An introduction to quantum probability, quantum mechanics, and quantum computation

Greg Kuperberg*
UC Davis
(Dated: October 8, 2007)

Quantum mechanics is one of the most surprising sides 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 both of these theories, above and beyond their mathematical formalism, have been a great source of ideas in mathematics, even for many questions that have nothing to do with physics or practical statistics. For example, the probabilistic method of Erdős and others [**?**] is a fundamental method in combinatorics to show the existence of combinatorial objects. In principle, the precepts of quantum mechanics could be similarly influential; there could easily be one or more kind of "quantum probabilistic method". But in practice the precepts of quantum mechanics are not very familiar to most mathematicians. Two subdisciplines of mathematics that have assimilated these precepts are mathematical physics and operator algebras. However, much of the intention of mathematical physics is the converse of our purpose, to apply mathematics to problems in physics. The theory of operator algebras is close to the spirit of this article; in this theory what we call quantum probability is often 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, quantum algorithms can be moderately faster or sometimes much faster than their classical and randomized alternatives. 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 technological implications, quantum computation is a beautiful subject that combines mathematics, physics, and computer science.

This article is an introduction to quantum probability theory, quantum mechanics, and quantum computation for the mathematically prepared reader. Chapters **??** and **??** 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 **??**.

This article owes a great debt to the textbook on quantum computation by Nielsen and Chuang [4], and to the Feynman Lectures, Vol. III [2]. Another good textbook written for physics students is by Sakurai [5].

Exercises

These exercises are meant to illustrate how empirical interpretations can lead to solutions of problems in pure mathematics.

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
$$R(n) \geq \frac{2^{(n-1)/2}}{(2(n!))^{1/n}}.$$
(The proof can be described as a couting argument. However, a solution phrased in terms of probabilitistic existence is more in the spirit of these notes.)

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

*Electronic address: greg@math.ucdavis.edu

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.)

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 generalization of classical probability theory that modifies the effects of physical forces. If you have firmly accepted classical probability, it is tempting to suppose that quantum mechanics is a set of probabilistic objects, in effect a special case of probability rather than a generalization. But this is not true in any reasonable sense; quantum probability violates certain inequalities that hold in classical probability (Section **??**). It is also tempting to view quantum mechanics as a a deterministic dynamical system that produces classical probabilities and is otherwise hidden. This interpretation is not reasonable either.

In physics courses, quantum mechanics is usually defined in terms of operators acting on Hilbert spaces. A state of a system is a vector of its Hilbert space, the vector evolves by unitary operators, the vector is measured by Hermitian operators, and the measured values have probability distributions.

Although we will discuss the vector-state model, we will emphasize the non-commutative probability model from operator algebras. In this model, a system can be fully quantum, or fully classical, or things in between. The fully quantum case corresponds to the vector-state model, but even in this case, the general state is described by an operator rather than a vector. The states that can be described by vectors are called pure; the others are mixed states.

The vector-state model of quantum mechanics was originally known as matrix mechanics and is due to Heisenberg. The historical alternative is Schrödinger's wave mechanics. Wave mechanics is best understood as a special case of matrix mechanics, and we will describe it this way. The probabilistic interpretation of quantum mechanics is due to Max Born and is known as the Copenhagen interpretation (Section **??**).

Since classical probability is a major analogy for us, it is reviewed in Section **??**. The point is that a classical probabilistic system (or measurable space) is an algebra of random variables that satisfies relevant axioms. One of the restrictions on the algebra is commutativity: If x and y are two real- or complex-valued random variables, then xy and yx are the same random variable. In quantum probability, this commutative algebra is replaced by a non-commutative algebra called a von Neumann algebra. The remaining definitions stay as much the same as possible.

We will mostly consider finite-dimensional quantum systems. These are enough to show most of the basic ideas of quantum probability, just as finite or combinatorial probability is enough to show most of the basic ideas of classical probability. Infinite-dimensional quantum systems are discussed in Section **??**.

To summarize, quantum probability is the most natural non-commutative generalization of classical probability. In this author's opinion, this description does the most to demystify quantum probability and quantum mechanics.

1.1. Quantum superpositions

We will begin by discussing part of the pure-state model of quantum mechanics in order to show the inadequacy of classical probability.

A pure state of a quantum mechanical system can be described as a vector of a complex vector space $\mathcal{H}$. If the system is finite, then we can say that the vector space is $\mathbb{C}^n$. It will be convenient to label the basis of this vector space by an arbitrary finite set A rather than by the numbers from 1 to n; we can then denote the vector space $\mathbb{C}^A$. The general state space $\mathcal{H}$ is not just a vector space but a Hilbert space, meaning that it has a positive-definite Hermitian inner product $\langle\cdot|\cdot\rangle$. When $\mathcal{H}$ is $\mathbb{C}^n$ or $\mathbb{C}^A$, then it has the standard inner product

$$\langle\phi|\psi\rangle = \sum_{a\in A} \overline{\phi_a}\psi_a.$$

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

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

This notation is due to Dirac [1] and is called "bra-ket" notation. Recall also that a linear map from a Hilbert space to itself is called an *operator*.

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". Classically, 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 **??**). *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 pure states is the vector space $\mathbb{C}^A$. In other words, a state of Alice is a vector

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

with complex coefficients α_a that are called *amplitudes*. The square norm $|\alpha_a|^2$ is interpreted as the probability that Alice is in the configuration $|a\rangle$. The total probability is therefore the sum

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

The state $|\psi\rangle$ is *normalized* if this sum is 1. The phase of α_a (*i.e.*, its argument or angle as a complex number) has no direct probabilistic interpretation, but it becomes important when we consider operators on $|\psi\rangle$. While 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 measurable, *i.e.*, it is not empirical. Indeed, the global phase of $|\psi\rangle$ is absent from the operator formalism that we will define in Section 1.3.

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

If A and B ("Alice" and "Bob") are the configuration sets of two classical systems, then an empirically allowed map from Alice's state to Bob's state is given by a stochastic linear map

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

also called a Markov map. The property that M is linear is the *classical superposition principle*: disjoint probabilities add. In addition, in order to be stochastic, M must have positive entries (so that probabilities remain positive) and its column sums must be 1 (to conserve probability).

In the quantum case, an empirical transition from Alice's vector states to Bob's vector states is a linear map

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

The requirement that U is linear is the *quantum superposition principle*. It appears to contradict the classical superposition principle, and it is thus an apparent paradox of quantum probability. (However, the treatment in Section 1.3 reconciles the two sides of this paradox.) The entries of U are also called amplitudes, just as the entries of a stochastic map are also probabilities. Since we have posited that $|\alpha_a|^2$ is a probability, U conserves total probability if and only if

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

for all $\psi \in \mathbb{C}^A$; *i.e.*, if U is a *unitary embedding*. If $A = B$ or at least $|A| = |B|$, then U is a *unitary operator*.

It will be convenient to consider maps that preserve or decrease probability. Such maps are called *extinction processes*; the model random walks that can terminate, experiments that can be scratched, etc. A classical map M of this kind is called *substochastic*. The corresponding quantum condition is

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

and such as U is *subunitary*.

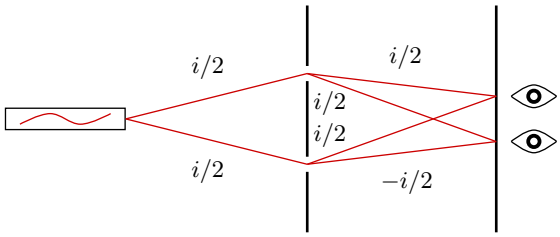

Figure 1: An idealized two-slit experiment.

One traditional, idealized setting for the quantum superposition principle is a diffraction apparatus known as the *two-slit 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

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

We can choose the matrices to be

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

so that

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

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.

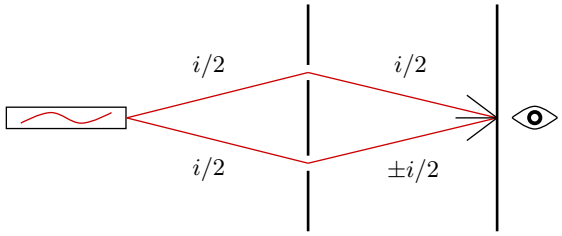

Figure 2: An angle-dependent detector in the two-slit experiment.

A natural reaction to the violation of classical superposition is to try to determine which slit the photon went through. For instance, the detector could be sensitive to the angle that the photon comes in, as in Figure 2. Or there could be a detector at one of the slits that notices that the photon passed through it. But in any such circumstance, the two paths then results in different final states (of the experiment as a whole) rather than in the same state. 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 lesson is that amplitudes of different trajectories of an object only add when there is no evidence of which trajectory it took. If the trajectory is recorded at all, 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 fire only one photon at a time.

Of course, a two-slit experiment is only an idealization of a real experiment; however, it is very similar to many actual experiments and even routine demonstrations. Note also that diffraction experiments can portray any operator and therefore any process in quantum mechanics, just as any classical stochastic map can be modelled by balls falling through chutes. The two-slit experiment can be demonstrated with photons, or electrons or even molecules, but it really describes general probabilistic rules. See Sections 1.5 and **??** for more discussion.

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, neither this configuration set nor any other is preserved by rotations. The resolution of this paradox is that rotated states appear as superpositions. For example, the spin left and spin right states are analogous to $|+\rangle$ and $|-\rangle$:

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

Although we will soon switch to a more general model of quantum probability, we can say a few words about presenting this vector space model in a basis-independent form. As we said, a quantum object can be assigned any Hilbert space $\mathcal{H}$ rather than the standard finite-dimensional vector space $\mathbb{C}^A$ for

[1] More precisely, significant photon-photon interactions require the extreme energies of cosmic rays and particle accelerators.

a configuration set A. Since states evolve by unitary operators, we can conjugate the standard basis of $\mathbb{C}^A$ by any unitary operator, to conclude that any orthonormal basis of any Hilbert space $\mathcal{H}$ can be called a configuration set. Likewise, if we accept a real-valued function f on A as a random variable, then we can model it by a diagonal matrix D who entries are the values of f. We can then conjugate that too by a unitary operator U, to conclude that any Hermitian operator $H = UDU^{-1}$ represents a real-valued random variable. If the eigenvalues of H are 0 and 1, so that $H = P$ is a Hermitian projection, then this corresponds to a Boolean random variable. These are the basic rules of vector-state quantum probability, with the exception of the crucial tensor product rule for joint states (Section 1.5).

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

Since our intention is to generalize classical probability, we will review some of the notions of this theory in the finite case.

A classical probabilistic system is most commonly modelled by a Boolean algebra Ω of random variables. (In the infinite case, it should be a σ-algebra; see Section **??**.) That the algebra is Boolean means that it is an algebra over $\mathbb{Z}/2$, and that every element is an idempotent, $x^2 = x$. The elements of Ω are called *events*. They correspond to random variables that take the values 1 and 0, or equivalently true and false, or yes and no. (Multiplication is Boolean AND, while adding 1 is Boolean complementation.) A *state* or *distribution* of Ω is then a function ρ from Ω to $[0, \infty]$ such that:

- $\rho(x) \geq 0$, and
- $\rho\left(\sum_j x_j\right) = \sum_j \rho(x_j)$ when $x_j x_k = 0$ for all j and k.

The value
$$P[x] = \rho(x)$$
represents the probability of the event x (the probability that x is true). So the axioms say that probabilities are positive, and probabilities of disjoint

events add. The state ρ is *normalized* if $\rho(1) = 1$, which means that the total probability is 1.

If the algebra Ω is finite, then it is isomorphic to $(\mathbb{Z}/2)^A$, the $(\mathbb{Z}/2)$-valued functions on some finite set A or the algebra of subsets of A. The set A is the set of *configurations* of Ω.

The complex-valued random variables over Ω or A form an algebra denoted $L^\infty(\Omega)$ or $\mathbb{C}^A = \ell^\infty(A)$. This algebra is generated (as an algebra over $\mathbb{C}$) by the elements of Ω, with addition in $\mathbb{Z}/2$ forgotten and multiplication retained. In other words, if $xy = z$ in Ω, then this is also imposed as a relation in $L^\infty(\Omega)$. (Technically, $L^\infty(\Omega)$ is only the bounded random variables and is a Banach-space completion of the algebra so generated, but these concerns are only important in the infinite case; see Section **??**.) The state ρ extends to a linear functional on $L^\infty(\Omega)$, so that

$$E[x] = \rho(x)$$

now represents the expected value of x as a random variable. Also, $L^\infty(\Omega)$ has an involution $x \mapsto x^*$ that conjugates the coefficients and values of x and that will be crucial later. The element x^* is called the *adjoint* of x and it is also written $x^\dagger$ in the physics literature.

An equivalent formulation is to write axioms for an algebra $\mathcal{M}$ which can be recognized as $L^\infty(\Omega)$ for some Boolean algebra Ω. In this approach, $\mathcal{M}$ is a *commutative, positive-definite $*$-algebra*. By definition:

- $\mathcal{M}$ is an associative algebra over the complex numbers $\mathbb{C}$.
- $\mathcal{M}$ has an anti-linear, anti-automorphism $*$:

$$(\alpha x)^* = \overline{\alpha} x^* \qquad (x+y)^* = x^* + y^* \qquad (xy)^* = y^* x^*$$

- $\mathcal{M}$ is positive-definite, meaning that if $x^*x = 0$, then $x = 0$.
- $\mathcal{M}$ is commutative; $xy = yx$ for all x and y.

If $\mathcal{M}$ is finite-dimensional, then these axioms imply that $\mathcal{M}$ is isomorphic to $\mathbb{C}^A$ for a finite set A, so that it is indeed equivalent to the other axiom set. (If $\mathcal{M}$ is infinite-dimensional, then these axioms should be strengthened, as we will discuss in Section **??**.)

Here are some other important definitions related to $\mathcal{M}$.

- An element $z \in \mathcal{M}$ is *self-adjoint* if $x = x^*$; it is *positive*, or $x \geq 0$, if $x = y^*y$ for some y; and it is *Boolean* if it is self-adjoint and if $x = x^2$.
- A *state* is a dual vector $\rho \in \mathcal{M}^\#$ which is positive on positive elements: $\rho(x) \geq 0$ if $x \geq 0$. The state ρ is *normalized* if $\rho(1) = 1$. (We write $\mathcal{M}^\#$ instead of $\mathcal{M}^*$ for the dual space because $*$ is already used internally to $\mathcal{M}$.)

If you know or suppose that $\mathcal{M} \cong \mathbb{C}^A$, then it is not hard to show that the self-adjoint elements are the real-valued random variables $\mathbb{R}^A$, the positive elements are the non-negative random variables $\mathbb{R}^A_{\geq 0}$, and the Boolean variables are the $0-1$-valued variables $\{0,1\}^A = (\mathbb{Z}/2)^A = \Omega$.

It is also not hard to show that the two definitions of a state are equivalent. Indeed, the set of normalized states of $\mathcal{M}$ or Ω is the simplex $\Delta_A \subset \mathbb{R}^A$ that consists of convex sums of elements of A. This simplex is shown for a two-state system (a randomized bit) and a three-state system (a randomized trit) in Figure **??**, together with an example element in each case. To support this picture, we define $[a]$ to be the state which is definitely a. For example, if a bit is 1 with probability p, then its state is

$$\rho = (1-p)[0] + p[1].$$

The notion of assigning a state or probability distribution to a probabilistic system has two different empirical interpretations, and the distinction between them will be important. [2] In the *frequentist* interpretation, the state of an object is always a configuration $a \in A$, although you may not know which one; and a distribution ρ is a summary of which configuration you witness in repeated trials. In the *Bayesian* interpretation, the state of an object is a probabilistic state $\rho \in \Delta_A$, which however is observer-dependent; it represents the observer's rational belief about which configuration $a \in A$ will be witnessed, whether or not repeated trials are possible.

Frequentism and Bayesianism are mathematically equivalent. They are only different philosophically, or they may lead to different practical advice. However, quantum probability required a degree of Bayesianism. Although frequentism will remain valid in some contexts, strict frequentism is untenable as the fundamental interpretation of quantum probability. So it is good practice to think of a randomized bit, for example, as living in an intermediate state between 0 and 1, *i.e.*, a classical superposition.

Finally, one fundamental operation on states, especially in the Bayesian interpretation, is the notion

[2] Actually there are several variations of both interpretations in this endless debate in statistics. We have chosen a fairly aggressive flavor of frequentism and a fairly conservative flavor of Bayesianism. This is not entirely fair, but it serves our pedagogical goals.

of a conditional state. If p is a Boolean random variable in $\mathcal{M}$ and ρ is a state, then as we said, the probability of p is $P[p] = \rho(p)$. If p is witnessed by an observer who knows or believes the prior state ρ, then afterwords $\mathcal{M}$ has an updated state ρ_p given by the formula

$$\rho_p(x) = \frac{\rho(px)}{\rho(p)}.$$

This is the state ρ conditioned on p, *i.e.*, what ρ becomes given that p was witnessed. The formula is not meaningful if $\rho(p) = 0$, which is to say, if p is impossible. We also define the unnormalized conditional state

$$\rho|_p(x) = \rho(px), \tag{1}$$

which is well-defined regardless of the probability of p. This state is the empirical posterior state if we view the measurement of p as an extinction process, by declaring extinction if p is false.

Exercises

1.3. Algebras and states

In this section we will define quantum probability as non-commutative probability. This is the other end from Section 1.1; it makes quantum probability seem as similar as possible to classical probability. We will conclude by showing that the the two descriptions are equivalent.

We chose the previous section's axioms for an algebra $\mathcal{M}$ of complex random variables so that they could be made quantum simply by dropping commutativity. To review, $\mathcal{M}$ is a *positive-definite $*$-algebra* if

- $\mathcal{M}$ is an associative algebra over the complex numbers $\mathbb{C}$.
- $\mathcal{M}$ has an anti-linear, anti-automorphism $*$:

$$(\alpha x)^* = \overline{\alpha} x^* \qquad (x+y)^* = x^* + y^* \qquad (xy)^* = y^* x^*$$

- $\mathcal{M}$ is positive-definite, meaning that if $x^*x = 0$, then $x = 0$.

If $\mathcal{M}$ is finite-dimensional, then these axioms are adequate; they are the main definition of a finite quantum system.

We can also repeat these related definitions without changes:

- An element $x \in \mathcal{M}$ is *self-adjoint* if $x = x^*$. Such elements form a real vector space $\mathcal{M}_{\mathrm{sa}}$.
- An element $x \in \mathcal{M}$ is *positive*, or $x \geq 0$, if $x = y^*y$ for some y. If $\mathcal{M}$ is finite-dimensional, the positive elements form a cone $\mathcal{M}_+$.
- An element $p \in \mathcal{M}$ is *Boolean* if it is self-adjoint and if $p = p^2$. Such a p is also called a *self-adjoint projection*. The Boolean elements form a set $\mathcal{M}_{\mathrm{bool}}$.
- A dual vector $\rho \in \mathcal{M}^\#$ has an adjoint defined by $\rho^*(x) = \rho(x^*)$, and it is *self-adjoint* if $\rho^* = \rho$. The set of self-adjoint dual vectors is the real vector space $\mathcal{M}^{\mathrm{sa}}$.
- A *state* is a dual vector $\rho \in \mathcal{M}^\#$ which is positive on positive elements: $\rho(x) \geq 0$ if $x \geq 0$. The set of states is a dual cone $\mathcal{M}^+$.
- The state ρ is *normalized* if $\rho(1) = 1$. The set of normalized states is the *state region* $\mathcal{M}^\Delta$.

These definitions yield the following elementary inclusions:

$$\mathcal{M}_{\mathrm{bool}} \subset \mathcal{M}_+ \subset \mathcal{M}_{\mathrm{sa}} \subset \mathcal{M}$$
$$\mathcal{M}^\Delta \subset \mathcal{M}^+ \subset \mathcal{M}^{\mathrm{sa}} \subset \mathcal{M}^\#.$$

Classically (*i.e.*, if $\mathcal{M}$ is commutative), $\mathcal{M}_{\mathrm{sa}}$ and $\mathcal{M}_{\mathrm{bool}}$ are both closed under multiplication, so that $\mathcal{M}_{\mathrm{sa}}$ is a real algebra and $\mathcal{M}_{\mathrm{bool}}$ is a Boolean algebra. However, quantumly neither one is closed under multiplication, so that at first glance, $\mathcal{M}_{\mathrm{sa}}$ is only a real vector space and $\mathcal{M}_{\mathrm{bool}}$ is only a set. Actually, $\mathcal{M}_{\mathrm{bool}}$ has somewhat more structure than that; see Exercise **??**. The most important extra structure at the moment is that $\mathcal{M}$ is partially ordered with respect to positivity: $x \geq y$ if $x - y \geq 0$. The inherited partial orderings of $\mathcal{M}_{\mathrm{sa}}$ and $\mathcal{M}_{\mathrm{bool}}$ are both important.

If $\mathcal{M}$ is finite-dimensional, then we can classify its structure using the Artin-Schreier theorem, because its positive-definite structure implies that it is semisimple (Exercise **??**). Since the complex numbers are algebraically closed, the theorem says that $\mathcal{M}$ is isomorphic to a direct sum of matrix algebras:

$$\mathcal{M} \cong \bigoplus_k \mathcal{M}_{n_k}.$$

In particular, if $\mathcal{M}$ is a matrix algebra $\mathcal{M}_n$, then it is as non-commutative as possible. We will call such an $\mathcal{M}$ and the system that it models *fully quantum*. In basis-independent form, a fully quantum system $\mathcal{M}$ is the algebra $\mathcal{B}(\mathcal{H})$ of operators on a finite-dimensional Hilbert space $\mathcal{H}$. (The "$\mathcal{B}$" is for "bounded", although in this context all operators are bounded; see Section **??**.)

If $\mathcal{M}$ is fully quantum, then we can use the matrix trace to convert a state ρ from a dual vector on $\mathcal{M}$ to an element.

$$\rho(x) = \mathrm{Tr}(\rho x).$$

(Actually this works in general, using the sum of the traces of the matrix summands.) Then ρ is positive as a dual vector if and only if it is positive as an element of $\mathcal{M}$, if and only if it is a positive-definite Hermitian matrix (Exercise **??**). Also ρ is normalized if and only if $\mathrm{Tr}(\rho) = 1$. Because such a ρ is a matrix and because its diagonal entries are probabilities, physicists also call it a *density matrix* or (in basis-independent form) a *density operator*. In this terminology, "density"means probability density, as in a probability distribution. The diagonal entries of a density matrix are in fact probabilities of configuration (Section **??**).

Example 1.3.1. The 2×2 matrix algebra $\mathcal{M}_2$, or a system that it models, is a second and better definition of a *qubit*. The Pauli spin matrices are a convenient basis for $(\mathcal{M}_2)_{\mathrm{sa}}$:

$$\sigma_0 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \sigma_1 = X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$\sigma_2 = Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

A state ρ of $\mathcal{M}_2$ is positive and normalized if and only if it is of the form

$$\rho = \frac{I + aX + bY + cZ}{2},$$

where a, b, and c are three real numbers that satisfy

$$a^2 + b^2 + c^2 \leq 1.$$

The state region $\mathcal{M}_2^\Delta$ is therefore a geometric sphere in the affine space of unit-trace, 2×2 Hermitian matrices. It is known as the *Bloch sphere*.

As in the example of a qubit, an important difference between quantum probability and classical probability is that the state region $\mathcal{M}^\Delta$ is not a simplex (except in the commutative case). But it is always convex, because it is defined by linear equalities and inequalities. This convex structure allows classical superpositions in a quantum setting. Empirically, if we have two states ρ_1 and ρ_2 of a quantum system, and if we prepare a new state ρ by choosing ρ_1 with probability p and ρ_2 with probability $1 - p$, then

$$\rho = p\rho_1 + (1 - p)\rho_2.$$

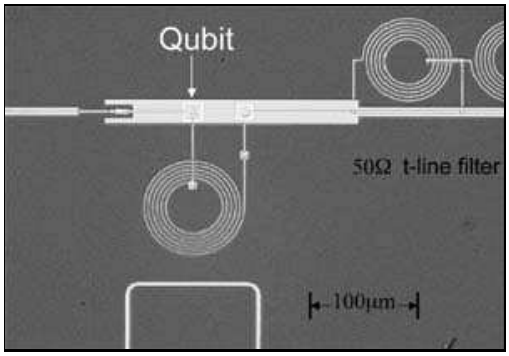

Figure 3: A Josephson junction qubit: superconducting aluminum on a silicon chip [3].

(This formula is consistent with the fact that all probabilities are linear in ρ.)

But what about quantum superpositions? If $\mathcal{M}$ is fully quantum, then they are also present in different guise from classical superpositions. To help separate the terminology, classical superpositions are in general called *mixtures*, while quantum superpositions (when they are defined) are often just called superpositions.

In general, if K is a convex set in a real vector space, then a point in K is *extremal* means that it is not a convex combination of two other points in K. An elementary theorem in convex geometry states that if K is compact and finite-dimensional, then every point in K is a convex linear combination of its extremal points. If $K = \mathcal{M}^\Delta$, then the extremal points are those states that are not mixtures. These states are called *pure* and other states are called *mixed*. By the theorem, every mixed state is a mixture of pure states. However, in a fully quantum system, the representation of a state as a mixture is never unique (Exercise **??**).

If $\mathcal{M} = \mathcal{M}_n$ is fully quantum, then a state ρ is pure if and only if it has rank 1 as a matrix (Exercise **??**). It then has the form

$$\rho = \psi \otimes \psi^* = |\psi\rangle\langle\psi|$$

for some vector $\psi \in \mathbb{C}^n$, since it is also Hermitian. If ρ is normalized, then in addition ψ is normalized, by the relation

$$\mathrm{Tr}\rho = \langle\psi|\psi\rangle.$$

In basis-independent form, if $\mathcal{M} = \mathcal{B}(\mathcal{H})$, and if a state ρ on $\mathcal{M}$ is pure, then it is described by a vector $|\psi\rangle \in \mathcal{H}$. A *configuration set* of $\mathcal{M}$ is, by definition, any orthonormal basis of $\mathcal{H}$. Any state $|\psi\rangle$ is a complex linear combination of the configurations, and such a linear combination can be called a *quantum superposition*.

To summarize, a pure state of a fully quantum $\mathcal{M}$ is represented by a vector in a Hilbert space, and it is a quantum superposition of any orthonormal basis of configurations. However, the transformation from a vector state $|\psi\rangle$ to the corresponding density operator $\rho = |\psi\rangle\langle\psi|$ is non-linear and erases the global phase of $|\psi\rangle$. Therefore empirical probabilities are a non-linear function of the vector state, and the global phase of a vector state is not directly empirical. (But relative phases are empirical, so the global phase of $|\psi\rangle$ is indirectly empirical, if $|\psi\rangle$ is used as a summand of another vector.)

Example 1.3.2. Since the state region $\mathcal{M}_2^\Delta$ of a qubit is a geometric sphere, every boundary point is a pure state $|\psi\rangle\langle\psi|$. It is not hard to check that any two opposite points form a (line) basis of the Hilbert space $\mathbb{C}^2$. In the context of quantum computation, the standard basis of $\mathbb{C}^2$ is called $|0\rangle$ and $|1\rangle$, and these states are assigned to the top and bottom of the Bloch sphere, as in Figure **??**. Another basis is

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

These states

In the middle is the *uniform state* (also called the *maximally mixed* or *maximum entropy* state) $\rho = I/2$.

Although probabilities are nonlinear functions of vector states $|\psi\rangle$, there are several important operations on a quantum system $\mathcal{M}$ which are linear on vector states. We can call such operations *coherent*; they serve to justify the quantum superposition principle in Section 1.1.

The most important coherent operation is an algebra isomorphism. Automorphisms and isomorphisms in general are the model of reverse dynamical systems and reversible physical transformations. The wrinkle is that algebra isomorphisms, and more generally algebra homomorphisms, are contravariant, meaning that they transfer states backward. If Alice and Bob have algebras $\mathcal{M}_A$ and $\mathcal{M}_B$, then a homomorphism

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

transfers states from Alice to Bob by means of its transpose:

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

If Alice and Bob are both fully quantum and have Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ of the same dimension, then every algebra isomorphism

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

is given by a unitary operator $u : \mathcal{H}_A \to \mathcal{H}_B$ by the formula

$$E(x) = uxu^{-1}.$$

In more concrete terms, the automorphism group of $\mathcal{M}_n$ as a $*$-algebra is the unitary group $U(n)$ (Exercise **??**). Note that u is conventionally covariant, so that it transfers pure states forward, from Alice to Bob.

Note also that this motivation for unitary operators does not support the analogy with Markov maps in Section 1.1. Rather, unitary operators are analogous to permutations of configurations of a classical system, since these are the reversible maps among Markov maps. This is another way to say that we have resolved the paradox of that section: Classical and quantum superposition do not contradict each other because in some ways, they are not analogous. Section **??** discusses the correct notion of quantum Markov maps, namely quantum operations. Classical and quantum superposition coexist for quantum operations, just as they do for states in the operator formalism.

Another coherent operation is conditioning a state with a Boolean random variable. If ρ is a state of $\mathcal{M}$ and $p \in \mathcal{M}_{\text{bool}}$ is Boolean, then the unnormalized conditional state is defined by

$$\rho|_p(x) = \rho(pxp).$$

This reduces to equation (**??**) when $\mathcal{M}$ is commutative, but it cannot be exactly the same formula as before because $\rho(px)$ is not positive as a dual vector in x, nor even self-adjoint. It is easy to check that if $\mathcal{M}$ is fully quantum, then the pure state $|\psi\rangle$ conditions to the pure state $p|\psi\rangle$. So conditioning a state is linear on pure states.

With a bit of modification, unitary operators and projections generate all subunitary maps between any two Hilbert spaces (Exercise **??**). These are all of the linear operators, or coherent operations, used in Section **??**. Their construction establishes the quantum superposition principle as a corollary of non-commutativity. In Section **??**, quantum superpositions appear in another way, as a corollary of the classical superposition principle.

Exercises

1.4. Measurements

In this section we will look more closely at the measuring quantum random variables. A random variable $x \in \mathcal{M}_{\text{sa}}$ (or more generally a classical domain $\mathcal{A} \subset \mathcal{M}$ as defined below) is also called a *measurable* or an *observable*. To *measure* it is to pass

to the conditional state, just as is done in classical probability.

We said that if p is a Boolean random variable in an algebra $\mathcal{M}$, then the unnormalized conditional state is

$$\rho|_p(x) = \rho(pxp).$$

The normalized conditional state is thus

$$\rho_p(x) = \frac{\rho(pxp)}{\rho(p)},$$

or in the vector-state case,

$$|\psi_p\rangle = \frac{p|\psi\rangle}{\sqrt{\langle\psi|p|\psi\rangle}}.$$

Conditioning on a measurement is also called "state collapse" or "wave function collapse", but in the context of non-commutative probability, this is an overly dramatic term. The concept of a conditional state is very natural in classical probability; and it is equally natural and not all that different in quantum probability. See Section **??**.

The behavior of non-commuting random variables ...

Example 1.4.1.

If any set of Boolean variables in $\mathcal{M}$ all commute, or indeed if any set of self-adjoint elements in $\mathcal{M}$ all commute, then they generate a commutative, positive-definite $*$-subalgebra $\mathcal{A} \subseteq \mathcal{M}$. We will call such an $\mathcal{A}$ a *classical realm* in $\mathcal{M}$. Elements in $\mathcal{A}$ are elements in $\mathcal{M}$, and states on $\mathcal{M}$ restrict to states on $\mathcal{A}$. While we work within $\mathcal{A}$, we are free to use any notion or result from classical probability without modification. In particular, the elements that we chose to generate $\mathcal{A}$ have a joint distribution, and the order that they are measured does not matter.

If a classical realm $\mathcal{A} \subseteq \mathcal{M}$ is finite-dimensional, then it is isomorphic to $\mathbb{C}^A$ for some set A. Thus a state ρ on $\mathcal{M}$ induces a probability for each outcome $a \in A$. Moreover, $\mathcal{A}$ has a basis $\{p_a\}$ of minimal projections indexed by the set A. We can then say that $\mathcal{A}$ is a model of an A-valued random variable with postconditioned states, using the same formulas as for a single Boolean variable:

$$P[a] = \rho(p_a) \qquad \rho_a(x) = \frac{\rho(p_a x p_a)}{\rho(p_a)}.$$

We can also run the construction backwards to build $\mathcal{A}$ from its minimal projections. Say that two Booleans p and q are *mutually exclusive* if $pq = 0$ (so that they necessarily commute). An A-valued random variable is then in general defined by a set of mutually exclusive Booleans that sum to 1:

$$\sum_{a\in A} p_a = 1 \qquad a \neq b \implies p_a p_b = 0.$$

If $\mathcal{M} = \mathcal{B}(\mathcal{H})$ is fully quantum, then this system of projections is equivalent to an orthogonal direct sum decomposition of the Hilbert space $\mathcal{H}$:

$$\mathcal{H} = \bigoplus_{a\in A} \mathcal{H}_a.$$

It will also be convenient to generalize a classical realm to allow some of the projections p_a to vanish. This is equivalent to making the realm a homomorphism $\mathcal{A} \to \mathcal{M}$ rather than a subalgebra. Section **??** discusses a much more significant generalization known as a POVM.

The most important case of a classical realm $\mathcal{A}$ is one generated by a single self-adjoint element $x \in \mathcal{M}_{\text{sa}}$. In this case the structure of $\mathcal{A}$ implies that x has a spectral decomposition,

$$x = \sum_{\lambda\in\sigma(x)} \lambda p_\lambda,$$

where the value set of the measurement, $A = \sigma(x)$, is also called the spectrum of x. The probability formula,

$$P[x = \lambda] = \rho(p_\lambda),$$

is then consistent with the expectation interpretation of the state ρ,

$$E[x] = \rho(x).$$

So $\mathcal{M}_{\text{sa}}$ is the space of real-valued random variables, just as it was classically. (Note that if $\mathcal{M}$ is fully quantum, then the structure theorem for this $\mathcal{A}$ is equivalent to the spectral theorem for Hermitian matrices.)

Another important type of classical realm $\mathcal{A}$ is a maximal commutative $*$-subalgebra of $\mathcal{M}$. (By definition, $\mathcal{A}$ is not contained in any commutative $*$-subalgebra $\mathcal{B}$.) If $\mathcal{M} = \mathcal{B}(\mathcal{H})$ is fully quantum, then it is easy to show that $\mathcal{A}$ is maximal if and only if $\mathcal{A}$ and $\mathcal{H}$ have the same dimension (Exercise **??**); indeed $\mathcal{A}$ consists of the diagonal matrices with respect to some orthonormal basis A of $\mathcal{H}$. Each minimal projection p_a of $\mathcal{A}$ has rank 1. It follows that the conditional state ρ_a does not depend on ρ; it is always the state

$$\rho_a = p_a = |a\rangle\langle a|.$$

Although the basis A need only be a line basis of $\mathcal{H}$, it is often convenient to make it a vector basis, so

that $\mathcal{H} = \mathbb{C}^A$. The set A is a *configuration set*, in keeping with Section 1.1. If $n = |A| = \dim \mathcal{H}$, then we say that $\mathcal{M}$ is an *n-state system*, even though technically n is the number of configurations rather than the number of states. For example, a qubit can also be described as any fully quantum 2-state system.

A maximal classical realm $\mathcal{A}$ is also called a *complete measurement*. The name evokes the fact that once any configuration $a \in A$ is measured, the conditional state is pure and determined by a, so there is no more left to learn from the state of $\mathcal{M}$. (Nonetheless, $\mathcal{M}$ has many different complete measurements; and as we said, even a pure state is typically still a source of perpetual randomness.) Note also that if $\mathcal{M}$ is fully quantum and we have chosen a basis so that $\mathcal{A}$ consists of diagonal matrices, then the diagonal entry ρ_{aa} of a state ρ is just the probability of the outcome $a \in A$. We can view ρ as a classical probability distribution on A, plus extra off-diagonal information.

If $\mathcal{M}$ is not fully quantum, then some of the above analysis has to be modified. Nonetheless, it is still true that all of the conditional states of a maximal classical realm $\mathcal{A}$ are pure, that a set A of such outcomes is called a configuration set, and that any two configuration sets have the same cardinality. If

$$\mathcal{M} \cong \bigoplus_k \mathcal{M}_{n_k},$$

then the cardinality of A is the total sum of the matrix sizes, $n = \sum_k n_k$.

Similar to a complete measurement, if ρ is a pure state, then there is an associated minimal Boolean p with the same matrix as ρ which answers whether the system is in the state ρ. If p and q are two such minimal Booleans, then $\mathrm{Tr}(pq)$ is both the probability that the state p will be found in the state q, and vice-versa; it can be called the *overlap* between p and q. If $\mathcal{M}$ is fully quantum, so that p and q have state vectors $|a\rangle$ and $|b\rangle$, then their overlap is $|\langle a|b\rangle|^2$. Two pure states are mutually exclusive if and only if they have no overlap (Exercise **??**).

Unlike real-valued random variables, there are two notions of complex-valued and vector-valued random variables. We can let z be any element of $\mathcal{M}$, which is the complexification of $\mathcal{M}_{\mathrm{sa}}$, so that

$$z = x + iy \qquad x = \frac{z + z^*}{2} \qquad y = \frac{z - z^*}{2i}.$$

Then z is a complex random variable in a weak sense, because x and y are both self-adjoint and are both therefore real random variables. This defines a complex random variable in the weak sense. The wrinkle is that x and y may not commute with each other, in which case z and a state ρ do not yield a distribution on $\mathbb{C}$. If the real and imaginary parts x and y do commute, or equivalently if z and z^* commute, then z is *normal*. A state ρ and a normal z generate a classical realm and a classical state on $\mathbb{C}$ as usual.

Likewise a vector-valued random variable in the weak sense is any $\vec{v} \in \mathcal{M} \otimes V$ for some vector space V. We can say that $\vec{v}$ is normal when its components commute in any basis of V, in which case it generates a classical realm and a classical state on V, given a state ρ on $\mathcal{M}$. An important example of the weak kind of vector-valued random variable is the angular momentum operator (Section **??**).

Note the set $\mathcal{M}_{\mathrm{nor}}$ of normal elements of $\mathcal{M}$ is not closed under either addition or multiplication (Exercise **??**). The same is true of the set $(\mathcal{M} \otimes V)_{\mathrm{nor}}$ of vector-valued measurements, or in general the A-valued measurements where the set A is an abelian group. Only the real random variables, $\mathcal{M}_{\mathrm{sa}}$, have the special property that they can be added even if they do not commute.

1.4.1. Exercises

1.5. Joint systems

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

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

[3] 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.

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

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