suppose that X and Y are Hermitian operators on a Hilbert space $\mathcal{H}$ with a state ρ. Recall that if X is a classical random variable,

$$V[X] = E[X^2] - E[X]^2$$

denotes the variance of X. Prove the generalized Heisenberg uncertainty relation:

$$V[X]V[Y] \geq \frac{E[i[X, Y]]^2}{4}.$$

(Hint: After subtracting the means from X and Y, show that the 2×2 matrix

$$\begin{pmatrix} E[X^2] & E[XY] \\ E[YX] & E[Y^2] \end{pmatrix}$$

is positive semi-definite. The expectation formula in Exercise 1.2.7 is reasonable for arbitrary operators, not just normal ones.)

1.3. Joint states

Up until this point, a skeptic could still view quantum "probability" as kind of a cloud model and not really a modification of probability theory itself. If a configuration set A of a particle is a set of positions, then perhaps the particle is merely diffuse, like a cloud. Quantum superposition, measurement, and equivalence between different orthonormal bases are all surprising, but they are not quite show stoppers. The topic of this section, namely the correct model of joint quantum states, radically contradicts the cloud interpretation. (Section 1.6 has a more conclusive result in this direction.)

If A and B are finite configuration sets for two classical systems, then the configuration set for the joint system is the Cartesian product $A \times B$. Equivalently, the state space of the joint system is a tensor product:

$$\mathbb{R}^A \otimes \mathbb{R}^B \cong \mathbb{R}[A \times B].$$

This definition extends to the quantum case: If two quantum systems have state spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, then the joint system has state space $\mathcal{H}_A \otimes \mathcal{H}_B$. In particular if A and B are orthogonal bases of $\mathcal{H}_A$ and $\mathcal{H}_B$ (*i.e.*, configuration sets for Alice and Bob), then $A \times B$ is a joint basis, just as in the classical case. (But see Section 2.4.)

If a quantum object were somehow a cloud of amplitudes or probabilities, we would expect Alice and Bob to have independent states $|\psi_A\rangle$ and $|\psi_B\rangle$, at least if they were physically separated. When this happens, their joint state is $|\psi_A\rangle \otimes |\psi_B\rangle$; this is also called a *product* state. But most states are not product states; these states are called *entangled*. Entangled quantum states are evidently similar to correlated classical states.

Examples 1.3.1. Since a qubit has the configuration set $|0\rangle$ and $|1\rangle$, a system of n qubits has configuration set $\{0, 1\}^n$. Thus the general state for this system has 2^n amplitudes; for example the general three-qubit state is

$$|\psi\rangle = a_{000}|000\rangle + a_{001}|001\rangle + a_{010}|010\rangle + a_{011}|011\rangle + a_{100}|100\rangle + a_{101}|101\rangle + a_{110}|110\rangle + a_{111}|111\rangle.$$

It may look as if an n-qubit state carries an exponential amount of information, namely its 2^n amplitudes, but this is only true in a weak sense. With respect to a reasonable definition of information (see Exercise 1.4.5 and Section 1.8), a quantum superposition is not a record of its list of amplitudes, just as a hand of poker is not a record its $\binom{52}{5}$ probabilities.

One important product state on n qubits is the *constant* state:

$$|\psi\rangle = |++\ldots+\rangle = 2^{-n/2} \sum_{s \in \{0,1\}^n} |s\rangle.$$

One important entangled state is the *cat* state (as in "Schrödinger's cat"):

$$|\psi\rangle = \frac{|00\ldots0\rangle + |11\ldots1\rangle}{\sqrt{2}}.$$

As another example, an EPR pair (see Section 1.6.2) is a pair of electrons or other elementary particles in the entangled spin

$$|\psi\rangle = \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}}.$$

It is similar to the cat state with $n = 2$. In general any state one two qubits of the form

$$|\psi\rangle = \frac{|a, b\rangle + |c, d\rangle}{\sqrt{2}}$$

with

$$\langle a|c\rangle = \langle b|d\rangle = 0$$

is called a *Bell state* or a *Bell pair*.

Unitary transitions and Hermitian measurements on a joint system $|\psi_A\rangle \otimes |\psi_B\rangle$ which affect only Alice (respectively Bob) take the form $X \otimes I$ (respectively $I \otimes X$), where X is unitary or Hermitian. This is exactly analogous to the classical case. Such operations are also called *local* to Alice or Bob.

The combined model of unitary transitions, Hermitian measurements, and tensor products for joint states describes an isolated quantum object whose state is measured after a period of evolution. It is the standard description of quantum mechanics in many physics courses. It also describes a unitary quantum computer that is alternately manipulated and interrogated by a classical controller. But it also has shortcomings and omissions which confuse its interpretation, namely:

1. Hermitian measurements are missing from the unitary category **U**. In physical terms, the model does not include observers, even though observers can also be observed. (But note that a Boolean measurement P is subunitary; conditioning without normalization does lie in **U**'.)

2. Many physical objects, including typical observers, are effectively classical, even if they are *prima facie* quantum. These are also missing from the category **U**.

3. The category **U** is only weakly connected: there is no strictly unitary map from $\mathcal{H}_A$ to $\mathcal{H}_B$ when

   $$\dim \mathcal{H}_A > \dim \mathcal{H}_B.$$

   The subunitary category **U**' is strongly connected, but a subunitary map from $\mathcal{H}_A$ to $\mathcal{H}_B$ then includes extinction. In other words, in the category **U**', if Alice has more states than Bob, she cannot transfer her state to Bob without the possibility that the world ends.

4. There is no notion of marginals: If Alice and Bob are in an entangled state, there is no vector state for Alice alone. In particular, Alice can entangle with the environment ("Eve").

5. Even though measurements are a source of randomness, the category **U**cannot express classical randomness. For example, if the spin state of an electron is prepared by randomly choosing between $|\uparrow\rangle$ and $|\downarrow\rangle$, what is its state? The model of probability distributions on the manifold of vector states of an object is suspect, and in the end, redundant.

Sections 1.4 and 1.5 describe another model of quantum probability, the category **Quant**, that addresses most of these problems. Section 1.8 describes a final model, the category **vN**, that settles them more completely.

Exercises

1. Another description of the EPR state in Examples 1.3.1 is via measurement. Let $\mathcal{H}$ have the spin basis $|\uparrow\rangle$ and $|\downarrow\rangle$ and define the operators

   $$J_x^{\text{tot}} = J_x \otimes I + I \otimes J_x \qquad J_z^{\text{tot}} = J_z \otimes I + I \otimes J_z$$

   on $\mathcal{H} \otimes \mathcal{H}$, where J_x and J_z are defined as in Example 1.2.1. Show that J_x^{tot} and J_z^{tot} have one common eigenstate, for which both eigenvalues vanish.

2. Show that if $\mathcal{H}$ is a Hilbert space of dimension at least 2, there does not exist a linear map

   $$U : \mathcal{H} \to \mathcal{H} \otimes \mathcal{H}$$

   that takes every state $|\psi\rangle$ to a state equivalent to $|\psi\rangle \otimes |\psi\rangle$. (Recall that two states are equivalent if they differ by a global phase.) This is the simplest of a series of *no cloning* theorems for quantum states. A harder version: Show that such a map U is not even approximately linear.

3. Show if $|\psi\rangle$ and $|\phi\rangle$ are two Bell states shared by Alice and Bob, then there is a unitary operator local to Alice (*i.e.*, of the form $U \otimes I$) which takes $|\psi\rangle$ to $|\phi\rangle$:

   $$(U \otimes I)|\psi\rangle = |\phi\rangle.$$

   Thus all Bell states are equivalent.

4. Show that if $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is a vector state and

   $$\dim \mathcal{H}_A \leq \dim \mathcal{H}_B,$$

   then it has the form

   $$|\psi\rangle = \sum_{a \in A} \alpha_a |a\rangle \otimes |f(a)\rangle$$

for some orthonormal bases A and B and some function $f : A \to B$. This presentation is called a *Schmidt decomposition* of $|\psi\rangle$. Show that the unordered set of numbers $|\alpha_a|^2$ is uniquely determined by $|\psi\rangle$.

5. A common error in quantum probability is to mistake the direct sum $\mathcal{H}_A \oplus \mathcal{H}_B$ for the joint state space of Alice and Bob. Provide an empirical interpretation for direct sums which, among other properties, would also work in classical probability.

6. Show that the cat states from Examples 1.3.1 are entangled.

1.4. Operator states

Let $\mathcal{H}$ be the (finite-dimensional) Hilbert space of a quantum object, Alice. We define an *operator state* of Alice (or more simply a *state*) to be a positive semi-definite Hermitian operator ρ on $\mathcal{H}$. (Positive semi-definiteness is denoted $\rho \geq 0$.) The state ρ is *normalized* if $\mathrm{Tr}(\rho) = 1$ and *subnormalized* if $\mathrm{Tr}(\rho) \leq 1$. In this section and the next one we will define a new model (the category $\mathbf{Quant}_{<\infty}$) of quantum probability by replacing vector states with operator states, and by replacing unitary operators by a more complete class of quantum operations, analogous to stochastic maps.

Let $\mathcal{M}(\mathcal{H})$ be the vector space of all operators on $\mathcal{H}$. (Later we will abbreviate the algebra of $n \times n$ matrices $\mathcal{M}(\mathbb{C}^n)$ as $\mathcal{M}_n$.) The set

$$\mathcal{M}^{+,1}(\mathcal{H}) \subset \mathcal{M}(\mathcal{H})$$

of all normalized states is the *Bloch region* of Alice. Also let $\mathcal{M}^{+,\leq 1}(\mathcal{H})$ be the set of all subnormalized states and let $\mathcal{M}^{+}(\mathcal{H})$ be the set of all states.

Proposition 1.4.1. *If $\mathcal{H}$ is an n-dimensional Hilbert space then $\mathcal{M}^{+,1}(\mathcal{H})$ is a compact and convex set of real dimension $n^2 - 1$. Its extremal points are rank 1 operators: If ρ is extremal, it has the form*

$$\rho = |\psi\rangle\langle\psi|$$

for a unit vector $|\psi\rangle \in \mathbb{C}^n$.

The Bloch region $\mathcal{M}^{+,1}(\mathcal{H})$ is analogous to the classical simplex Δ_A of probability distributions on a finite set A. (Section 1.8 will discuss a mutual generalization.) First, the positivity and normalization conditions that define the two regions are both mathematically similar and have similar interpretations. If we choose an orthonormal basis A for $\mathcal{H}$, then an operator state ρ becomes a matrix; it can be written

$$\rho = \sum_{a,a' \in A} p_{a,a'} |a\rangle\langle a'|.$$

The diagonal entry $p_{a,a}$ is the probability of the configuration $|a\rangle$. Thus the positivity condition $\rho \geq 0$ asserts that the probability of any configuration (in any basis) is non-negative. The normalization condition asserts that the total probability in any basis is 1:

$$\mathrm{Tr}(\rho) = \sum_{a \in A} p_{a,a} = 1;$$

evidently this condition is basis-independent. Because the diagonal entries of ρ are probabilities, it is often called a *density matrix* or a *density operator* in physics.

The geometric features of $\mathcal{M}^{+,1}(\mathcal{H})$ and Δ_S are also similar, albeit with some important differences as well. Both regions are convex in order to allow classical superpositions. More precisely, if ρ_1 and ρ_2 are two states and $0 < p < 1$ is a probability, then the state

$$\rho = p\rho_1 + (1-p)\rho_2 \tag{1}$$

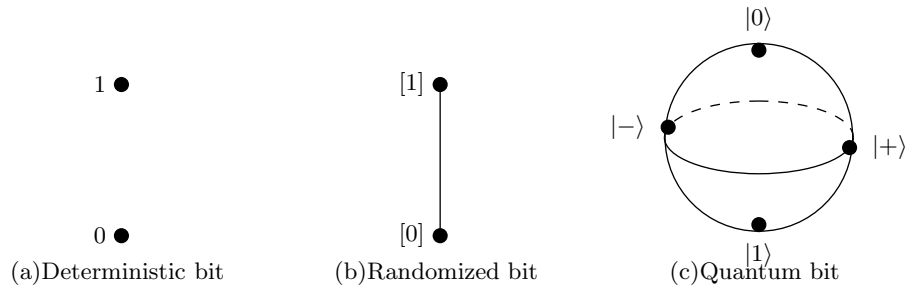
is a *classical superposition* or *mixture* of ρ_1 and ρ_2; it can be prepared by choosing randomly between them. If ρ is a mixture, *i.e.*if it is not an extremal point of $\mathcal{M}^{+,1}(\mathcal{H})$, then it is also called a *mixed state*.

If a state $\mu \in \Delta_S$ is extremal, then it is an element of S itself. It can then be called *definite* in the sense μ possesses no randomness: the probability of every event is either 0 or 1. If a state $\rho \in \mathcal{M}^{+,1}(\mathcal{H})$ is extremal, then it is called *pure*. By Proposition 1.4.1, pure states correspond to vector states, except that $|\psi\rangle\langle\psi|$ does not depend on the global phase of $|\psi\rangle$. As in Example 1.2.1, every state $\mathcal{M}^{+,1}(\mathcal{H})$ is a source of randomness; all states are indefinite.

Example 1.4.1. Our third and final understanding of a qubit is set of states is the Bloch region $\mathcal{M}^{+,1}(\mathcal{H})$. In this case $\mathcal{M}^{+,1}(\mathcal{H})$ is a round ball and is called the *Bloch sphere*, as shown in Figure 4. Two pure states are orthogonal if and only if they are antipodal as points on the Bloch sphere. The state in the middle,

$$\rho = \frac{|0\rangle + |1\rangle}{2},$$

is the *uniform state*; it is the equal mixture of any two orthonormal states.



(a)Deterministic bit (b)Randomized bit (c)Quantum bit

Figure 4: The space of states for three different types of bits.

Example 1.4.1 hints at a more general fact: Every mixed state is a mixture of pure states in many different ways. A mixed state encodes all of the statistical information that can be extracted by measurements and other operations. Thus a probability distribution on pure states is a highly redundant description of a mixed state.

Also following Example 1.4.1, if $\mathcal{H}$ has d states, then the state $\rho = I/d$, where I is the identity operator, is called the *uniform state*. It is the uniform mixture of all configurations in any orthonormal basis, hence a strong analogue of the uniform distribution on a finite set in classical probability.

All of the operations defined for vector states readily extend to operator states. If

$$U : \mathcal{H}_A \to \mathcal{H}_B$$

is a unitary or subunitary operator, its induced action on operators is given by

$$\mathcal{U}(\rho) = U\rho U^*.$$

If U is unitary, then

$$\mathcal{U}(\mathcal{M}^{+,1}(\mathcal{H}_1)) \subset \mathcal{M}^{+,1}(\mathcal{H}_2),$$

while if U is subunitary, then

$$\mathcal{U}(\mathcal{M}^{+,1}(\mathcal{H}_1)) \subset \mathcal{M}^{+,1}(\mathcal{H}_2).$$

If

$$X : \mathcal{H} \to \mathcal{H}$$

is a Hermitian measurement, then its expectation value is defined as

$$E_\rho[X] = \mathrm{Tr}(\rho X).$$

If $H = P$ is a Hermitian projection, *i.e.*a Boolean measurement, then the probability of P is defined as

$$P_\rho[P] = \mathrm{Tr}(\rho P)$$

and the conditional state is

$$\widehat{\rho|_P} = \frac{P\rho P}{\mathrm{Tr}(\rho P)}.$$

These rules for probabilities and conditional states also apply to set-valued measurements, using the projection P_s corresponding to an outcome $s \in S$.

Recall that if $\mathcal{H}_A$ and $\mathcal{H}_B$ are the Hilbert spaces of Alice and Bob, then their joint Hilbert space is $\mathcal{H}_A \otimes \mathcal{H}_B$. If Alice and Bob have independent states ρ_A and ρ_B, then their joint state is $\rho_A \otimes \rho_B$, a product state. General non-product joint states are a non-trivial mutual generalization of classical correlation and quantum entanglement, and their nomenclature reflects some of their surprising properties. A joint state ρ is called *separable* if it is a mixture of independent states. Non-independent separable states are roughly analogous to classical correlated states, but even these have some interesting quantum properties [5]. If ρ is not separable, then it is *entangled*. In some crucial respects, entanglement of mixed states is a weaker condition than entanglement of pure states. Current research is devoted to relating these two forms of entanglement.

As promised, there is a way to express marginals of joint states using operator states. If ρ is a joint state on $\mathcal{H}_A \otimes \mathcal{H}_B$, its marginal states on $\mathcal{H}_A$ and $\mathcal{H}_B$ are defined as partial traces:

$$\rho_A = \mathrm{Tr}_B(\rho) \qquad \rho_B = \mathrm{Tr}_A(\rho).$$

More explicitly, suppose that A and B are orthonormal bases of $\mathcal{H}_A$ and $\mathcal{H}_B$. Then

$$\rho_A = \mathrm{Tr}_B(\rho) = \sum_{a,a' \in A; b \in B} |a\rangle\langle a, b|\rho|a', b\rangle\langle a'|$$

$$\rho_B = \mathrm{Tr}_A(\rho) = \sum_{a \in A; b,b' \in B,} |b\rangle\langle a, b|\rho|a, b'\rangle\langle b'|.$$

Example 1.4.2. Suppose that Alice and Bob are qubits in the entangled vector state

$$|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

The operator form of this state is then

$$\rho = |\psi\rangle\langle\psi| = \frac{1}{2}\left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right).$$

Both of its marginals are the uniform state:

$$\rho_A = \rho_B = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Whereas in classical probability, a marginal of a definite state is definite, in quantum probability the marginal of a pure state need not be pure.

In general a linear map

$$\mathcal{E} : \mathcal{M}(\mathcal{H}_A) \to \mathcal{M}(\mathcal{H}_B)$$

is called a *superoperator*. If we interpret a unitary map (including both dimension-preserving operators and dimension-increasing embeddings) as a superoperator, and we have described a partial trace as another kind of superoperator. Both of these operations are empirical, and we can naively consider the category that they generate inside the category of all superoperators. For the moment we will call it **Quant**; in the next section we will show that it includes all maps of states that could reasonably be empirical.

Exercises

1. Verify that a local measurement $X \otimes I$ applied to a state ρ on a joint system $\mathcal{H}_A \otimes \mathcal{H}_B$ has the same probabilities as the measurement X applied to the marginal state $\mathrm{Tr}_B(\rho)$, and that the conditioned states are also consistent.

2. Prove Proposition 1.4.1.

3. Show, as Example 1.4.1 claims, that $\mathcal{M}_2^{+,1}$ is a round 3-dimensional ball and that pure states are orthonormal if and only if they are antipodal. Show that the probability of any Boolean measurement on a state ρ is proportional to the displacement of ρ from some hyperplane passing through the center of $\mathcal{M}_2^{+,1}$.

4. Show that every state $\rho \in \mathcal{M}_n^{+,1}$ is a convex combination of at most n pure states that have the same diagonal entries as ρ.

5. The *entropy* $S(\rho)$ of a state ρ is defined as $\mathrm{Tr}(\rho(\log \rho))$. Show that the uniform state I/d maximizes the entropy S on $\mathcal{M}_n^{+,1}$. Show that a state is pure if and only if it has no entropy.

6. Verify that a pure state conditioned on a measurement is still pure. More generally, show that measurement does not increase entropy: For any projection P and any state ρ,
$$S(P\rho P) \leq S(\rho).$$

7. Show that the each marginal of a pure joint state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is pure if and only if $|\psi\rangle$ is unentangled.

8. A *purification* of a state ρ on $\mathcal{H}$ is a pure state on a joint system $\mathcal{H} \otimes \mathcal{H}'$ whose left marginal is ρ. Show that every state on $\mathcal{H}$ has a purification in $\mathcal{H} \otimes \mathcal{H}$, and that it is unique up to a unitary operator local to the second factor.

9. The *support* of a state ρ on $\mathcal{H}$ is its image in $\mathcal{H}$ as a linear operator. Show that if ρ has full support, then every outcome of a projective measurement has non-zero probability.

10. Show that the uniform state on $\mathcal{H}$ is the only one which is invariant under all unitary operators on $\mathcal{H}$. Show, following Exercise 1.1.5, that the uniform spin-$\frac{1}{2}$ state is the only state that is not direction-dependent.

11. Show that every state in an open neighborhood of the uniform state on $\mathcal{H}_A \otimes \mathcal{H}_B$ is separable.

12. Given a joint Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, compute the dimension (in terms of the dimensions of $\mathcal{H}_A$ and $\mathcal{H}_B$) of:

    a) the space of all joint states on $\mathcal{H}_A \otimes \mathcal{H}_B$.

    b) the space of all joint pure states.

    c) the space of all product states.

    d) the space of all pure product states.

13. This exercise requires knowledge of some topology and differential geometry. Show that the space of pure states of an n-dimensional Hilbert space $\mathcal{H}$ is a $2n-2$-dimensional manifold, explicitly the manifold $\mathbb{C}P^{n-1}$. Show that the Riemannian metric that it inherits from its embedding in $\mathcal{M}(\mathcal{H})$ is two-point homogeneous, meaning that isometries act trasitively on unit tangent vectors. Show that the Riemannian metric (which is called the *Fubini-Study metric*) is positively curved.

1.5. Quantum operations

This section is mathematically more challenging than previous sections in Chapter 1. Our goal is to characterize all maps

$$\mathcal{E} : \mathcal{M}(\mathcal{H}_A) \to \mathcal{M}(\mathcal{H}_B)$$

that satisfy relatively weak conditions that we might want from empirical operations. A map that satisfies them will be called a "quantum operation". We will abbreviate $\mathcal{M}(\mathcal{H}_A)$ as $\mathcal{M}_A$ for Alice's operators, $\mathcal{M}_B$ for Bob's, etc.

By the classical superposition principle, an empirical map

$$\mathcal{E} : \mathcal{M}_A \to \mathcal{M}_B$$

should first be a linear map, *i.e.*, a superoperator. If $\mathcal{E}$ is linear, it is called *positive* if

$$\rho \geq 0 \implies \mathcal{E}(\rho) \geq 0$$

and *trace-preserving* if

$$\mathrm{Tr}(\rho) = 1 \implies \mathrm{Tr}(\mathcal{E}(\rho)) = 1.$$

Thus the condition that

$$\mathcal{E}(\mathcal{M}^{+,1}(\mathcal{H}_A)) \subset \mathcal{E}(\mathcal{M}^{+,1}(\mathcal{H}_B))$$

says that $\mathcal{E}$ is positive and trace-preserving or TPP. (Likewise $\mathcal{E}$ is positive if it preserves all states and positive, sub-trace-preserving or PSTP if it preserves subnormalized states.) By analogy with stochastic matrices, it is tempting to propose TPP maps as quantum operations. However, the tensor product of two TPP maps need not be positive, so the category of TPP maps is not compatible with joint states as they are defined in Section 1.3.

A map $\mathcal{E} : \mathcal{M}_A \to \mathcal{M}_B$ is called *completely positive* (CP) if for every quantum system C, the map

$$\mathcal{E} \otimes \mathcal{I} : \mathcal{M}_A \otimes \mathcal{M}_C \to \mathcal{M}_B \otimes \mathcal{M}_C$$

is positive, where $\mathcal{I}$ is the identity on $\mathcal{M}_C$.

Example 1.5.1. The transpose map $\mathcal{T} : \rho \mapsto \rho^T$ on $\mathcal{M}_n$ for $n \geq 2$ is positive but not completely positive.

Completely positive, trace-preserving (TPCP) maps do form a tensor category which for the moment we will call $\overline{\mathbf{Quant}}$. Every quantum operation should be TPCP; the category $\overline{\mathbf{Quant}}$ should contain the empirical class **Quant** of quantum operations. (If extinction is allowed, then every quantum operation should be STPCP.) In Section 1.4, we defined a category $\underline{\mathbf{Quant}}$ generated by unitary maps and partial traces; it should be contained in the empirical class **Quant**. The important result is that

$$\overline{\mathbf{Quant}} = \underline{\mathbf{Quant}},$$

which justifies either one as a definition of **Quant**. We can likewise define $\mathbf{Quant}'$ as the category of STPCP maps and $\mathbf{Quant}^+$ as the category of CP maps.

Theorem 1.5.1 (Stinespring,Kraus)**.** *Let*

$$\mathcal{E} : \mathcal{M}_A \to \mathcal{M}_B$$

be a superoperator. Then $\mathcal{E}$ is completely positive if and only if there exist operators

$$E_1, \ldots, E_N : \mathcal{H}_A \to \mathcal{H}_B$$

such that

$$\mathcal{E}(\rho) = \sum_{k=1}^{N} E_k \rho E_k^*. \tag{2}$$

Equivalently $\mathcal{E}$ factors as

$$\mathcal{M}_A \overset{\mathcal{U}}{\to} \mathcal{M}_B \otimes \mathcal{M}_C \overset{\mathrm{Tr}_C}{\to} \mathcal{M}_B,$$

where

$$\mathcal{D}(\rho) = D\rho D^*$$

and Tr_C is a partial trace.

The map $\mathcal{E}$ is trace-preserving if and only if

$$\sum_{k=1}^{N} E_k^* E_k = I \in \mathcal{M}_A, \tag{3}$$

in which case D is unitary.

Often Theorem 1.5.1 is called Stinespring's theorem [23]. Equation (2) is called the *operator-sum representation* or the *Kraus decomposition* [16]. The operation $\mathcal{D}$, or the corresponding operator D, is a *dilation* of the CP map $\mathcal{E}$.

Theorem 1.5.1 justifies the quantum superposition principle as a consequence of the classical superposition principle and complete positivity. These two assumptions alone imply that every quantum operation is a sum (or classical superposition) of subunitaries (which are quantum superpositions.) In this sense, the radical element of quantum probability is not quantum superposition itself, but rather replacing the simplex of states Δ_A with the Bloch region $\mathcal{M}^{+,1}(\mathcal{H})$.

This point of view is further supported by the following corollary. Say that a CP map is *coherent* if it is a single Kraus term. In particular, a unitary map is coherent.

Corollary 1.5.2. *If a CP map $\mathcal{E} : \mathcal{M}_A \to \mathcal{M}_B$ takes pure states to pure states, then either it is either coherent, or all states in its image are proportional. If $\mathcal{E}$ is invertible in the category **Quant**, then it is unitary.*

Note that in physics, an invertible process is usually called *reversible*.

We will prove these results at the end of this section; we first consider some particular classes of quantum operations.

A state ρ on a Hilbert space $\mathcal{H}$ can be interpreted as a quantum operation from the 1-state Hilbert space $\mathbb{C}$ to $\mathcal{H}$. In the other direction, the trace map

$$\mathrm{Tr} : \mathcal{M}(\mathcal{H}) \to \mathbb{C}$$

is also a quantum operation. These two operations can be thought of as *creation* and *destruction* of states. The composition $\rho \circ \mathrm{Tr}$ can be thought of as *initializing* an object in the state ρ. A partial trace

$$\mathrm{Tr}_A : \mathcal{M}(\mathcal{H}_A) \otimes \mathcal{M}(\mathcal{H}_B) \to \mathcal{M}(\mathcal{H}_B)$$

is also completely positive.

Suppose that an orthogonal decomposition

$$\mathcal{H} = \bigoplus_{s \in S} \mathcal{H}_s$$

represents a set-valued measurement. We noted in the previous section that the probability of the outcome s is given by

$$\mathrm{Tr}(\rho P_s)$$

and that the conditional state is

$$\widehat{\rho|_{P_s}} = \frac{P_s \rho P_s}{\mathrm{Tr}(\rho P_s)}.$$

If we imagine a hidden observer, Eve, performing this measurement, she will effect the operation

$$\mathcal{P}(\rho) = \sum_{s \in S} P_s \rho P_s \tag{4}$$

on the state ρ. This is evidently a quantum operation, one that expresses blind or hidden measurement. It has an explicit dilation

$$D : \mathcal{H} \to \mathcal{H} \otimes \mathbb{C}^S$$

given by the formula

$$D\psi = \oplus_{s \in S} P_s \psi \otimes |s\rangle.$$

We can interpret this dilation as a visible measurement, because the factor $\mathbb{C}^S$ could belong to Eve and does record the measurement outcome.

The main shortcoming of the dilation D as a model of of measurement is that Eve must possess quantum memory — she cannot be a classical computer or a human being. Section 1.8 discusses a better model with both quantum and classical objects. Nonetheless the model is very useful. An object can be measured by its environment; one electron or other particle can measure another one; a quantum computer can measure some of its qubits and place the outcome in other qubits; etc. Whenever two objects become entangled, we can say that each one is measuring the other. We can also say that decoherence is generally equivalent (by dilation) to entanglement with the environemtn. In the limit, one description of a non-quantum physical object is that it is a quantum object which is constantly being measured, or becoming entangled with, its environment.

Proof of Theorem 1.5.1. The proof here is based on a characterization of CP maps due to Jamiołkowski and Choi [7, 15, 22]. First, any superoperator

$$\mathcal{E} : \mathcal{M}_A \to \mathcal{M}_B$$

can be interpreted as an element

$$X_{\mathcal{E}} \in \mathcal{M}_A \otimes \mathcal{M}_B = \mathcal{M}(\mathcal{H}_A \otimes \mathcal{H}_B).$$

The point is that $\mathcal{E}$ is a tensor with four indices (Section 1.11), two for $\mathcal{H}_A$ and two for $\mathcal{H}_B$. In indices, its usual interpretation as a map is given by the expression:

$$\mathcal{E}(\rho)^b_{b'} = \mathcal{E}^{ba'}_{b'a} \rho^a_{a'}.$$

But we can also pair the indices differently as follows:

$$X_{\mathcal{E}}(\chi)^{a'}_{b'} = \mathcal{E}^{ba'}_{b'a} \chi^a_b.$$

Here $|\chi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. In reference to the alternate pairing of indices, we will call $X_{\mathcal{E}}$ the *sideways action* of $\mathcal{E}$.

We claim that $\mathcal{E}$ is completely positive as a superoperator if and only if $X_{\mathcal{E}} \geq 0$ as a Hermitian operator. This identification is known as the *Jamiołkowski criterion* or (in greater generality) the *Choi isomorphism*. We will rephrase the completely positivity condition to establish the logical equivalence. The map $\mathcal{E}$ is completely positive if and only if for any $\mathcal{M}_C$,

$$(\mathcal{E} \otimes \mathcal{I})(\rho) \geq 0$$

for all states $\rho \in \mathcal{M}_A \otimes \mathcal{M}_C$. The lemma that $\mathcal{E}$ (and therefore $\mathcal{E} \otimes \mathcal{I}$) preserves the Hermitian property of ρ if and only if $X_{\mathcal{E}}$ is Hermitian is left to Exercise 1.5.2. The more interesting positive semidefiniteness condition says that

$$\langle\psi|(\mathcal{E} \otimes \mathcal{I})(\rho)|\psi\rangle \geq 0$$

for all vectors $|\psi\rangle \in \mathcal{H}_B \otimes \mathcal{H}_C$. This numerical inequality is linear in ρ, so we may assume that ρ is

extremal, *i.e.*, pure. Thus by Proposition 1.4.1, complete positivity may be written more symmetrically as

$$\langle\psi|(\mathcal{E}\otimes\mathcal{I})(|\phi\rangle\langle\phi|)|\psi\rangle \geq 0$$

for all

$$|\psi\rangle \in \mathcal{H}_B\otimes\mathcal{H}_C \qquad |\phi\rangle\in\mathcal{H}_A\otimes\mathcal{H}_C.$$

In indices,

$$\psi^{bc}\psi_{b'c'}\mathcal{E}^{ab'}_{a'b}\phi^{a'c'}\phi_{ac}\geq 0.$$

If

$$\dim\mathcal{H}_C \geq \min(\dim\mathcal{H}_A,\dim\mathcal{H}_B),$$

then

$$\chi^a_b = \psi^{bc}\phi_{ac}$$

is an arbitrary vector in $\mathcal{H}_A\otimes\mathcal{H}_B$. With this abbreviation, complete positivity of $\mathcal{E}$ is the condition

$$\chi^{b'}_{a'}\mathcal{E}^{ab'}_{a'b}\chi^a_b \geq 0$$

for all χ. This is precisely positivity of $X_\mathcal{E}$.

The operator $X_\mathcal{E}$ is extremal among positive operators if and only if it has rank 1, *i.e.*,

$$X_\mathcal{E} = |E\rangle\langle E|$$

for some $E\in\mathcal{H}_A\otimes\mathcal{H}_B$. In indices,

$$\mathcal{E}^{ab'}_{a'b} = E^{a'}_{b'}E^b_a.$$

In operator form, this says that

$$\mathcal{E}(\rho) = E\rho E^*.$$

In other words, $\mathcal{E}$ is a single Kraus term if (and only if) it is extremal among CP maps. Therefore the general CP map is a sum of such terms.

The further assertions when $\mathcal{E}$ is trace-preserving are left to Exercise 1.5.3. □

Proof of Corollary 1.5.2. Let $\mathcal{D}$ be a dilation of $\mathcal{E}$, and let

$$D:\mathcal{H}_A\to\mathcal{H}_B\otimes\mathcal{H}_C$$

be its operator form. If $|\psi\rangle\in\mathcal{H}_B\otimes\mathcal{H}_C$ is a vector state, then its marginal

$$\mathrm{Tr}_C(|\psi\rangle\langle\psi|)$$

is pure if and only if $|\psi\rangle$ is a product state (Exercise 1.4.7):

$$|\psi\rangle = |\psi_B\rangle\otimes|\psi_C\rangle.$$

By hypothesis, every vector in the image of D must have this form. Now let

$$|\psi\rangle = |\psi_B\rangle\otimes|\psi_C\rangle \qquad |\psi'\rangle = |\psi'_B\rangle\otimes|\psi'_C\rangle$$

be two inequivalent states (*i.e.*, non-proportional vectors) in the image of D. If the sum $|\psi\rangle+|\psi'\rangle$ is also a product state, then either the left factors $|\psi_B\rangle$ and $|\psi'_B\rangle$ or the right factors $|\psi_C\rangle$ and $|\psi'_C\rangle$ are proportional — but not both, because then $|\psi\rangle$ and $|\psi'\rangle$ would be proportional. If this relationship holds for every inequivalent pair of states in $\mathrm{im}\,D$, then they must all have either the same left factor or the same right factor.

If all vectors in $\mathrm{im}\,D$ have the same left factor, respectively the same right factor, then

$$D|\psi\rangle = |\psi_B\rangle\otimes(E|\psi\rangle),$$

respectively

$$D|\psi\rangle = (E|\psi\rangle)\otimes|\psi_C\rangle,$$

for some linear map E. In the first case, states in the image of $\mathcal{E}=\mathrm{Tr}_C\circ\mathcal{D}$ are proportional to $|\psi_B\rangle\langle\psi_B|$. In the second case,

$$\mathcal{E} = \langle\psi_C|\psi_C\rangle\mathcal{D},$$

hence it is coherent.

If $\mathcal{E}$ is invertible, then it must send extremal points of $\mathcal{M}_A^{+,1}$ to $\mathcal{M}_B^{+,1}$. (This is generally true of any invertible map in the category of linear maps between convex bodies.) *I.e.*, it must send pure states to pure states. In this case E is unitary for two independent reasons: $\mathcal{E}$ is invertible, and $\mathcal{E}$ preserves trace. □

Exercises

1. Show directly from the definition of complete positivity that every state ρ on a Hilbert space $\mathcal{H}$ is $\mathcal{E}_\rho(1)$ for a completely positive map

   $$\mathcal{E}_\rho:\mathbb{C}\to\mathcal{M}(\mathcal{H}).$$

   Show that dilation of $\mathcal{E}$ is equivalent to purification of ρ.

2. Establish a missing step of Theorem 1.5.1: The map $\mathcal{E}$ commutes with the Hermitian adjoint operation if and only if $X_\mathcal{E}$ is Hermitian.

3. Establish the other missing step of Theorem 1.5.1: $\mathcal{E}$ is TPCP if and only if Equation (3) holds, if and only if D is unitary. Modify Equation (3) to the case when $\mathcal{E}$ is STPCP, and show that in this case D is subunitary.

4. Show that if
$$\mathcal{E} : \mathcal{M}_A \to \mathcal{M}_B$$
is STPCP, then there is an STPCP map
$$\mathcal{F} : \mathcal{M}_A \to \mathbb{C}$$
such that
$$\mathcal{E} \oplus \mathcal{F} : \mathcal{M}_A \to \mathcal{M}_B \oplus \mathbb{C}$$
is TPCP. Compare with Exercises 1.1.3 and **??**.

5. Find Kraus elements for a partial trace map
$$\mathrm{Tr}_B : \mathcal{M}_A \otimes \mathcal{M}_B \to \mathcal{M}_A.$$

6. Show that every blind measurement quantum operation (4) can be expressed as a convex combination of unitary quantum operations.

7. Show that the uniform state on $\mathcal{H}$ is sent to itself by every blind measurement quantum operation (4), and that it is the only state with this property.

8. A quantum operation $\mathcal{E}$ is *doubly stochastic* if and only if it is both trace-preserving and preserves the uniform state. For example, unitary quantum operations are doubly stochastic. Doubly stochastic quantum operations for a fixed Hilbert space $\mathcal{H}$ form a convex region, and unitary quantum operations are extremal points (check). Show that if $\dim \mathcal{H} = 2$, then all extremal doubly stochastic quantum operations are unitary, but that this is not true when $\dim \mathcal{H} > 2$. Compare with Exercise **??**.

1.6. Empiricism

1.6.1. Interpretation and evidence

Having defined the category **Quant** of quantum operations, we can now state its empirical interpretation:

1. State: Every observer in the universe can model external reality as a quantum system with a Hilbert space $\mathcal{H}$ that carries some particular state in the Bloch region $\mathcal{M}^{+,1}(\mathcal{H})$ at each point in time.

2. Independence: Reality decomposes into approximately disjoint subsystems whose joint Hilbert spaces are tensor products such as $\mathcal{H}_A \otimes \mathcal{H}_B$. An observer is an approximately independent subsystem whose residual non-independence is described by measurements such as Hermitian operators.

3. Evolution: After an observer performs a measurement, the new state of reality is given by projecting its state. More generally the state of reality evolves by quantum operations.

4. Statistics: An observer's experiences are interpreted as independently repeatable experiments. The probability of a measured value is the fraction of times that it occurs in repeated trials of the experiment.

This interpretation is exactly parallel to the one for classical probability theory at the end of Section 1.10. The first person to understand it clearly was Max Born in 1926 [6], an insight for which he eventually won the Nobel Prize. (Our presentation with mixed states is due to von Neumann and Hellwig-Kraus [13, 14, 16, 24].) It is intellectually healthy to have trouble accepting the Copenhagen interpretation. It is not healthy to reject it outright, even though this fate befell two disappointed parents of the interpretation, Einstein and Schrödinger. In this section we will discuss some of the overwhelming physical evidence for this interpretation, and a mathematical result in support of its radical nature.

First the evidence:

1. Quantum probability and quantum mechanics were originally developed to understand molecular, atomic, and subatomic structure and processes. It is a vast edifice that makes quantum probability truly irrefutable. For example, the structure of a hydrogen molecule is grossly arbitrary is understood in great detail This is in the same sense that an expert game of backgammon can be understood in great detail with classical probability, but seems grossly arbitrary without it.

2. A variety of real experiments and demonstrations match the thought experiments of quantum probability. This includes the examples in this article. For one, the two-slit "experiment" in Section 1.1 is a qualitatively correct model of laser speckle (scattering interference) and holography (photographic interference). Laser speckle is familiar as the twinkle in the dot of a laser pointer; see Figure 5. At the same time, light is composed of discrete, non-interacting photons. This is an unavoidable aspect of low-intensity X-ray photography, as shown in Figure 6.

   Photons are not the only particles that exhibit quantum superposition; in principle every physical object does. Figure 7 shows is an image of electrons obeying quantum superposition. Recently it has been demonstrated for C_{60} carbon molecules (buckyballs) [1].

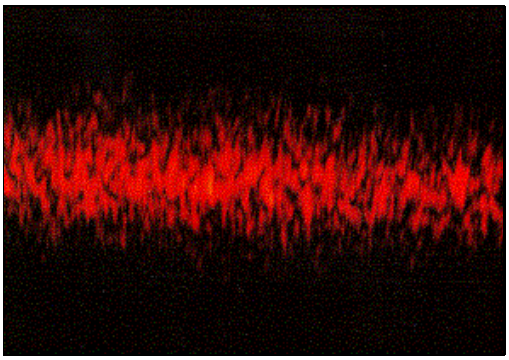


Figure 5: Laser speckle [25].

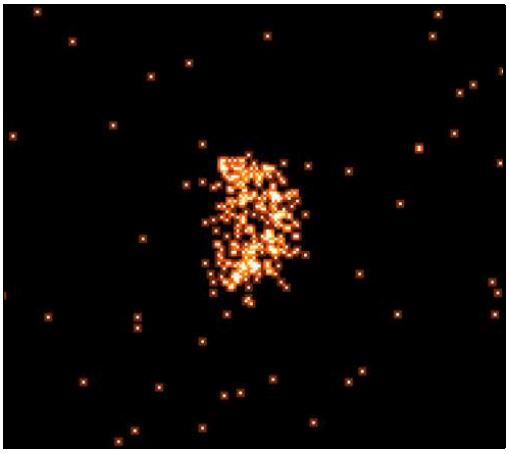


Figure 6: An X-ray image of Venus comprised of discrete photons (from the Chandra telescope) [9].

3. As mentioned in Section 1.3, and as discussed further in Section 1.6.2, the radical aspects of quantum probability require entangled joint states. Entanglement has also been demonstrated by a variety of experiments; see Exercise 1.6.3.

4. The known fundamental laws of physics are reversible, or in the quantum language, unitary. Unitary quantum probability does not encompass determinism or classical probability as a special case. Thus if the laws of physics are reversible and any physical objects are quantum, then the entire universe must be quantum. Among fundamental forces, the only one without a satisfactory quantum model is gravity[2].

[2] And string theory is a promising attempt at such a model.

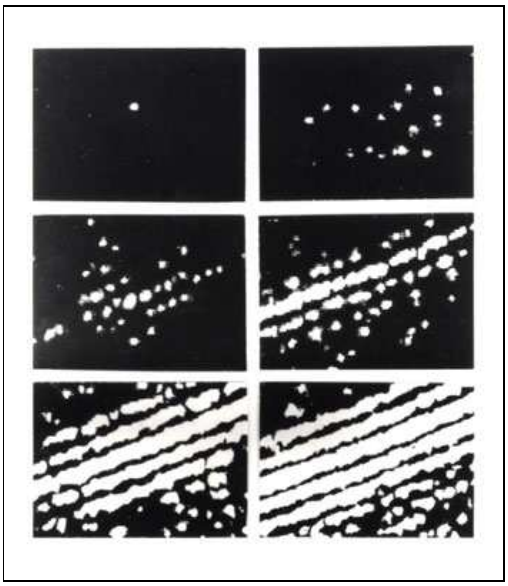


Figure 7: Quantum interference of individual electrons [19].

Even if quantum probability is irrefutable, is it necessary? Classical probability theory is (to some extent) unnecessary in the sense that it can be reproduced by hidden determinism. On the other hand, there is no reasonable reduction from quantum probability to classical probability or hidden determinism; see Section 1.6.2. Some entirely new theory could conceivably arise to "explain" quantum probability, but there is no reason to expect that a hypothetical successor would spare anyone from disbelief. Even irrefutable scientific facts can be refined; but if they are irrefutable, there is no turning back[3]. As it happens, the other known fundamental laws of physics do not modify quantum probability at all: relativity is geometric, while quantum field theory postulates specific physical forces.

If quantum probability is true and necessary, why is most macroscopic experience (on biological length scales and above) classical? In fact, length does not directly determine whether a physical system is described by classical or quantum probability. Rather, the system's relevant attribute is the number of accessible states. If the system has many states, then different evolutionary paths in the sense of Section 1.1 are likely to arrive at different final states, whence total probability is given by classical rather than quantum path summation. In quantum theory,

[3] For example, if you do not want to believe that the Earth orbits the sun, it does not help to learn that its orbit is an ellipse rather than a Copernican circle.

"microscopic" and "macroscopic" properly refer to amounts of entropy rather than to distances.

Another way to say it is that macroscopic objects typically evolve by highly decoherent quantum operations. They therefore constantly export entanglement to the environment. This is why the mixed-state model is useful for empirical interpretations. The macroscopic world consists of physical systems whose quantum state is strongly coupled to a common sea of thermal entropy, but which retain approximately independent classical states.

In particular the "paradox" of Schrödinger's cat, which Schrödinger offered as a criticism of the Copenhagen interpretation, is misleading. (But it is a useful antecedent of the notion of a cat state; see Examples 1.3.1.) The claim is that if a cat is at risk of death from a vial of poison that is controlled by a radioactive decay, then the cat is in a quantum superposition of life and death. But for thermal reasons, any room-temperature state of a cat is massively mixed, and typical superpositions are effectively classical. Such mixed states are also unaffected by typical blind measurement operations (Exercise 1.5.6). Only a frozen cat could be prepared in a pure state well enough to demonstrate non-commutativity of measurements.

Finally we caution against over-interpreting quantum probability. The best reason to believe or interpret anything in science is to understand it better. The basic statistical interpretation — the Copenhagen interpretation — is very helpful for understanding quantum mechanics and almost mandatory for understanding quantum computation. It is useful in the theory of operator algebras and potentially useful in some other areas of mathematics. One claimed alternative, the Everett "many worlds" interpretation, is narrowly relevant to path summation (Exercise 1.1.4). Another alternative, the Bohm interpretation, makes the narrow point that quantum probability can be viewed as a non-local deterministic system. (Non-local means that the model sacrifices any notion of independence in joint systems.) These alternative interpretations are not broadly useful.

1.6.2. Entanglement paradoxes

Einstein was a more inspired critic of the Copenhagen interpretation than Schrödinger. In a joint paper with Podolsky and Rosen [11], he noted that commuting measurements on an EPR pair,

$$|\psi\rangle = \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}},$$

possess classically implausible correlations. Their argument was sharpened by John Bell [3]. He established a simple inequality in classical probability, Bell's theorem, that is violated by quantum measurements on the EPR state. (Bell was also unsatisfied with the Copenhagen interpretation [4].)

We first give an informal description of Bell's theorem. Suppose that Alice and Bob are two suspects in prison together who are taken apart for separate questioning. In questioning, they are allowed to use notes and even electronic organizers, but they are not supposed to communicate by any means. Each of the suspects is given a sequence of questions (which may continue for several interrogation sessions). There are only three distinct questions, "X", "Y", and "Z", and only two answers, say "yes" and "no". The suspects are not expected to give consistent answers, but the authorities still hope to glean some information from the pattern of the answers. For simplicity, the questions are random and independent.

Suppose that the authorities notice that if the nth question posed to Alice and Bob is the same, they always give the same answer; but when the nth question posed is different, they only give the same answer $\frac{1}{4}$ of the time. Can they conclude that Alice and Bob are secretly communicating during the interrogations, or that they have advance access to the question lists, despite efforts to isolate them? If they are classical entities, then they must be cheating. If they always give the same answer when asked the same question in the nth round, then they must have prepared common answers lists to all three questions in advance. But if the nth question differs, then at least two of the three prepared answers are equal, if they are not all equal, so the probability of giving the same answer is at least $\frac{1}{3}$.

But if Alice's and Bob's electronic organizers can store entangled EPR pairs, then they can reduce the rate of agreement for distinct questions to $\frac{1}{4}$. It is convenient to re-express the EPR pair as the qubit cat state

$$|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

Alice and Bob can each answer one the three "questions" by performing the corresponding measurements

$$X = J_{2\pi/3} \qquad Y = J_{-2\pi/3} \qquad Z = J_0,$$

where

$$J_\theta = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}$$

Let

$$X_A = X \otimes I \qquad X_B = I \otimes X$$

be the corresponding factor measurements for Alice and Bob, and likewise for Y and Z. Then (Exercise 1.6.1):

1. Each of the six variables is an unbiased ± 1-valued random variable.
2. The variables X_A and X_B (and likewise for Y and Z) agree with probability 1.
3. The variables X_A and Y_B (and likewise the other pairs) agree with probability $\frac{1}{4}$.

(Note that each pair of questions converts an EPR pair to a product state; the EPR pair cannot be reused.) If the interrogators witness these classically impossible correlations, they might be tempted to seize Alice's and Bob's electronic devices and try to use them to communicate with each other. But they would not succeed, because no quantum operation on Alice's qubits affects the marginal state on Bob's qubits, or vice-versa.

More formally, Bell's theorem is an inequality concerning correlations of two-valued classical random variables which does not hold for quantum random variables:

Theorem 1.6.1 (Bell)**.** *If X, Y, and Z are three classical random variables taking values in $\{\pm 1\}$, then*

$$E[XY] + E[XZ] + E[YZ] \geq -1.$$

Proof. We would like to show, equivalently, that

$$E[XY + XZ + YZ] \geq -1.$$

It is easy to check that

$$XY + XZ + YZ = \begin{cases} 3 & \text{if } X = Y = Z \\ -1 & \text{otherwise} \end{cases}.$$

Since the random variable $XY + XZ + YZ$ is always at least -1, its expectation is at least -1. □

A variant of the Bell-EPR paradox (with a different set of classically impossible correlations) was famously demonstrated in an experiment by Aspect et al [2], and since then by others. In the experiment the two halves of Bell-state photon pairs were interrogated at almost simultaneously, so that there was not enough time for a message to travel from one photon to the other. These experiments should not be taken as self-contained proof of that quantum probability is true, because they have possible "loopholes" that could allow the photons to communicate. At the same time, there is no evidence of any genuine interaction between the photons in these demonstrations, much less non-quantum interactions that would present an illusion of quantum non-interaction.

The original purpose of the Bell-EPR paradox was the simple conclusion that quantum operations do not admit a deterministic or classically random simulation that preserves locality. In hindsight, it is a first step in the direction of quantum algorithms (Section **??**) and especially quantum security (Section **??**), since these can also be viewed as entanglement paradoxes. The problem of communication security is for two parties (Alice and Bob) to share information with some confidence that there is no eavesdropper (Eve). The shared information is ideally random, because it can then be used to mask arbitrary messages. In Bell's protocol, the same argument that Alice's and Bob's answers are classically impossible also shows that there cannot be an Eve who knows their answers in advance. Thus the shared answers are also shared secrets.

The relation to quantum algorithms is less formal. Intuitively, quantum algorithms exploit entanglement as a kind of communication. For example, the result of Grover's search algorithm (Section **??**) can be described as a guessing game: If Alice thinks of a number from 1 to N and only responds "yes" or "no" depending one whether Bob guesses correctly, then Bob can guess it with $O(\sqrt{N})$ guesses, provided that he can guess in quantum superposition and Alice's consideration of each guess is unitary. Grover's algorithm is a classically impossible form of communication afforded by quantum entanglement.

Exercises

1. Establish that Bell's operators X_A, Y_A, Z_A, X_B, Y_B, and Z_B applied to a Bell state violate Bell's inequality.
2. The quantum violation of Bell's theorem can be called a "no hidden variables" theorem: Quantum operations cannot be simulated by hidden structure which is deterministic or classically random. One rigorous (but possibly limited) interpretation of this principle can be phrased as category theory: There does not exist a non-trivial linear tensor functor from the category $\mathbf{Quant}_{<\inf}$ to the category **Prob**. Prove this result using Bell's theorem and measurements of EPR pairs.
3. The Aspect experiment employs the inequality

$$E[X_A X_B] + E[X_A Y_B] + E[Y_A Y_B] - E[Y_A X_B] \leq 2$$

for ±1-valued classical random variables, due to Clauser, Horne, Shimony, and Holt [8]. (This avoids the assumption in Bell's theorem that when Alice and Bob perform the same measurement, they will agree with probability one.) Prove this inequality, and then find a violation using the operators J_θ for four particular values of θ.

1.7. Infinite systems

The immediate way to extend the finite-state theory to infinite quantum systems is to allow the Hilbert space $\mathcal{H}$ to be infinite-dimensional (but usually separable). Section 1.8 discusses a better and more general extension due to von Neumann, but much can be learned from the this less creative approach.

We can use various definitions from operator theory [**?**] to adapt various objects such as states, random variables, and quantum operations to infinite Hilbert spaces. Once these are defined, we can define the category **Quant** to be the category of Hilbert spaces (both finite and infinite) with TPCP maps as the morphisms or quantum operations.

First, a (normal) state ρ is defined as a positive semi-definite trace-class operator with trace 1. In other words, the Bloch region $\mathcal{B}^{+,1}(\mathcal{H})$ is defined as the trace 1 subspace of $\mathcal{B}_t(\mathcal{H})$, the algebra of trace-class operators. The spectral theorem for compact operators implies that such a state ρ can be expressed as:

$$\rho = \sum_{s\in S} p_s |s\rangle\langle s|$$

for some orthonormal basis S of $\mathcal{H}$. Thus as in the finite case, pure states (by definition the extremal elements of the Bloch region $\mathcal{M}^{+,1}(\mathcal{H})$) correspond to vector states (by definition unit vectors in $\mathcal{H}$) up to a global phase.

A real-valued, bounded random variable X on $\mathcal{H}$ is defined as a self-adjoint bounded operator. This matches the definition of states in that $\mathcal{B}(\mathcal{H})$, the algebra of bounded operators, is the Banach space dual of $\mathcal{B}_t(\mathcal{H})$. This duality means that for any state ρ and any bounded variable X, the trace $\mathrm{Tr}(\rho X)$ is well-defined as a finite real number. Thus we can define the expectation

$$E_\rho[X] \stackrel{\text{def}}{=} \mathrm{Tr}(\rho X)$$

as before. More generally, a state ρ and a real-valued random variable X produce a probability measure on $\mathbb{R}$, the *distribution* of X, by the spectral theorem for bounded operators. This theorem expresses X as an integral with respect to an operator-valued measure μ_P whose value on any interval is a projection that commutes with X:

$$X = \int_{\mathbb{R}} \lambda d\mu_P. \tag{5}$$

If we pair the measure μ_P with the state ρ, the result is the desired scalar-valued measure on $\mathbb{R}$, indeed on the spectrum of X.

A (projective) measurement X is again defined as a direct sum decomposition

$$\mathcal{H} \cong \bigoplus_{s\in S} \mathcal{H}_s$$

for some outcome set S, which may now be infinite. Probabilities and conditional states have the same formulas:

$$P_\rho[X = s] = \mathrm{Tr}(P_s\rho) \qquad \widehat{\rho|_{X=s}} = \frac{P_s\rho P_s}{P[X=s]}.$$

Not every real-valued random variable defines a measurement of this type. Rather, the spectral theorem says that a Hermitian operator X has a point spectrum and a continuous spectrum. Only the point spectrum possesses eigenspaces, so X must have a pure-point spectrum in order to define a measurement. However, there are various ways to approximately measure a continuous-spectrum values of an operator. The point spectrum is usually discrete, meaning that the eigenvalues are isolated, while the continuous spectrum usually consists of intervals, which in quantum mechanics are called *bands*. But there are other possibilities for both parts of the spectrum.

An unbounded random variable is defined as a self-adjoint unbounded operator, although such an operator is an artifice in the sense that it is only defined on a dense subset of $\mathcal{H}$. By definition it is a densely defined function whose graph is a closed vector subspace of $\mathcal{H} \oplus \mathcal{H}$ which is invariant under switching the two summands. The definition is chosen so that self-adjoint operators satisfy the spectral theorem. If X is a self-adjoint operator, then

$$U(t) = e^{itX}$$

is a one-parameter group of unitary operators, and conversely every strongly continuous one-parameter group of unitary operators defines a possibly unbounded operator. Either the spectral theorem or the unitary operator model could be taken as an alternate definition of an unbounded self-adjoint operator.

Example 1.7.1. The function spaces $L^2(\mathbb{R}^d)$, with $1 \le d \le 3$, are very common in quantum mechanics. Pure states are naturally referred to as wave or

amplitude functions. Technically they are half densities, meaning that the square norm of a wave function is a probability density function with a volume form factor. For example, the wave function

$$\psi(x) = \frac{e^{-x^2/2}}{\pi^{1/4}}\sqrt{dx},$$

here written as a half density, is called a *coherent state* on $\mathcal{H} = L^2(\mathbb{R})$ (notwithstanding that elsewhere every pure state is called coherent). It will appear later as the ground state of the harmonic oscillator.

To give an example of a mixed state on $L^2(\mathbb{R})$, let $a \geq 1$ and let

$$\rho(x,y) = \frac{1}{\sqrt{\pi a}} e^{a(x-y)^2 + a^{-1}(x+y)^2)/4}$$

be a kernel (in the sense of integration, not null spaces). The corresponding operator

$$\rho(f)(x) = \int_{\mathbb{R}} \rho(x,y) f(y) dy$$

is trace-class with trace 1 and is called a *quasifree state*; when $a = 1$ it is the coherent state.

Example 1.7.2. If $f(x)$ is a continuous (or even integrable) function on $\mathbb{R}$, then multiplication by f is an operator on $\mathcal{H}$ which is given the same name. For example, x is an unbounded, continuous-spectrum operator whose distribution with respect to the pure state $\psi(x)$ has the density function $|\psi(x)|^2$. Another example is the operator

$$p = -i\frac{\partial}{\partial x}$$

which has the same spectrum as x, namely all of $\mathbb{R}$ as a continuous spectrum. They are both Gaussian random variables with respect to coherent and quasi-free states. We will see later that the operator

$$H = \frac{p^2 + x^2}{2}$$

has discrete spectrum $\mathbb{Z}_{\geq 0} + \frac{1}{2}$. In the standard coherent state, H is definite with value $\frac{1}{2}$, while in a standard quasi-free state, it has a discrete exponential distribution.

Exercises

1.8. Operator algebras

Following von Neumann, we can represent a quantum object not as a Hilbert space $\mathcal{H}$, but as an abstract algebra $\mathcal{A}$ whose elements can be called "operators". Such an operator algebra should satisfy suitable axioms so that we can define states, random variables, and quantum operations. Von Neumann defined two types of algebras for this purpose, C^*-algebras and W^*-algebras; the latter are now called *von Neumann algebras*. Von Neumann algebras are actually just C^*-algebras with a stronger topological closure property.

The algebras $\mathcal{M}(\mathcal{H})$ of all bounded operators on a Hilbert spaces $\mathcal{H}$ are one class of von Neumann algebras that happen to contain all von Neumann algebras as subalgebras. But considering only $\mathcal{M}(\mathcal{H})$ is a very restricted view of the theory of operator algebras, just as considering only symmetric groups is a very restricted view of finite group theory. Quantum physics has also drifted towards considering specific algebras of operators, although not usually with von Neumann's axioms.

A C^*-algebra $\mathcal{A}$ is, first, a complex vector space with an associative and bilinear multiplication law. It also has an abstract anti-linear, product-reversing adjoint operation denoted "$*$":

$$(X+Y)^* = X^* + Y^* \qquad (\lambda XY)^* = \overline{\lambda} Y^* X^*.$$

Finally $\mathcal{A}$ is also a Banach space a norm $||\cdot||$ that satisfies the relation

$$||X^* X|| = ||X||^2.$$

This last axiom, the "C^* axiom" is coy and has many consequences for the structure of $\mathcal{A}$. Among other things, it means that the norm $||\cdot||$ is completely determined by the algebra structure of $\mathcal{A}$ and that

$$||X^*|| = ||X||.$$

Intuitively $\mathcal{A}$ consists of bounded operators and $||X||$ behaves as the spectral radius of X. For simplicity we will assume that every C^*-algebra $\mathcal{A}$ has a unit, even though non-unital C^*-algebras are also an interesting class.

Possessing a unit is traditionally an optional axiom for C^*-algebras; we will assume it for simplicity.

Theorem 1.8.1 (Gelfand,Naimark)**.** *If $\mathcal{A}$ is a (unital) commutative C^*-algebra, then it is isomorphic to an algebra of continuous functions $C(A)$ on a compact Hausdorff topological space A.*

By Theorem 1.8.1, and since $C(A)$ is a C^*-algebra for every compact Hausdorff space A, a C^*-algebra can be thought of as a "non-commutative topological space". In particular if A is a finite set, then $C(A) = \mathbb{C}^A$ is exactly the model of finite probability described in Section 1.10 — its set of normalized states is Δ_A.

Theorem 1.8.1 also implies that if $X \in \mathcal{A}_{\text{sa}}$ (the self-adjoint subspace of $\mathcal{A}$) and $f : \mathbb{R} \to \mathbb{R}$ is a continuous function, then there is a well-defined element

$f(X) \in \mathcal{A}_s a$ (Exercise **??**). For example, $\sin X$ and $|X|$ are well-defined. We will need a slight generalization of this principle: An element $X \in \mathcal{A}_{\text{sa}}$ is *positive*, or $X \geq 0$, if $X = Y^*Y$ for some Y. If $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is continuous and $X \geq 0$, then $f(X)$ is well-defined; if $f \geq 0$ as a function, then $f(X) \geq 0$.

A *representation* of a C^*-algebra $\mathcal{A}$ is a homomorphism from $\mathcal{A}$ to the C^*-algebra $\mathcal{B}(\mathcal{H})$ of bounded operators on a Hilbert space. (A homomorphism between two C^*-algebras is a linear map that respects multiplication, $*$, and is continuous with respect to the Banach norm.) Crucially, the algebra $\mathcal{B}(\mathcal{H})$ has other topologies besides the one coming from its Banach norm, namely the strong and weak operator topologies. If $\mathcal{M}$ is a C^*-algebra which is closed with respect to the weak operator topology in some faithful representation $\mathcal{H}$, then $\mathcal{M}$ is a *von Neumann algebra*.

Theorem 1.8.2. *If $\mathcal{M}$ is a commutative von Neumann algebra, then it is isomorphic to the algebra $L^\infty(M)$ for some σ-field M.*

By Theorem 1.8.2, and since $L^\infty(M)$ is a von Neumann algebra for many natural σ-fields M, a von Neumann algebra $\mathcal{M}$ can be thought of as a "noncommutative measure space".

Theorem 1.8.2 also implies that if $X \in \mathcal{M}_{\text{sa}}$ (the self-adjoint subspace of $\mathcal{A}$) and $f : \mathbb{R} \to \mathbb{R}$ is a measurable function, then there is a well-defined element $f(X) \in \mathcal{M}_s a$. This closure property is called "functional calculus".

The traditional definition of von Neumann algebra via a faithful action on a Hilbert space $\mathcal{H}$ is contrary to our intention of emphasizing operators over vectors. Happily there are other characterizations of von Neumann algebras within the class of C^*-algebras:

Theorem 1.8.3 (???)**.** *A C^* algebra $\mathcal{M}$ is a von Neumann algebra if and only if it has a pre-dual ${}^\#\mathcal{M}$ as a Banach space. The pre-dual, if it exists, is unique up to isometry.*

In particular, the algebra $\mathcal{B}(\mathcal{H})$ of bounded operators on a Hilbert space $\mathcal{H}$ is a von Neumann algebra with pre-dual $\mathcal{B}_t(\mathcal{H})$. Theorem 1.8.3 interplays with the general fact that every Banach space $\mathcal{B}$ embeds isometrically in its second dual $\mathcal{B}^{\#\#}$. Thus the pre-dual ${}^\#\mathcal{M}$ can be viewed as subspace of the dual $\mathcal{M}^\#$. Also, by a construction of ???, if $\mathcal{A}$ is a C^*-algebra, its second dual $\mathcal{A}^{\#\#}$ has the natural structure of a von Neumann algebra, the *universal enveloping von Neumann algebra* of $\mathcal{A}$.

If $\mathcal{A}$ and $\mathcal{B}$ are two C^* algebras, there is a natural tensor product C^*-algebra $\mathcal{A} \otimes \mathcal{B}$ which is a topological completion of the algebraic tensor product. If $\mathcal{M}$ and $\mathcal{N}$ are von Neumann algebras, there a von Neumann algebra $\mathcal{M} \otimes \mathcal{N}$ which is a further topological completion than the C^*-algebra completion.

After accepting these preliminaries (perhaps on faith), we can proceed to define the basic structures of quantum probability. A C^*-algebra $\mathcal{A}$ or a von Neumann algebra $\mathcal{M}$ can be assigned. The algebra is termed its *algebra of observables*, because the self-adjoint elements ($\mathcal{A}_{\text{sa}}$ or $\mathcal{M}_{\text{sa}}$) will be interpreted as real-valued random variables. If $\mathcal{A}$ is a C^*-algebra, a *state* is a dual vector $\rho \in \mathcal{A}^\#$ which is *positive*, meaning that

$$X \geq 0 \implies \rho(X) \geq 0.$$

The state ρ is *normalized* if

$$\rho(I) = 1.$$

Previously we took ρ to be an operator rather than a dual vector; this is not really different if we define

$$\rho(X) = \text{Tr}(\rho X).$$

In particular we can call $\rho(I)$ the "trace" of ρ.

A state ρ on a von Neumann algebra $\mathcal{M}$ is *normal* if it lies in the pre-dual ${}^\#\mathcal{M}$. The commutative case illustrates the reason to take states from the dual of a C^*-algebra but from the pre-dual of a von Neumann algebra. If $\mathcal{A} = C(A)$, then by the ??? theorem, states are equivalent to finite Borel measures on A. If $\mathcal{M} = L^\infty(M)$, then general states are equivalent to finitely additive, finite measures on M; normal states are equivalent to countably additive measures and are hence more empirical. Note also that the states of the C^*-algebra $\mathcal{A}$ are the normal states of the von Neumann algebra $\mathcal{A}^{\#\#}$.

A pair $(\mathcal{M}, \rho)$ consisting of a von Neumann algebra $\mathcal{M}$ and a normalized, normal state ρ is also called a *quantum probability space* or a *noncommutative probability space*. Each random variable $X \in \mathcal{M}_{\text{sa}}$ has a well-defined spectrum $\text{Spec}\, X \subset \mathbb{R}$ (which depends crucially on the structure of $\mathcal{M}$), and each state ρ induces a probability distribution on $\text{Spec}\, X$. If $\mathcal{M} \subseteq \mathcal{B}(\mathcal{H})$ is defined as an algebra of operators on a Hilbert space, one way to construct the spectrum and distribution of X is by Equation (5). The point is that ${}^\#\mathcal{M}$ is a quotient of $\mathcal{B}_t(\mathcal{H})$, and we can use any lift of ρ to a trace-class operator on $\mathcal{H}$.

As usual, if $\mathcal{A}$ and $\mathcal{B}$ are Alice's and Bob's algebras of observables, their joint algebra of observables is $\mathcal{A} \otimes \mathcal{B}$.

If $\mathcal{A}$ and $\mathcal{B}$ are C^*-algebras, then a linear map

$$\mathcal{E} : \mathcal{B} \to \mathcal{A}$$

is *positive* if it takes positive elements of $\mathcal{A}$ to positive elements of $\mathcal{B}$; it is *completely positive* if

$$\mathcal{E} \otimes \mathcal{I} : \mathcal{B} \otimes \mathcal{C} \to \mathcal{A} \otimes \mathcal{C}$$

is positive for every C^*-algebra $\mathcal{C}$; and it is *unital* if

$$\mathcal{E}(I) = I.$$

If $\mathcal{M}$ and $\mathcal{N}$ are von Neumann algebras, a linear map

$$\mathcal{E} : \mathcal{N} \to \mathcal{M}$$

is *normal* if it has a pre-transpose

$$^{\#}\mathcal{E} : {}^{\#}\mathcal{M} \to {}^{\#}\mathcal{N}.$$

In general $\mathcal{E}$ is unital if and only if the pre-transpose $^{\#}\mathcal{E}$ (or the transpose $\mathcal{E}^{\#}$) is trace-preserving. The category **vN** is defined as the category of pre-duals of von Neumann algebras with TPCP maps as its morphisms. Equivalently it is the category of von Neumann algebras with normal UCP (unital and CP) maps as contravariant morphisms (Exercise **??**). The category **vN** is the most satisfactory model of infinite quantum probability, and its morphisms can be called quantum operations. An interesting also-ran is the category $\mathbf{C}^*$ of C^* algebras with arbitrary UCP maps as contravariant morphisms. Although **vN** can be viewed as a subcategory of $\mathbf{C}^*$, $\mathbf{C}^*$ can also be viewed as a subcategory of **vN**, by taking each C^*-algebra $\mathcal{A}$ to its enveloping algebra $\mathcal{A}^{\#\#}$.

1.9. Classical and quantum coexistence

Because the category **vN** includes commutative algebras, which are the corresponding models of classical probability, they can model coexistence of classical and quantum objects and interactions between them. We can then use these categories to study measurements of quantum systems by classical observers, and classical behavior of quantum systems. (As discussed in Section 1.6.2, the converse of the latter is impossible.)

As a first case, suppose that Alice is classical and finite, so that her von Neumann algebra is

$$\mathcal{A} = \mathbb{C}^A$$

for some finite set A. Suppose that Bob has some other von Neumann algebra $\mathcal{B}$. Then we can axiomatically define a *destructive measurement* to be an arbitrary quantum operation

$$\mathcal{E} : {}^{\#}\mathcal{B} \to {}^{\#}\mathcal{A}.$$

Then (Exercise **??**) the general form of $\mathcal{P}$ is

$$\mathcal{E}(\rho) = \bigoplus_{a \in A} \rho(E_a),$$

where each $E_a \geq 0$ and

$$\sum_{a \in A} E_a = I.$$

This structure is also known concretely as a *positive, operator-valued measure*, or *POVM*, because it is a probability measure on A that takes values in the positive cone $\mathcal{B}^+$ rather than in the real numbers.

We can construct a sensible conditional state on $\mathcal{B}$ in the same setting. It would be given by a quantum operation

$$\mathcal{F} : {}^{\#}\mathcal{B} \to {}^{\#}\mathcal{B} \otimes {}^{\#}\mathcal{A}$$

such that the composition $\mathrm{Tr}_B \circ \mathcal{F}$ (in which Alice applies $\mathcal{F}$ and then destroys Bob) is a POVM

$$\mathcal{E} : {}^{\#}\mathcal{B} \to {}^{\#}\mathcal{A}.$$

We further suppose that $\mathcal{F}$ is *initial* among all such factors of $\mathcal{E}$, in the sense that Bob retains as much information as possible about his previous state. Then (Exercise **??**) one form for $\mathcal{F}$ is

$$\mathcal{F} = \bigoplus_{a \in A} [a] \otimes \rho^{\sqrt{E_a}},$$

where in general the state ρ^X is defined as

$$\rho^X(Y) = \rho(X^* Y X);$$

also if $X \geq 0$, $\sqrt{X}$ is its unique positive square root. In light of this possible structure for $\mathcal{F}$, the *conditional state* of the POVM $\mathcal{E}$ applied to ρ with outcome a is

$$\rho|_{\mathcal{E}=a} = \rho^{\sqrt{E_a}}.$$

Finally the POVM $\mathcal{E}$ has mutually exclusive outcomes if and only if its non-destructive lift $\mathcal{F}$ is a projection, *i.e.*, $\mathcal{F}^2 = \mathcal{F}$. In this case (Exercise **??**) the E_as are projections such that

$$E_a E_{a'} = 0$$

when $a \neq a'$. Thus the projective measurements defined in Sections **??** and **??** are exactly those POVMs with mutually exclusive outcomes.

The above discussion can be extended to the case where Alice is classical but infinite. In this case her von Neumann algebra is $L^\infty(A)$ for some σ-field A, and a POVM is a measure on A that takes values in $\mathcal{B}^+$. This structure is closely related to a normalized measure on $\mathcal{B}^+$ itself.

We turn to a result from the theory of operator algebras that shows how decoherence can reduce a quantum object to a classical one. Suppose that $\mathcal{A}$

is Alice's algebra of observables and that she evolves according to a quantum operation

$$\mathcal{P} : \mathcal{A} \to \mathcal{A}$$

in a unit period of time. Very often we can suppose that $\mathcal{P}$ is an idempotent, *i.e.*, $\mathcal{P}^2 = \mathcal{P}$ (Exercise **??**). Intuitively this means that Alice stabilizes in a short period of time and, once stable, her temporal evolution does not further affect her state.

Theorem 1.9.1 (Choi-Effros)**.** *If $\mathcal{A}$ is a C^*-algebra and $\mathcal{P}$ is an SUCP idempotent on $\mathcal{A}$, then the image of $\mathcal{P}$ is a C^*-algebra $\mathcal{B}$ with a modified product*

$$X \circ Y = \mathcal{P}(XY).$$

Theorem 1.9.1 says that if Alice's evolution is an idempotent $\mathcal{P}$, then she can be modelled by a smaller effective C^*-algebra $\mathcal{B}$. In particular, $\mathcal{B}$ can be commutative even when $\mathcal{A}$ is not. The algebra $\mathcal{B}$ is a vector subspace of $\mathcal{A}$, but not in general a subalgebra. Rather, the modified product defined by Choi and Effros matches the process of applying $\mathcal{P}$ between external interactions with Alice. Note also that if $\mathcal{A}$ is a von Neumann algebra and if $\mathcal{P}$ is normal, then $\mathcal{B}$ is also a von Neumann algebra.

Example 1.9.1. The blind projective measurement operation $\mathcal{P}$ defined in Equation (4) is an idempotent. In this case $\operatorname{im}\mathcal{P}$ is closed under multiplication and is the algebra

$$\bigoplus_{s\in S} \mathcal{M}(\mathcal{H}_s).$$

Finally, the classification of finite-dimensional C^*-algebras, or von Neumann algebras, also fits the theme of classical and quantum coexistence. Say that a $*$-algebra $\mathcal{A}$ is *positive* if

$$X^*X = 0 \implies X = 0.$$

If $\mathcal{A}$ is positive and finite-dimensional, then it is automatically a C^*-algebra, indeed a von Neumann algebra. This can be seen from the following classification theorem.

Theorem 1.9.2 (Artin-Schreier)**.** *If $\mathcal{A}$ a positive, finite-dimensional $*$-algebra, then it is a direct sum of matrix algebras:*

$$\mathcal{A} \cong \bigoplus_{k=1}^{n} \mathcal{M}_{\lambda_k}$$

for some integers

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

The vector $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ in Theorem 1.9.2 can be called the *shape* of the algebra $\mathcal{A}$.

Examples 1.9.2. A joint system consisting of a finite classical Alice with algebra $\mathbb{C}^n$ and a finite quantum Bob with algebra $\mathcal{M}_k$ has the rectangular shape $(k, k, \ldots, k)$. For example, Bob could be a quantum computer with qubit memory and Alice could be a classical controllers with classical bit memory. The first finite quantum system which is not of this form is the hybrid trit $\mathcal{M}_2 \oplus \mathbb{C}$. The author [**?**] has analyzed the storage properties of finite quantum systems such as the hybrid trit.

Exercises

1.10. Appendix: A classical review

Consider a classical probabilistic system with a finite set A of configurations. The general probability distribution or measure μ on A can be written in the form

$$\mu = \sum_{a\in A} p_a[a],$$

where each probability $p_a \geq 0$ and

$$\sum_{a\in A} p_a = 1.$$

The symbol $[a]$ represents both the configuration a and the probabilistic state in which a is certain. (This state is also called an atom or a Kronecker delta function.) The set of all normalized states forms a simplex Δ_A whose vertices are the set A. These simplices are the objects of a reasonable if slightly non-standard definition of finite, classical probability theory. The other elements of this definition are as follows.

An *event* is a subset $E \subseteq A$. A state μ assigns a probability to E between 0 and 1 by the formula

$$P[E] = P_\mu[E] = \sum_{a\in A} p_a.$$

If $P[E] > 0$, then E and μ induce a conditional state on E given by the formula

$$\widehat{\mu|_E} = \frac{1}{P[E]} \sum_{a\in E} p_a\ [a].$$

A *random variable* or *measurement* is a function $X : A \to B$ for some set B. It is interpreted as a partition of A into disjoint events: the equation $X = b$, as an event, is the set of all a such that

$X(a) = b$. If X takes values in $\mathbb{R}$ or $\mathbb{C}$, then it also has an *expectation*, defined as

$$E[X] = \sum_a p_a X(a).$$

If A and B are two finite configuration sets, then a function

$$M : \Delta_A \to \Delta_B$$

is a *stochastic map* (or a *Markov map*) if it comes from a linear map

$$M : \mathbb{R}^A \to \mathbb{R}^B.$$

The linearity of M is the *classical superposition principle*. Another way to describe a stochastic map is to start with a linear map M and impose the positivity condition $M(\Delta_A) \subseteq \Delta_B$. If M is viewed as a matrix, the positivity condition says that the entries of M are non-negative and each column sums to 1. In this case M is called a *stochastic matrix*. Stochastic maps are the natural empirical class of maps in probability theory.

A non-negative vector μ in $\mathbb{R}^A$ is a *subnormalized state* if the total probability is at most 1, and a linear map

$$M : \mathbb{R}^A \to \mathbb{R}^B$$

that preserves subnormalized states is called *substochastic*. A substochastic map is also called an *extinction process*. It can be converted to a stochastic map by adding an extinction state to the target set B.

If A and B are the configuration sets of two probabilistic systems, then the two systems together have a joint configuration set $A \times B$. The joint simplex $\Delta_{A\times B}$ lies in a tensor product:

$$\Delta_{A\times B} \subset \mathbb{R}^(A \times B) = \mathbb{R}^A \otimes \mathbb{R}^B.$$

A joint state $\mu \in \Delta_{A\times B}$ is *independent* or a *product state* if it factors as a tensor product:

$$\mu = \mu_A \otimes \mu_B.$$

More typically μ does not have this form, in which case it is *correlated*. A stochastic map that affects only one of A and B has the form $M \otimes I$ or $I \otimes M$; likewise an event that depends on only one of A or B has the form $E \times B$ or $A \times E$. Whether or not μ is independent, it induces states on A and B called *marginals*. They are defined as:

$$\mu_A = \sum_{a\in A} \big(\sum_{b\in B} p_{(a,b)}\big)[a] \qquad \mu_B = \sum_{b\in B} \big(\sum_{a\in A} p_{(a,b)}\big)[b].$$

Evidently, if μ is a product, then it is the product of its marginals.

Example 1.10.1. Two fair dice are rolled. If we take the dice as subsystems A and B with a joint state, then $A = B = \{1, \ldots, 6\}$ and the state μ is the *uniform* state

$$\mu = \frac{1}{36} \sum_{1\le s,t\le 6} [(s, t)].$$

The event of rolling 7 is the set $E = \{(s, t)|s+t = 7\}$. The chance of E is $P[E] = \frac{1}{6}$. The state μ is an independent state, but the conditional state $\mu|_E$ is correlated. A protocol such as "if you rolled 7, pick one die at random and roll it again" is modelled by a stochastic map.

The rules of classical probability theory are the basis of a (Bayesian) statistician's model of ordinary human experience, as well as most scientific experiments, according to the following interpretation:

1. State: Each observer can model external reality as a measure space, such as a finite set A, that carries has probability distribution at each point in time.

2. Independence: Reality decomposes into approximately disjoint subsystems modelled by Cartesian products of measure spaces. An observer is an approximately independent subsystem whose residual non-independence is described by witnessed information.

3. Evolution: After an observer witnesses an event, the new state of reality is given by conditional expectation. More generally the state of reality evolves by stochastic and substochastic maps.

4. Statistics: An observer's experiences are viewed as independently repeatable experiments. The probability of an event is the fraction of times that it occurs in repeated trials of the experiment.

These rules can be accepted in one of two ways: (1) they hold empirically; (2) they can be mimicked by deterministic systems with hidden information. The first reason, but not the second, applies to the quantum analogue of these rules.

Exercises

1.11. Appendix: Categories and tensors

Much of the presence of category theory in mathematics is in the spirit of Moliere: You can use categories without knowing it. In many areas of

mathematics there is a distinguished class of objects (*e.g.*, sets, groups, topological spaces, vector spaces) and a distinguished class of functions between them (*e.g.*all functions, group homomorphisms, continuous functions, linear maps). In some cases two objects are related by a transformation which isn't strictly a function but behaves a lot like one; the historically important example was a function between two topological spaces considered up to homotopy.

A *category* $\mathcal{C}$ is a possibly abstract class of set-like *objects* and a class of function-like relations, or *morphisms*. Every morphism f has a *domain* A and a *target* B (both of them objects in $\mathcal{C}$) and is then denoted

$$f : A \to B.$$

There is a partial composition law: If the target of f is the same object as the domain of g, or

$$f : A \to B \qquad g : B \to C,$$

then there is a composition

$$f \circ g : A \to C.$$

The composition law for $\mathcal{C}$ is required to be associated, and for every object A there should be an identity morphism

$$i : A \to A$$

which satisfies the obvious identity axiom with respect to composition. These axioms are not all that restrictive and there are a wide variety of different kinds of categories.

Examples 1.11.1. The category of sets, with all functions as the morphisms, is called **Set**. The category of vector spaces over a field $\mathbb{F}$, with linear transformations as the morphisms, can be called $\mathbf{Vect}_{\mathbb{F}}$. The category of topological spaces with continuous functions as the morphisms is called **Top**.

Two different categories can have the same objects but different morphisms. For example, the obejcts of the category **iSet** are sets, but the morphisms are just the injective functions.

As an example of a more abstract category, if G is a group, it yields a category with one object whose morphisms are the elements of G.

A *tensor category* is a category with a multiplication law, denoted "$\otimes$" and called a "tensor product", for both objects and morphisms. The multiplication law should be associative: $A \otimes (B \otimes C)$ should be the same object as $(A \otimes B) \otimes C$[4]. The axioms for taking tensor products of morphisms is more complicated because the axioms include compatibility relations between the operations of composing and tensoring morphisms. The tensor categories of interest in this article are all *symmetric*, meaning that "$\otimes$" is commutative as well as associative, both for objects and morphisms.

The most important example of a tensor category is Vect, the category of vector spaces.

[4] There is a distinction between a *strict* tensor category, in which $A \otimes (B \otimes C)$ and $(A \otimes B) \otimes C$ are exactly the same object, and a *lax* tensor category, in which they are isomorphic and the isomorphisms are meta-associative; this meta-associativity is called "coherence". We don't have to worry about lax tensor categories here, but note that they arise in physics in the guise of spin calculus; the main coherence axiom is known as the Biedenharn-Elliott identity.

2. MECHANICS

Basic quantum mechanics assumes the rules of quantum probability plus a single additional rule. If a physical system is independent and autonomous (or *closed*), then quantum mechanics postulates that it has a pure state space $\mathcal{H}$ and its state evolves according to a one-parameter group $U(t)$ of unitary operators. The group has the form

$$U(t) = \exp(-itH)$$

for some self-adjoint (but often unbounded) operator H, which is called the *Hamiltonian* of the system. If the state at time t is $|\psi\rangle = |\psi(t)\rangle$, it evolves by the *Schrödinger equation*,

$$i\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle.$$

Besides the system's autonomous behavior, it can also be measured. In particular, H itself is a measurement and its value is called the *energy* of the system. A state $|\psi\rangle$ with definite energy E satisfies the eigenvalue equation

$$E|\psi\rangle = H|\psi\rangle,$$

which is also called the time-independent Schrödinger equation.

The operator $\frac{\partial}{\partial t}$ has physical units of inverse time, while the Hamiltonian H has units of energy. Therefore quantum mechanics requires a conversion factor $\hbar$, known as *Planck's constant*, to relate the two operators. Written with units, the Schrödinger equation is

$$i\hbar\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle.$$

If $\hbar$ is small compared to the scale of a physical system, its quantum phase will oscillate wildly, which leads to classical probabilistic behavior. In metric units,

$$\hbar \approx 6.6262 \times 10^{-34} Js,$$

where J is joules and s is seconds. So Planck's constant is extremely small on the human scale.

The fact that the Hamiltonian H often has a discrete spectrum points to the origin of quantum mechanics and its name. The word "quantum" now often means simply "non-commutative", but the original meaning is "discrete". The original purpose of quantum mechanics was to explain why many physical measurements unexpectedly take discrete values. For example, the energy spectrum of the quantum harmonic oscillator is $\mathbb{Z}_{\geq 0} + \frac{1}{2}$. The energy of a classical harmonic oscillator can of course be any non-negative real number.

2.1. Wave mechanics

Wave mechanics is a model of a single particle moving in $\mathbb{R}^d$ with $1 \leq d \leq 3$. To give the particle something to do, it interacts with a potential $V(\vec{x})$. Then wave mechanics posits the specific Hamiltonian

$$H = \frac{\vec{p}\cdot\vec{p}}{2} + V(\vec{x}),$$

where $\vec{x}$ and $\vec{p}$ are two vector-valued operators. The component x_k of $\vec{x}$ is interpreted as multiplication by the coordinate x_k, while $\vec{p} = -i\partial/\partial\vec{x}$. The Schrödinger equation then becomes the linear partial differential equation

$$i\frac{\partial\psi}{\partial t} = -\frac{\Delta\psi}{2} + V(\vec{x})\psi,$$

where

$$\Delta = \frac{\partial}{\partial\vec{x}}\cdot\frac{\partial}{\partial\vec{x}}$$

is the Laplacian. Since this is a wave equation, it illustrates half of *particle-wave duality*. The state of a particle travels through space as a wave, which is why it is called a *wave function* (but see exercise 2.1.3).

The operator $\vec{x}$ is interpreted as position and has physical units of distance. The potential $V(\vec{x})$ has units of energy, the same as H. With a factor of $\hbar$,

$$\vec{p} = -i\hbar\frac{\partial}{\partial\vec{x}},$$

the operator $\vec{p}$ is interpreted as the particle's linear momentum. The kinetic energy of a particle with momentum $\vec{p}$ is $(\vec{p}\cdot\vec{p})/2m$, where m is its mass, so with units the Hamiltonian becomes

$$H = \frac{\vec{p}\cdot\vec{p}}{2m} + V(\vec{x}).$$

The mass m, an energy scale E, and Planck's constant $\hbar$ fully determine the dimensional scale of Schrödinger's equation. For example, the length scale of the particle's quantum behavior is $d = \hbar/\sqrt{Em}$, which decreases as m increases. Thus if several particles interact on a common length and energy scale, heavier particles behave more classically than lighter ones.

Example 2.1.1. A particle in the potential $V(\vec{x}) = 0$ is called *free*. In this case the Hamiltonian has no point spectrum and the time-independent Schrödinger equation has no solutions with finite Hilbert norm. But the plane wave

$$\psi(\vec{x}) = e^{i\vec{k}\cdot\vec{x}}$$

is an interesting infinite-norm solution. It has definite momentum $\vec{p} = \vec{k}$ as well as definite energy. It can be approximated by a wave packet (or wavelet) that, according to the Schrödinger equation with physical units, travels with velocity $\vec{k}/m$.

The behavior of a free particle illustrates an interpretation of the continuous spectrum of an arbitrary Hamiltonian. A state $|\psi\rangle$ is *unbound* if it wanders under unitary evolution:

$$\lim_{t\to\infty} \langle\psi|U(t)|\psi\rangle = 0.$$

It is *bound* if it recurs:

$$\limsup_{t\to\infty} |\langle\psi|U(t)|\psi\rangle| = 1.$$

These are the quantum analogues of the notions of wandering and non-wandering points in (reversible) classical dynamical systems. The spectral theorem implies that point-spectrum states are bound and continuous-spectrum states are unbound. In particular, every bound state is a quantum superposition of fixed-energy or stationary states.

Example 2.1.2. The simple harmonic oscillator is a 1-dimensional system with Hamiltonian

$$H = \frac{p^2}{2} + \frac{x^2}{2}.$$

The coherent state

$$\psi(x) = \frac{e^{-x^2/2}}{(2\pi)^{1/4}}\sqrt{dx}$$

is an eigenstate with energy $\frac{1}{2}$. The rest of the spectrum can be found with *ladder operators*. This is a Lie-algebraic method that begins with the commutation relations

$$[x,p] = i \qquad [H,x] = -ip \qquad [H,p] = ix.$$

Define the *lowering* and *raising* operators

$$a = \frac{x+ip}{\sqrt{2}} \qquad a^* = \frac{x-ip}{\sqrt{2}}.$$

Then H, a, and a^* satisfy

$$[a,a^*] = 1 \qquad [H,a] = -a \qquad [H,a^*] = a.$$

These commutation relations imply that if a state $|\psi\rangle$ has energy E, then $a|\psi\rangle$ has energy $E-1$ and $a^*|\psi\rangle$ has energy $E+1$, provided that either vector is non-zero. The coherent state spans the kernel of a and we rename it $|0\rangle$. The commutation relations imply, by induction, that

$$\langle 0|a^n(a^*)^n|0\rangle = n!,$$

so the state

$$|n\rangle = \frac{(a^*)^n|0\rangle}{\sqrt{n!}}$$

is normalized and is an eigenstate with energy $n+\frac{1}{2}$.

Thus the harmonic oscillator has a sequence of eigenstates $|0\rangle, |1\rangle, \ldots$. To show that they span $L^2(\mathbb{R})$, observe that $H \geq 0$ as an operator, so its spectrum is non-negative. If the spectrum of a state $|\psi\rangle$ lies in the interval $[0,n]$, then

$$a^{n+1}|\psi\rangle = 0.$$

It follows that

$$|\psi\rangle = P(a^*)|\psi\rangle$$

for some polynomial P of degree at most n. Thus $|\psi\rangle$ is a linear combination of $|0\rangle, \ldots, |n\rangle$.

Wave mechanics also postulates a multiparticle Schrödinger equation. The state space of n particles in d dimensions is

$$L^2(\mathbb{R}^d)^{\otimes n} \cong L^2(\mathbb{R}^{dn}).$$

The dn coordinates of this Hilbert space are divided into n vector-valued position operators $\vec{x}_1, \ldots, \vec{x}_n$, and there are n corresponding momentum operators $\vec{p}_1, \ldots, \vec{p}_n$. The Hamiltonian has the general form

$$H = \sum_{k=1}^{n} \frac{\vec{p}_k \cdot \vec{p}_k}{2} + V(\vec{x}_1, \ldots, \vec{x}_n).$$

The corresponding Schrödinger equation is a PDE in dn dimensions. It is usually intractable, even with the aid of computers (but see Chapter 3). Almost the only case with a satisfactory solution is the one in which the potential factors,

$$V(\vec{x}_1, \ldots, \vec{x}_n) = \prod_{k=1}^{n} V_k(\vec{x}_k),$$

which means that the particles do not interact with each other.

Exercises

1. Show that

$$\psi(\vec{x}) = \frac{\sin k|\vec{x}|}{|\vec{x}|}$$

is an infinite-norm solution to the free-particle Schrödinger equation. It represents a spherically radiating particle.

2. The function
$$\psi(\vec{x}) = \frac{1}{|\vec{x}|}$$
is not considered an infinite norm solution to the free-particle Schrödinger equation. For example, $f(\vec{x})$ is a smooth bump function, then $f(\vec{x})\psi(\vec{x})$ is not an approximate eigenstate of H. Why not?

3. A *nonlinear Schrödinger equation* is a partial differential equation
$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2}\frac{\partial^2\psi}{\partial x^2} + V(x,\psi)$$
for some function V which is not linear in ψ. Similarly, given a pair of wave functions
$$(\psi_1,\psi_2) \in L^2(\mathbb{R}) \oplus L^2(\mathbb{R}),$$
we can write coupled Schrödinger equations (either linear or nonlinear):
$$i\frac{\partial\psi_1}{\partial t} = -\frac{1}{2m_1}\frac{\partial^2\psi_1}{\partial x^2} + V_1(x,\psi_1,\psi_2)$$
$$i\frac{\partial\psi_2}{\partial t} = -\frac{1}{2m_2}\frac{\partial^2\psi_2}{\partial x^2} + V_2(x,\psi_1,\psi_2).$$
Explain why a nonlinear Schrödinger equation cannot model a quantum particle and coupled Schrödinger equations cannot model coupled quantum particles.

4. Prove that if a state lies in the point spectrum of H, then it recurs, while if it lies in the continuous spectrum, then it wanders. Prove that in the Schrödinger wave equation, if the potential $V(x)$ is non-negative, then an unbound particle must escape to infinity.

5. Prove that the harmonic oscillator state $|n\rangle$ has the form
$$H_n(x)e^{-x^2/2}$$
for some polynomial $H_n(x)$ of degree n. Since the polynomials are orthogonal, they are Hermite polynomials up to rescaling x and multiplying by a constant factor.

6. Let ρ be the standard quasifree state with parameter a, defined in Example 1.7.1. Show that the harmonic oscillator Hamiltonian H has a discrete exponential distribution with respect to the state ρ with parameter
$$t = \frac{a-1}{a+1}.$$
I.e., show that
$$\rho = \sum_{n=0}^{\infty}(t-1)t^n|n\rangle\langle n|.$$

2.2. A classical limit

In this section we will reformulate wave mechanics as a version of Hamiltonian mechanics. In classical physics, Hamiltonian mechanics produces a dynamical system (Hamilton's equations) on the configurations of a physical object for every smooth function H (the Hamiltonian) on the configuration space. The configuration space must have a symplectic or Poisson structure and is also called phase space. The quantum version is an important generalization of wave mechanics, and it also limits to classical Hamiltonian mechanics as $\hbar \to 0$.

We first restate quantum mechanics so that measurement operators evolve and states does not. Given a general Hamiltonian H acting on an arbitrary Hilbert space $\mathcal{H}$, let $|\psi(t)\rangle$ be the state at time t and let
$$|\psi\rangle = |\psi(0)\rangle.$$
If A is a measurement operator to be applied at time t, define
$$A(t) = U(-t)AU(t).$$
Then evolving the operator A and fixing the state ψ is statistically equivalent to fixing the operator and evolving the state:
$$\langle\psi|A(t)|\psi\rangle = \langle\psi(t)|A|\psi(t)\rangle.$$
This equivalence is a special case of the conjugation principle in group theory: If a group G acts on a set S, then ghg^{-1} does to $g(s)$ what h does to s. As applied to quantum mechanics, it is called the *Heisenberg picture*. The differential form of the Heisenberg picture is an operator-valued differential equation called the *Heisenberg equation*:
$$i\frac{\partial A}{\partial t} = [A,H].$$

We can now rename variables to match Hamilton's equations in symplectic $\mathbb{R}^{2n}$. Substitute $\vec{q}$ for $\vec{x}$ and n for d and drop the restriction $d \leq 3$; the state space becomes $L^2(\mathbb{R}^n)$. The operator $\vec{q}$ consists of the coordinates $q_1,\ldots,q_n$ on $\mathbb{R}^n$, while $\vec{p}$ is defined by
$$p_k = -i\frac{\partial}{\partial q_k}.$$
These operators satisfy the commutation relations
$$[q_j,p_k] = i\delta j,k.$$
Now assume that the Hamiltonian operator H is a "function" $H(\vec{p},\vec{q})$ of $\vec{p}$ and $\vec{q}$. For example it might

be a non-commutative polynomial in the coordinates of $\vec{p}$ and $\vec{q}$ or a suitably convergent power series. Then formally

$$[q_k, H] = i\frac{\partial H}{\partial p_k} \qquad [p_k, H] = -i\frac{\partial H}{\partial q_k}.$$

Combining this with the Heisenberg equation yields an operator form of Hamilton's equations for conjugate variables:

$$\frac{\partial q_k}{\partial t} = \frac{\partial H}{\partial p_k} \qquad \frac{\partial p_k}{\partial t} = -\frac{\partial H}{\partial q_k}.$$

To see the classical limit, assume for simplicity that $n = 1$ and the conjugate variables are just p and q. In units of Planck's constant, their commutator is

$$[q, p] = i\hbar.$$

Now take the limit $\hbar \to 0$. The Heisenberg uncertainty relation (Exercise 1.2.8) yields the inequality

$$V[q]V[p] \geq \frac{\hbar^2}{4}.$$

In fact the inequality is sharp: Gaussian wave packets (*i.e.*, coherent states) achieve equality. If $\hbar$ is small, then p and q can both be nearly definite in the initial state $|\psi(0)\rangle$. A Hamiltonian expressed in terms of p and q is also nearly definite and the quantum dynamics of the system preserves near definiteness at least for a while. In conclusion, classical Hamiltonian dynamics is a valid short-term approximation to quantum Hamiltonian dynamics.

Exercises

1. Let $|\psi\rangle$ be a state with respect to which

$$V[q]V[p] = \frac{\hbar^2}{4}.$$

Show that $|\psi\rangle$ is pure and has the form

$$\psi(x) = \frac{e^{a(x-b+ic)^2/2}}{(\pi a)^{1/4}}$$

for some real constants a, b, and c. (The converse of this exercise is also worthwhile and is much easier.)

2.3. Symmetry and spin

If a physical system can be rotated in space, its state space $\mathcal{H}$ becomes a representation of the Lie group SO(3). By Noether's theorem for quantum systems, if the system's Hamiltonian H is invariant under rotation, the Lie generators, if multiplied by i, are Hermitian operators with conserved values. The angular momentum operators in the x, y, and z directions in $\mathbb{R}^3$ are written J_x, J_y, and J_z; the angular momentum in the direction of a general unit vector $\vec{v}$ is $\vec{v} \cdot \vec{J}$. These operators do not commute with each other, so they are not simultaneously definite. There is another twist as well. Since global phase is not statistically meaningful, the state space $\mathcal{H}$ might only be a projective representation of any given symmetry group. This can happen with spatial rotations, so the true quantum rotation group is the double cover

$$\widetilde{\mathrm{SO}(3)} \cong \mathrm{SU}(2).$$

Whether $\mathcal{H}$ is a projective representation or a linear one, the analysis will show that angular momentum takes discrete values that differ by multiples of $\hbar$, even though there is a continuous family of directions in which to measure it. Moreover, the rotational state space of a physical system is typically finite.

The operators J_x, J_y, and J_z satisfy the commutation relations

$$[J_x, J_y] = iJ_z \qquad [J_y, J_z] = iJ_x \qquad [J_z, J_x] = iJ_y.$$

Angular momentum has the same physical units as Planck's constant, so for example $[J_x, J_y] = i\hbar J_z$ in units. Since SU(2) is compact, every unitary representation decomposes as an orthogonal direct sum of finite-dimensional irreducible representations (irreps). By Cartan-Weyl theory, it has a unique irrep of dimension $n+1$ for every integer $n \geq 0$, which we will call V_j with $j = n/2$. It is also called a *spin-j* system. The standard basis is

$$|-j\rangle, |1-j\rangle, \ldots, |j\rangle,$$

where $|m\rangle$ is an eigenstate of J_z with eigenvalue m. The proof that the irreps of SU(2) have this structure uses the same ladder operator method as in the harmonic oscillator system. In this case the ladder operators are

$$J^+ = J_x + iJ_y \qquad J^- = J_x - iJ_y.$$

By convention their action on the standard basis is

$$J^+|m\rangle = \sqrt{(j-m)(j+m+1)}|m+1\rangle$$
$$J^-|m\rangle = \sqrt{(j+m)(j-m+1)}|m-1\rangle.$$

Another important operator is

$$J^2 = J_x^2 + J_y^2 + J_z^2,$$

which in representation theory is called a Casimir operator. Its sole eigenvalue on the spin-j system is $j(j+1)$.

The term "spin" often refers to the intrinsic angular momentum of a particle. Electrons, protons, neutrons, and neutrinos all have spin $\frac{1}{2}$, while the cobalt-60 nucleus, for example, has spin 5. The spin-$\frac{1}{2}$ spin states are also written $|\uparrow\rangle$ and $|\downarrow\rangle$, and they are also called *right-handed spin* and *left-handed spin* when measured in the direction of the particle's motion. A photon is a spin-1 particle, but the spin state $|0\rangle$ never occurs in the direction of motion. (This is only possible because of special relativity.)

Although some of the most interesting behavior of a particle is due to its spin, it is worth remembering that most of its state is in its position. In general if a particle has spin-j, its total state space (in a flat universe) is $V_j \otimes L^2(\mathbb{R}^3)$.

There are two ways that V_j arises extrinsically. First, if a particle lies on the unit 2-sphere S^2, or if it exists in $\mathbb{R}^3$ but its radial state is an independent factor, then it possesses *orbital angular momentum*. (Orbital angular momentum operators are often denoted $L_z, L^+, \ldots$, but we will stay with J.) The rotation group SO(3) acts on S^2 and by extension $L^2(S^2)$. It decomposes as

$$L^2(S^2) \cong V_0 \oplus V_1 \oplus V_2 \oplus \cdots,$$

where a vector in the summand V_j is a spherical harmonic of degree j. If the particle moves freely on S^2, then its Hamiltonian is

$$H = -\frac{\Delta}{2}.$$

The Laplacian Δ equals the action of the Casimir operator J^2 on the representation $L^2(S^2)$. Thus the spectrum of H is $\{\frac{j(j+1)}{2}\}$, and the jth eigenvalue (numbered from 0) has multiplicity $2j+1$. Especially in atomic physics, the first several harmonic spaces are often denoted by letters: s, p, d, f, g, ...

Second, when two or more spin systems are combined, their joint state space decomposes as a direct sum of spin systems. The Clebsch-Gordan rule gives the decomposition of two spin systems:

$$V_j \otimes V_k \cong V_{j+k} \oplus V_{j+k-1} \oplus \cdots \oplus V_{|j-k|}.$$

Thus a spin-j particle and a spin-k particle can together be in a spin-ℓ state with

$$\ell \in \{|j-k|, \ldots, j+k\}.$$

Higher tensor products also decompose, of course, but in a more complicated way. The main rule to remember is that the total angular momentum operators match the Leibniz rule for the diagonal action of a Lie algebra on a tensor product. *I.e.*, the action of $\vec{J}$ on

$$V_{j_1} \otimes V_{j_2} \otimes \cdots \otimes V_{j_n}$$

is given by

$$\vec{J} = \sum_{k=1}^{n} \vec{J}^{(k)} = \sum_{k=1}^{n} I^{\otimes k-1} \otimes J_{\vec{v}} \otimes I^{\otimes N-k}.$$

Exercises

2.4. Identical particles

Following Section 1.1, two coherent trajectories of a physical system obey quantum superposition if they arrive at the same state and classical superposition if they arrive at different states. Thus in quantum mechanics there is a difference between a pair of particles (or any other two physical systems) that are logically identical and a pair that merely appear the same. If they are identical, then a coherent beam apparatus that conditionally switches them,

(*picture*)

exhibits quantum interference. By this test and its consequences, many classes of particles are known to be identical. All electrons are identical, all helium nuclei are identical, etc.

More explicitly, suppose that two particles have the same state space $\mathcal{H}$. Then by the independence rule, their joint state space is $\mathcal{H}^{\otimes 2}$. This state space has an operator X that switches the two factors:

$$X(|\psi_1\rangle \otimes |\psi_2\rangle) = |\psi_2\rangle \otimes |\psi_1\rangle.$$

Since X is a Hermitian involution, it has two eigenspaces. The space $S^2\mathcal{H}$ of symmetric tensors has eigenvalue 1, while the space of antisymmetric tensors $\Lambda^2\mathcal{H}$ has eigenvalue -1. If the two particles are identical, then X must fix the density operator $|\psi\rangle\langle\psi|$, but not necessarily the state vector $|\psi\rangle$ itself. Thus $|\psi\rangle$ can lie in either eigenspace of X, so particles can be identical in two different ways. If

$$X|\psi\rangle = |\psi\rangle,$$

then the particles are called *bosons* and obey *Bose-Einstein statistics*. If

$$X|\psi\rangle = -|\psi\rangle,$$

then the particles are called *fermions* and obey *Fermi-Dirac statistics*. More generally, given n particles with the same state space $\mathcal{H}$, if they are identical bosons, then their joint state space is the symmetric space

$$S^n\mathcal{H} \subseteq \mathcal{H}^{\otimes n}.$$

If they are identical fermions, then their joint state space is the antisymmetric space

$$\Lambda^n \mathcal{H} \subseteq \mathcal{H}^{\otimes n}.$$

Here is a completely explicit description in terms of amplitudes. If n particles each have state space $\mathcal{H}$ with a basis A, then a joint pure state can be written

$$|\psi\rangle = \sum \alpha_{a_1,a_2,\ldots,a_n} |a_1, a_2, \ldots, a_n\rangle.$$

If the particles are identical bosons, then

$$\alpha_{a_1,a_2,\ldots,a_n} = \alpha_{a_{\sigma(1)},a_{\sigma(2)},\ldots,a_{\sigma(n)}}$$

for any permutation $\sigma \in S_n$. If they are identical fermions, then

$$\alpha_{a_1,a_2,\ldots,a_n} = (-1)^\sigma \alpha_{a_{\sigma(1)},a_{\sigma(2)},\ldots,a_{\sigma(n)}}.$$

In this case, any amplitude α with a repeated index must vanish, so identical fermions cannot appear in the same state. This conclusion is called the *Pauli exclusion principle*.

Example 2.4.1. Let $\mathcal{H} = L^2(S^1)$ and let

$$\psi(\theta_1, \theta_2) = \frac{\sin(\theta_1 - \theta_2)}{\sqrt{2}\pi}$$

be a joint wave function for two particles on the circle S^1. It is antisymmetric and can describe identical fermions, by the natural continuous analogue of amplitude antisymmetry. At the same time, the state is invariant under rotation of the circle, so the measured value of θ_1 has the uniform distribution.

2.5. Atomic structure

In this section we will draw together the ideas of Sections 2.1, 2.3, and 2.4 to analyze the hydrogen atom. We will also qualitatively predict some of the features of other atoms and molecules.

2.6. Quantum field theory

3. COMPUTATION

Quantum computation is a computational model based on quantum probability as described in Chapter 1. It was first proposed by Feynman [**?**] that artificial quantum systems — quantum computers — could be used to simulate natural quantum systems (Chapter 2). A sequence of further ideas [**? ? ?**] culminated in Shor's discovery of polynomial-time quantum to factor integers [**?**]. It is now reasonable to conjecture that quantum computation is sometimes exponentially faster than deterministic (or classically randomized) computation. This possibility is as almost as surprising as quantum probability itself, and it would give a new reason that quantum probability does not reduce to classical probability (Section 1.6).

Quantum algorithms to simulate many natural Hamiltonians [**?**] close the circle with Feynman's proposal. They indicate that quantum probability, and not any further features of quantum mechanics, provides the likely acceleration of quantum computers. Indeed, since experimental quantum computation is very difficult in practice, we can say that the realistic physics that we might exploit for quantum computation is both a friend and a foe.

Since quantum computation depends on continuous state, it is fair ask whether it unrealistically exploits analog precision. This is related to the problem that quantum computation is more fragile than classical computation and must be protected from decoherence. These issues are addressed by the theory of quantum error correction and fault-tolerant computation Section 3.6. Although theory leans to the conclusion that quantum computation is possible, useful quantum computers have not yet been built. There could conceivably be some unknown practical barrier to quantum computation.

Quantum secrecy is a parallel development [**?**] which, unlike quantum computation, has been convincingly demonstrated [**?**]. The idea is to use the violation of Bell's inequalities and related phenomena to detect eavesdropping in communication, rather than to defy the eavesdropper's computational power. Since quantum secrecy assumes that eavesdroppers have unlimited computational power, it is more trustworthy than classical cryptography, but it is also more limited. Regardless, the first practical use of quantum computers could be as repeaters for quantum secrecy protocols.

3.1. Computational models

To put quantum probability into a computational form, we begin with the *qubit*, which has a 2-

dimensional state space $\mathbb{C}^2$ with standard basis $|0\rangle$ and $|1\rangle$. The memory of a quantum computer could consist of n qubits with state space $\mathcal{H} = \mathbb{C}^{2^n}$. The general pure state is given by a unit vector in $\mathcal{H}$, while the general state is a matrix in the Bloch region $\mathcal{B}_{2^n} \subset \mathcal{M}_{2^n}$. A computation could consist of a sequence of *2-qubit, unitary gates* applied to this 2^n-dimensional state space $\mathbb{C}^{2^n}$, followed by a measurement. By definition such a gate is a unitary operator $U : \mathbb{C}^4 \to \mathbb{C}^4$ applied to some pair of the qubits. More generally, a (j, k)-*qubit gate* is a quantum operation

$$\mathcal{E} : \mathcal{M}_{2^j} \to \mathcal{M}_{2^k}.$$

The general *quantum circuit* is a sequence of such quantum operations applied to j-tuples of qubits in a memory with n qubits. There is usually a uniform bound on the gates such as $j, k \leq 2$. If $j \neq k$, then n changes by $k - j$ as the operation is applied. As in classical Boolean circuitry, a quantum circuit is equivalent to an acyclic digraph with each node labelled by a quantum operation.

Example 3.1.1.

Before considering quantum algorithms in this model, we can compare qubits to randomized bits. Since a state of n qubits is a quantum superposition of the 2^n basis states, or a mixture of these, quantum computation is a form of parallel computation. Since the coefficients are complex numbers, it is also a form of analog computation. However, randomized classical computation shares both of these features. Indeed, the simplex Δ_{4^n} of states of $2n$ classical bits has the same dimension as the Bloch region $\mathcal{B}_{2^n}$ of n qubits. Qubits are more useful than randomized bits because they allow more operations, not because they carry more state. (But see Section 3.5.)

We can also place quantum computation in the context of standard complexity classes. To review, a complexity class is a set of computational decision problems (YES-NO-valued functions on the set of finite bit strings) that can be answered with specific computational resources. The computational resources defining a particular complexity class may or may not be realistic. Here are some standard complexity classes with the most natural quantum class included:

1. P is the set of problems that can be solved in deterministic polynomial time.

2. NP is problems that can be solved in non-deterministic polynomial time. A non-deterministic computer program is one with blank conditionals. Its execution history is a tree rather a sequence. A non-deterministic program answers a question in NP if at least one leaf of the computation tree says YES when the true answer is YES, but all leaves say NO when the true answer is NO, and if the tree has polynomial depth.

3. BPP is probabilistic polynomial time with bounded error. It can be defined using the same non-deterministic model as NP, except that each conditional is assigned a computed probability. The program answers a question in BPP if a random computation path provides the true answer with probability at least $\frac{2}{3}$.

4. BQP is the quantum analogue of BPP. A problem is in BQP if it is solved by a polynomial quantum circuit. As in other circuit models, the circuit must be efficiently precomputed (say with a classical computer) using the length of the input. The final stage is a boolean-valued measurement, which must be correct with probability $\frac{2}{3}$ as in the definition of BPP.

5. PSPACE is deterministic polynomial space with no restriction on computation time. It is equivalent to a non-deterministic polynomial-time computation model in which each node of the computation tree is assigned an arbitrary binary function of its child nodes, and the final answer is the boolean value assigned to the root. It is also equivalent to polynomial-time parallel computation with exponentially many networked processors.

A complexity class can also be modified by an oracle, which is a black-box function that a program can invoke in one step, or that can be used as a large gate in a circuit.

The inclusions

$$\text{BPP} \subseteq \text{BQP} \subseteq \text{PSPACE}.$$

are elementary. BPP can also be defined by circuits of stochastic maps, which places it inside BQP. For the other inclusion, any quantum or stochastic circuit can be evaluated by path summation (Section **??**), which requires only as much memory as the number of nodes in the circuit. In fact quantum amplitudes can be estimated in a space-efficient way with stochastic sampling. Thus quantum randomness only saves time and not space over classical randomness.

It is reasonable to conjecture that P = BPP, that BQP is somewhat larger than BPP, and that PSPACE is vastly larger than BQP. It is also reasonable to conjecture that BQP does not contain NP. For one reason, it is easy to find an oracle A such that BQP^{A} does not contain NP^{A}.

3.2. Low-level operations

A first step in constructing quantum algorithms is to identify a convenient set of universal quantum gates. At the unitary level, it is convenient to include multiplication by a global phase even though it is irrelevant. The group of unitary operators on a single qubit is then $U(2)$. Some important examples of single-qubit (or *unary*) gates are the NOT gate X, the Hadamard gate H, and the phase rotation $Z(\theta)$:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$
$$Z(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}$$

The gate $Z = Z(\pi)$ is a *phase flip*. The phase subgroup $\{Z(\theta)\}$ and the Hadamard gate H generate $U(2)$.

If $U \in U(2)$ is unitary, then there is a corresponding *coherently controlled* extension $C(U)$ of U to 2 qubits. It applies U to the right qubit when the left qubit is in the state $|1\rangle$. As a matrix,

$$C(U) = \begin{pmatrix} I & 0 \\ 0 & U \end{pmatrix}.$$

Coherently controlled unitaries applied to both qubits generate the 2-qubit group $U(4)$, and 2-qubit unitaries applied to arbitrary pairs of qubits in an n-qubit memory generate its full unitary group $U(2^n)$.

Strictly speaking the unitary group on n qubits is not full quantum computation, because the latter also includes decoherent quantum operations. By Theorem 1.5.1, unitary operators together with qubit erasure (taking the marginal from n qubits to $n-1$) and qubit creation in the state $|0\rangle$ generate the entire category of quantum operations.

Furthermore, any quantum computation can be dilated to a unitary computation. If the output is classical, the unitary part is followed by the measurement of output qubits. Such a dilation changes computation time by only a constant factor, but it can be expensive in space. Nonetheless, many quantum algorithms are unitary with classical pre- and postprocessing. Decoherent operations often (but not always) squander the quantum acceleration.

Quantum operations also need not be exact in order to work. Given two states ρ and ρ', the probability that two measurements might give a different answer is bounded by the trace distance

$$d(\rho, \rho') = \frac{1}{2}||\rho' - \rho||_1.$$

If $\mathcal{F}$ and $\mathcal{F}'$ are two quantum operations with the same domain and target, their distance $d(\mathcal{F}, \mathcal{F}')$ is defined as the Lipschitz constant of their difference $\mathcal{F}' - \mathcal{F}$ with respect to trace distance on states. This distance behaves predictably with respect to composition and tensor products. Its relevance is that each step in a quantum algorithm only needs to be approximated to within a tolerance. Among other uses, approximate computation is needed for the fault-tolerance problem in Section 3.6.

3.3. Dilations and Grover's algorithm

It is useful to explicitly construct unitary dilations of classical algorithms. Suppose for simplicity that f is boolean-valued with classical input s and that it is computed by a sequence of classical gates. Then we can dilate the computation one gate at a time. For example, the Toffoli gate is a $(3,3)$-qubit gate defined by

$$T|a, b, c\rangle = |a, b, c + ab\rangle$$

with $a, b, c \in \mathbb{Z}/2$. Here and below we assume the abbreviations

$$|a, b\rangle = |a\rangle \otimes |b\rangle \qquad |a^n\rangle = |a\rangle^{\otimes n}.$$

The Toffoli is a unitary dilation of the classical AND gate applied to a and b. In general a dilation D_f of f takes as input $|s\rangle$, n scratch qubits initialized to $|0\rangle$, and a receiving qubit also initialized to $|0\rangle$. (Extra qubits used in dilation are also called *ancillas*.) The output can be written

$$D_f|s, 0^{n+1}\rangle = |s, x(s), f(s)\rangle,$$

where $|x(s)\rangle$ is the scratch work.

Many quantum algorithms require a more convenient dilation of f that we can call *minimal unitary form*. It is a unitary operator U_f that preserves the input and adds the output (in $\mathbb{Z}/2$-arithmetic) to an extra qubit:

$$U_f|s, b\rangle = |s, b + f(b)\rangle.$$

It is implemented as the conjugation

$$U_f = D_f^{-1} \circ N \circ D_f$$

where trivial tensor factors are suppressed and

$$N|f(s), b\rangle = |f(s), f(s) + b\rangle.$$

is a controlled NOT. This trick is also called *uncomputing* the function f.

A related construction is a *phase rotation controlled by* f, denoted $Z_f(\theta)$. It is defined as

$$Z_f(\theta) = D_f^{-1} \circ Z(\theta) \circ D_f,$$

where the phase rotation $Z(\theta)$ acts on $|f(s)\rangle$. It is also given by the formula

$$Z_f(\theta)|s\rangle = e^{if(s)\theta}|s\rangle.$$

One significant use of these unitary tools is Grover's general quantum search algorithm. The input to this algorithm is a black-box function, or oracle,

$$f : S \to \{0, 1\},$$

where S is some finite set. The function f comes with the promise that $f(s) = 1$ for a single $s \in S$ and the task is to find the solution s. Clearly any classical algorithm must evaluate f at least $N/2$ times on average in order to find s. But if the oracle is available in minimal unitary form U_f, then Grover's algorithm can find s using $O(\sqrt{N})$ quantum queries.

For simplicity assume that $S = [0, 2^n)$ for some n. Grover's algorithm is a loop that alternates between the standard qubit basis, $\{|0\rangle, |1\rangle\}$, and the basis $\{|+\rangle, |-\rangle\}$. We can either implement this alternation with Hadamard gates, or assume that quantum gates are available in both bases. The algorithm requires n qubits and is initialized in the state

$$|\psi\rangle = |+^n\rangle.$$

Relative to the standard basis, the qubits are then in the *constant pure state*

$$|S\rangle = \frac{1}{|S|} \sum_{s \in S} |s\rangle.$$

Note that $|S\rangle$ can be defined for any finite orthonormal set of states of any quantum system.

Grover's algorithm then consists of $\lfloor \frac{\pi\sqrt{N}}{4} \rfloor$ iterations of the following two steps, with $|\psi\rangle$ as the state of the qubits at each step.

1. In the standard basis, apply the phase flip $Z_f = Z_f(\pi)$ controlled by f. This reflects the vector $|\psi\rangle$ in the hyperplane perpendicular to $|s\rangle$.
2. By a similar computation in the $\{|+\rangle, |-\rangle\}$ basis, reflect $|\phi\rangle$ through the hyperplane perpendicular to $|+^n\rangle$ and formally negate $|\psi\rangle$.

The effect of both reflections together is to rotate $|\psi\rangle$ from $|+^n\rangle$ to $|s\rangle$ by an angle θ given by

$$\sin\theta = \langle +^n | s \rangle = \frac{1}{\sqrt{N}}.$$

When N is large,

$$\theta \approx \frac{4}{\pi\sqrt{N}}$$

and $|+^n\rangle$ and $|s\rangle$ are nearly orthogonal. Thus the full course of Grover's algorithm brings $|\psi\rangle$ close to the state $|s\rangle$.

3.4. Shor's algorithm

Grover's algorithm and its variations represent one of two main families of existing quantum algorithms. The other family consists of algorithms related to Shor's algorithm for finding the period of a periodic function.

The input to Shor's algorithm is a black-box function

$$f : \mathbb{Z} \to S,$$

where S is some target set. It comes with the promise that

$$f(x) = f(x + p)$$

for some period p, and that otherwise f takes distinct values. The cost of computing $f(x)$ is polynomial in $\log x$, and the task is to find the period p in polynomial time in $\log p$. Since f is given as an oracle, this is classically impossible just from counting necessary queries (exercise **??**). But if the oracle is available in minimal unitary form, then $O(\log p)$ quantum queries suffice. (Technically $O(\log p)$ classical queries suffices, but they must be very large queries.)

The main computational step of Shor's algorithm is the *quantum Fourier transform* on a cyclic group $\mathbb{Z}/N$. This is a unitary operator F_N defined by

$$F_N|x\rangle = \frac{1}{\sqrt{N}} \sum \omega^{xy} |y\rangle$$

with $x, y \in \mathbb{Z}/N$ and

$$\omega = e^{2\pi i/N}.$$

This map has the same formula as the discrete Fourier transform and the algorithm for it when $N = 2^n$ is very similar to the classical FFT. But the interpretation is very different, because the FFT transforms a list of 2^n stored numbers, while the QFT transforms the amplitudes of n stored qubits. (See Exercise **??**.)

To compute F_{2^n}, first express the residue x as a high bit plus a low remainder,

$$x = 2^{n-1} x_{n-1} + x',$$

with $0 \leq x_{n-1} \leq 1$ and $0 \leq x' < 2^{n-1}$. Second, use the remainder x' to control a unitary operator applied to the high qubit $|x_q\rangle$:

$$U(x')|x_{n-1}\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle + (-1)^{x_{n-1}}\omega^{x'}|1\rangle\big) = \frac{|0\rangle + \omega^x|1\rangle}{\sqrt{2}}$$

This operator can be expressed as

$$U(x') = Z(\frac{x_{n-2}\pi}{2})Z(\frac{x_{n-3}\pi}{4})\dots Z(\frac{x_0\pi}{2^{n-1}})H,$$

where x_k is the kth bit of x. In words, $U(x')$ is a Hadamard gate followed by a sequence of phase rotations controlled by each bit of x' separately. To obtain F_n, recursively apply the QFT operator F_{n-1} to $|x'\rangle$ and rechristen the high qubit with input $|x_{n-1}\rangle$ as the low qubit with output $|y_0\rangle$.

Shor's algorithm is as follows. First, choose a number N ... Prepare an integer in the constant pure state $|[0, N)\rangle$, where

$$[0, N) = \{0, \dots, N-1\}.$$

Second, supply this state to the minimal unitary form U_f and discard the output. This step partially measures the input according to the value of $f(x)$, which only depends on the residue of x mod p. The result ρ is a mixture of the corresponding residue classes within $[0, N)$.

The final step of Shor's algorithm is to reveal the residual coherence of ρ with a quantum Fourier transform on $\mathbb{Z}/N = [0, N)$. To understand its effect, first fictitiously suppose that p divides N. (This is of course infeasible given that p is not known.) In this case

$$\rho = \frac{1}{p}\sum_{k=0}^{p-1} |k + p\mathbb{Z}/(N/p)\rangle\langle k + p\mathbb{Z}/(N/p)|$$

and

$$F_N(\rho) = \frac{1}{p}\sum_{k=0}^{p-1} |\frac{kN}{p}\rangle\langle\frac{kN}{p}|.$$

The measurement of a few copies of this state in the standard basis reveals N/p and therefore p.

To use Shor's algorithm to factor a number M, define the periodic function

$$f(n) = a^n$$

for some prime residue $a \in \mathbb{Z}/M$. It can be computed quickly for any fixed n by repeated squaring. Its period divides the exponent of $(\mathbb{Z}/M)^\times$ and once this exponent is known it is easy to factor M.

3.5. Feasibility

Like a bit, a qubit can in principle be any 2-state physical system. The operational difference is that a randomized bit only needs to be suitably independent from other memory bits, while a qubit state decoheres if it is leaked to any physical observer, even an accidental one. A randomized computation can proceed as intended even if its entire state is monitored by the programmer; a quantum computation can be ruined if one qubit is witnessed by a stray atom.

3.6. Error correction

3.7. Quantum secrecy

[1] Markus Arndt, Olaf Nairz, Julian Vos-Andreae, Claudia Keller, Gerbrand van der Zouw, and Anton Zeilinger, *Wave-particle duality of c_{60} molecules*, Nature **401** (1999), 680–682.

[2] Alain Aspect, Jean Dalibard, and Gérard Roger, *Experimental test of bell's inequalities using time-varying analyzers*, Phys. Rev. Lett. **49** (1982), no. 25, 1804–1807.

[3] J. S. Bell, *On the einstein-podolsky-rosen paradox*, Physics **1** (1964), no. 3, 195–200.

[4] ———, *Speakable and unspeakable in quantum mechanics*, Cambridge University Press, 1987, Collected papers on quantum philosophy.

[5] Charles H. Bennett, David P. DiVincenzo, Christopher A. Fuchs, Tal Mor, Eric Rains, Peter W. Shor, John A. Smolin, and William K. Wootters, *Quantum nonlocality without entanglement*, Phys. Rev. A (3) **59** (1999), no. 2, 1070–1091.

[6] Max Born, *Zur quantenmechanik der stossvorgange*, Z. Physik **38** (1926), 803–827.

[7] Man Duen Choi, *Completely positive linear maps on complex matrices*, Linear Algebra and Appl. **10** (1975), 285–290.

[8] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, *Proposed experiment to test local hidden-variable theories*, Phys. Rev. Lett. **23** (1969), no. 15, 880–884.

[9] K. Dennerl, V. Burwitz, J. Englhauser, C. Lisse, and

S. Wolk, *Discovery of x-rays from venus with chandra*, Astronom. and Astrophys. **386** (2002), 319–330.

[10] Paul A. Dirac, *Principles of quantum mechanics*, Oxford University Press, 1930.

[11] Albert Einstein, Boris Podolsky, and Nathan Rosen, *Can quantum-mechanical description of physical reality be considered complete?*, Phys. Rev. **47** (1935), no. 10, 777–780.

[12] Richard P. Feynman, Robert B. Leighton, and Matthew Sands, *The Feynman lectures on physics. Vol. 3: quantum mechanics*, Addison-Wesley, 1965.

[13] K.-E. Hellwig and K. Kraus, *Pure operations and measurements*, Comm. Math. Phys. **11** (1968/1969), 214–220.

[14] ———, *Operations and measurements. II*, Comm. Math. Phys. **16** (1970), 142–147.

[15] A. Jamiołkowski, *Linear transformations which preserve trace and positive semidefiniteness of operators*, Rep. Mathematical Phys. **3** (1972), no. 4, 275–278.

[16] K. Kraus, *General state changes in quantum theory*, Ann. Physics **64** (1971), no. 2, 311–335.

[17] K. M. Lang, S. Nam, J. Aumentado, C. Urbina, and John M. Martinis, *Banishing quasiparticles from josephson-junction qubits: why and how to do it*, IEEE Trans. Appl. Superconduct. **13** (2003), no. 2, 989–993.

[18] Saunders Mac Lane, *Categories for the working mathematician*, second ed., Springer-Verlag, New York, 1998.

[19] Pier Giorgio Merli, Gian Franco Missiroli, and Giulio Pozzi, *On the statistical aspect of electron interference phenomena*, Amer. J. Phys. **44** (1976), no. 3, 306–307.

[20] Michael A. Nielsen and Isaac L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge, 2000.

[21] Jun John Sakurai, *Modern quantum mechanics*, 2nd ed., Benjamin/Cummings, 1985.

[22] D. Salgado, J. L. Sanchez-Gomez, and M. Ferrero, *A simple proof of the Jamiolkowski criterion for complete positivity of linear maps*.

[23] W. Forrest Stinespring, *Positive functions on C^*-algebras*, Proc. Amer. Math. Soc. **6** (1955), 211–216.

[24] John von Neumann, *Wahrscheinlichkeitstheoretischer aufbau der quantenmechanik*, Göttinger Nachr. (1927), 242–272.

[25] Diederik Wiersma and Ad Lagendijk, *Laser action in very white paint*, Phys. World **10** (1997), no. 1, 33–37.