

Quantum Computation Lecture Notes

Jyh Ying Peng
(based on lecture notes by John Preskill 1998)

Summer 2003

Chapter 1

Quantum States and Ensembles

1.1 Axioms of Quantum Mechanics

Quantum theory (as all physical theories) can be characterized by how it represents (physical) states, observables, measurements, and dynamics (evolution in time).

1. **States.** A state is a complete description of a physical system. In quantum mechanics, a state is a **ray** in a **Hilbert space**.

A Hilbert space is a vector space over the field of complex numbers \mathbf{C} , with vectors denoted by $|\psi\rangle$ (Dirac's ket notation). A ket $|\psi\rangle$ is represented as a $n \times 1$ matrix (or a n element vector), where n is the dimension of the Hilbert space, and its corresponding bra $\langle\psi|$ is the transpose conjugate of the ket. It has an inner product $\langle\psi|\varphi\rangle$ (may be seen as matrix multiplication) that maps an ordered pair of vectors to \mathbf{C} , with the properties:

- (a) Positivity. $\langle\psi|\psi\rangle > 0$ for $|\psi\rangle \neq 0$
- (b) Linearity. $\langle\varphi|(a|\psi_1\rangle + b|\psi_2\rangle) = a\langle\varphi|\psi_1\rangle + b\langle\varphi|\psi_2\rangle$
- (c) Skew symmetry. $\langle\varphi|\psi\rangle = \langle\psi|\varphi\rangle^*$

The Hilbert space is **complete** in the norm $\|\psi\| = \langle\psi|\psi\rangle^{\frac{1}{2}}$.

A ray in a Hilbert space is an equivalent class of vectors that differ by multiplication by a nonzero complex scalar. That is, a ray is represented by a given vector and all its (complex) multiples, and two different rays treated as vectors will never be “parallel” (in the Hilbert space). We represent rays (the equivalent classes) by a vector with unit norm $\langle\psi|\psi\rangle = 1$, so $e^{i\alpha}|\psi\rangle$ (where $\alpha \in \mathfrak{R}$) represent the same physical

state for all α . The real number α can be seen as an **overall phase**, which is physically insignificant. We can form new states by **superposition** $a|\varphi\rangle + b|\psi\rangle$, and here the **relative phase** between the two components *are* physically significant. That is, whereas $a|\varphi\rangle + b|\psi\rangle$ and $e^{i\alpha}(a|\varphi\rangle + b|\psi\rangle)$ represent the same physical state, $a|\varphi\rangle + e^{i\alpha}b|\psi\rangle$ is generally a different physical state.

2. **Observables.** An observable is a property of a physical system that in principle can be measured. In quantum mechanics, an observable is a **self-adjoint operator (matrix)**. An operator is a linear map taking vectors to vectors, which can be represented by matrices. For an operator \mathbf{A} , $\mathbf{A}: |\psi\rangle \rightarrow \mathbf{A}|\psi\rangle$ and $\mathbf{A}(a|\psi\rangle + b|\varphi\rangle) = a\mathbf{A}|\psi\rangle + b\mathbf{A}|\varphi\rangle$. And the adjoint of an operator \mathbf{A}^\dagger is defined by $\langle\varphi|\mathbf{A}\psi\rangle = \langle\mathbf{A}^\dagger\varphi|\psi\rangle$, where $|\mathbf{A}\psi\rangle$ denotes $\mathbf{A}|\psi\rangle$, for all vectors $|\psi\rangle, |\varphi\rangle$. In matrix representation the adjoint is the transpose conjugate. \mathbf{A} is self-adjoint if $\mathbf{A} = \mathbf{A}^\dagger$. The eigenstates of an observable (eigenvectors of the corresponding self-adjoint matrix) form a complete orthonormal basis in the Hilbert space \mathcal{H} . We can express an observable \mathbf{A} as $\mathbf{A} = \sum_n a_n \mathbf{P}_n$, where each a_n is an eigenvalue of \mathbf{A} , and \mathbf{P}_n is the orthogonal projection to the space spanned by the corresponding eigenvectors. The \mathbf{P}_n 's are self-adjoint and satisfy $\mathbf{P}_n \mathbf{P}_m = \delta_{nm} \mathbf{P}_n$, $\mathbf{P}_n^\dagger = \mathbf{P}_n$.
3. **Measurements.** In quantum mechanics, the numerical outcome of the measurement of the observable \mathbf{A} is an eigenvalue of \mathbf{A} , and right after the measurement, the quantum state becomes the eigenstate of \mathbf{A} corresponding to the measurement result. If the quantum state before measurement is $|\psi\rangle$, then outcome a_n is obtained with probability $\text{Prob}(a_n) = \|\mathbf{P}_n|\psi\rangle\|^2 = \langle\psi|\mathbf{P}_n|\psi\rangle$, and the (normalized) quantum state becomes in this case

$$\frac{\mathbf{P}_n|\psi\rangle}{\langle\psi|\mathbf{P}_n|\psi\rangle^{\frac{1}{2}}}. \quad (1.1)$$

Which means if the measurement is immediately repeated, the same result would be obtained with probability one.

4. **Dynamics.** Time evolution of a quantum state is unitary (unitary transformation can be seen as a rotation in Hilbert space); it is generated by a self-adjoint operator, called the **Hamiltonian** of the system. In the **Schrödinger picture** of dynamics, the vector (state) describing the system evolves in time according to the **Schrödinger equation** $\frac{d}{dt}|\psi(t)\rangle = -i\mathbf{H}|\psi(t)\rangle$, where \mathbf{H} is the Hamiltonian. Using the definition of function derivatives, the equation can be reexpressed as $|\psi(t + dt)\rangle = (\mathbf{1} - i\mathbf{H}dt)|\psi(t)\rangle$. The operator $\mathbf{U}(dt) \equiv$

$\mathbf{1} - i\mathbf{H}dt$ is unitary to linear order in dt , because $\mathbf{U}(dt)^\dagger\mathbf{U}(dt) = (\mathbf{1} + i\mathbf{H}dt)(\mathbf{1} - i\mathbf{H}dt) = \mathbf{1} + (\mathbf{H}dt)^2 \approx \mathbf{1}$. Since a product of unitary operators is finite, time evolution over a finite interval is also unitary $|\psi(t)\rangle = \mathbf{U}(t)|\psi(0)\rangle$. If the Hamiltonian is time-independent, the Schrödinger equation can be directly solved to give $\mathbf{U}(t) = e^{-it\mathbf{H}}$ (which is likewise unitary).

In the formulations in this section, there is an obvious dualism between how a quantum state evolves when “left to itself”, and when it has been “measured” by “something”. In the former case the state evolves according to the Schrödinger equation, which is deterministic; whereas in the latter case the evolution is probabilistic. Even without the discrepancy between different evolutions, the existence of undeterministic physical processes is itself troubling. We will attempt an explanation of the matter in the next chapter, where the quantum measurement process is explained in detail.

1.2 The Qubit

The indivisible unit of classical information is the **bit**, which can take one of two values $\{0, 1\}$. The corresponding unit of quantum information is the “quantum bit” or **qubit**, which can be represented as a ray in 2-d Hilbert space. We may denote an orthonormal basis for this Hilbert space $\{|0\rangle, |1\rangle\}$, any state of a qubit may be expressed as

$$a|0\rangle + b|1\rangle, \quad (1.2)$$

with $|a|^2 + |b|^2 = 1$ (normalization). As noted in section 1.1, the overall phase is irrelevant. Since $a, b \in \mathbf{C}$, a qubit can be specified by four real parameters, but normalization and the irrelevance of overall phase reduce this to two real parameters. This means that a qubit can also be represented as a vector in a 2-d real vector space with suitable range limitations.

We can perform a measurement that projects the unknown state of a qubit onto the basis $\{|0\rangle, |1\rangle\}$. Then $|0\rangle$ and $|1\rangle$ is obtained with probability $|a|^2$ and $|b|^2$ respectively. The state would be disturbed unless a or b is initially zero, and the prior state of the qubit cannot be known by any measurement. The fact that the state would be known after measurement implies that a “measurement” can also be seen as a “preparation” of a particular quantum state. In fact, data is *erased* when the qubit is measured (prepared), much as we “erase” classical bits by writing zeros in them, ignoring their initial values. But with classical bits we can acquire information about their values easily without disturbing their states.

Consider a probabilistic classical bit, it has a definite value, unknown to us. All we know is that there is a probability p_0 for it to be 0, and probability p_1 to be 1, where the probabilities sum to one. What is the difference between a probabilistic bit and a qubit? One of the differences is that besides the absolute values of a and b in eq. (1.2), their relative phase is also of significance. This is reflected in the fact that a qubit takes two real parameters to specify, whereas a probabilistic classical bit need only one.

1.3 Representations of the Qubit

In this section we seek to find useful (mathematical) representations of the state and evolution of a qubit, based on real physical systems. We discuss two such systems, the spin states of a spin- $\frac{1}{2}$ object and the polarization states of a photon.

1.3.1 Spin- $\frac{1}{2}$

In physics, the qubit can be interpreted as the spin state of an object with spin- $\frac{1}{2}$ (such as an electron). Then $|0\rangle$ and $|1\rangle$ in eq. (1.2) can be represented by the spin up $|\uparrow\rangle$ and spin down $|\downarrow\rangle$ states along a particular axis in the physical spin- $\frac{1}{2}$ system (the z-axis will generally be used). Speaking geometrically in the 3-d real vector space, by setting

$$a = e^{-\frac{i\varphi}{2}} \cos \frac{\theta}{2} \text{ and } b = e^{\frac{i\varphi}{2}} \sin \frac{\theta}{2} \quad (1.3)$$

in eq. (1.2), and using spin along z-axis as basis, the state of a qubit can be represented as a unit vector in a 3-d real vector space, parameterized by (θ, φ) , where θ is the angle between the vector and z-axis (the polar angle), and φ is the angle between the x-axis and the vector's projection onto x-y plane (the azimuthal angle). We see that the angle between the z-axis and the vector determines the probabilities of obtaining spin up or spin down along the z-axis, and the azimuthal angle represents the relative phase. The vector points in +z when $a = 1, b = 0$ and -z when $a = 0, b = 1$, as expected. So we have established a correspondence between the general representation in 2-d Hilbert space of a qubit and representation in 3-d real vector space of the spin state of a spin- $\frac{1}{2}$ object, which means that the spin state of any spin- $\frac{1}{2}$ object can in principle be used to physically implement a quantum bit.

To describe the time evolution and measurement of a quantum bit, we also need to represent a general unitary transformation (rotation) of the

state. Using 2-d Hilbert space vector representation, the rotation of the spin state of a spin- $\frac{1}{2}$ object can be represented with the use of complex 2×2 Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (1.4)$$

A rotation in 3-d real vector space through angle θ about the axis $\hat{n} = (n_1, n_2, n_3)$ (where $n_1^2 + n_2^2 + n_3^2 = 1$), represented (in 2-d Hilbert space) as a 2×2 unitary matrix, is

$$\mathbf{U}(\hat{n}, \theta) = \exp\left(-i\frac{\theta}{2}\hat{n} \cdot \vec{\sigma}\right) = \exp\left(-i\frac{\theta}{2}(n_1\sigma_1 + n_2\sigma_2 + n_3\sigma_3)\right). \quad (1.5)$$

In eq. (1.5) we parameterized a unitary transformation in 2-d Hilbert space by the corresponding rotation in 3-d real vector space. Since the 3-d real vector space rotation parameters \hat{n} and θ describes the most general rotation or unitary transformation in 3-d real vector space, we expect $\mathbf{U}(\hat{n}, \theta)$ to represent the most general unitary transformation in the corresponding 2-d Hilbert space.

The Pauli matrices have the properties of being mutually anti-commuting and squaring to the identity:

$$\sigma_k\sigma_l + \sigma_l\sigma_k = 2\delta_{kl}\mathbf{1}, \quad (1.6)$$

so we see that $(\hat{n} \cdot \vec{\sigma})^2 = \sum_{k,l} n_k n_l \sigma_k \sigma_l = \sum_k n_k^2 \mathbf{1} = \mathbf{1}$, and for finite rotations (unitary transformations)

$$\begin{aligned} \mathbf{U}(\hat{n}, \theta) &= \exp\left(-i\frac{\theta}{2}\hat{n} \cdot \vec{\sigma}\right) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \left(-i\frac{\theta}{2}\hat{n} \cdot \vec{\sigma}\right)^n \\ &= \sum_{n=0,2,4,\dots} \frac{1}{n!} \left((-1)^{\frac{n}{2}} \left(\frac{\theta}{2}\right)^n \mathbf{1}\right) + \sum_{n=1,3,5,\dots} \frac{1}{n!} \left(-i(-1)^{\frac{n-1}{2}} \left(\frac{\theta}{2}\right)^n \hat{n} \cdot \vec{\sigma}\right) \\ &= \mathbf{1} \cos \frac{\theta}{2} - i\hat{n} \cdot \vec{\sigma} \sin \frac{\theta}{2}. \end{aligned} \quad (1.7)$$

This is indeed the form of the most general 2×2 unitary matrix with determinant 1.

A rotation by 2π about any axis is $\mathbf{U}(\hat{n}, 2\pi) = -\mathbf{1}$, yet in 3-d real vector space this would be an identity transformation, so shouldn't $\mathbf{U} = \mathbf{1}$? But there is nothing wrong here if we notice that $-\mathbf{1}$ only changes the overall

phase, and that is physically irrelevant, so the physical state remains unchanged. Yet when such an operation is applied to more than one qubit, there may be a physically detectable effect. Consider an operation that acts on two qubits by rotating the second qubit by 2π when the first qubit is spin down, but otherwise does nothing. Then

$$\begin{aligned} & \frac{1}{\sqrt{2}} (|\uparrow\rangle_1 + |\downarrow\rangle_1) \otimes |\uparrow\rangle_2 = \frac{1}{\sqrt{2}} (|\uparrow\rangle_1 \otimes |\uparrow\rangle_2 + |\downarrow\rangle_1 \otimes |\uparrow\rangle_2) \\ \rightarrow & \frac{1}{\sqrt{2}} (|\uparrow\rangle_1 \otimes |\uparrow\rangle_2 - |\downarrow\rangle_1 \otimes |\uparrow\rangle_2) = \frac{1}{\sqrt{2}} (|\uparrow\rangle_1 - |\downarrow\rangle_1) \otimes |\uparrow\rangle_2. \end{aligned}$$

The first qubit is initially in an equal superposition between spin up and spin down, and the second qubit is spin up, the whole system is thus in an equal superposition between $|\uparrow\rangle_1|\uparrow\rangle_2$ and $|\downarrow\rangle_1|\uparrow\rangle_2$. When the operation is applied *on the whole system*, the second component undergoes a sign change (phase flip, the phase is changed by π), so the relative phase between the two components is changed.

Using eq. (1.3) and setting $|0\rangle \equiv |\uparrow_z\rangle \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|1\rangle \equiv |\downarrow_z\rangle \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, we see that spin up and spin down along x-axis ($\theta = \frac{\pi}{2}, \varphi = 0$ and $\theta = \frac{\pi}{2}, \varphi = \pi$ respectively) is

$$|\uparrow_x\rangle = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle), |\downarrow_x\rangle = \frac{1}{\sqrt{2}} (|0\rangle - |1\rangle). \quad (1.8)$$

And the spin up and spin down along y-axis ($\theta = \frac{\pi}{2}, \varphi = \frac{\pi}{2}$ and $\theta = \frac{\pi}{2}, \varphi = -\frac{\pi}{2}$ respectively) is

$$|\uparrow_y\rangle = \frac{1}{\sqrt{2}} (|0\rangle + i|1\rangle), |\downarrow_y\rangle = \frac{1}{\sqrt{2}} (|0\rangle - i|1\rangle). \quad (1.9)$$

Note that the Pauli matrices σ_1, σ_2 , and σ_3 in eq. (1.4) are self-adjoint and have the spin up and spin down states along x-axis, y-axis, and z-axis as eigenstates respectively, with eigenvalue 1 for the spin up states and -1 for the spin down states. So the eigenstates of matrix $\hat{n} \cdot \vec{\sigma} = n_1\sigma_1 + n_2\sigma_2 + n_3\sigma_3$ are $|\uparrow_{\hat{n}}\rangle$ with eigenvalue 1, and $|\downarrow_{\hat{n}}\rangle$ with eigenvalue -1 . By the definition of observables in section 1.1 the Pauli matrices can be seen as the observables of the spin states along the three axes. The Pauli matrices are also unitary, but because their determinant is -1 , they cannot be seen as rotations in the form of equation (1.7), in fact, treated as evolution operators they represent errors that can occur with a qubit. σ_1 represents a bit flip error:

$$\begin{aligned} \sigma_1 (a|0\rangle + b|1\rangle) &= a \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= a \begin{pmatrix} 0 \\ 1 \end{pmatrix} + b \begin{pmatrix} 1 \\ 0 \end{pmatrix} = b|0\rangle + a|1\rangle, \end{aligned} \quad (1.10)$$

similarly σ_3 represents a phase flip:

$$\begin{aligned}\sigma_3(a|0\rangle + b|1\rangle) &= a \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= a \begin{pmatrix} 1 \\ 0 \end{pmatrix} - b \begin{pmatrix} 0 \\ 1 \end{pmatrix} = a|0\rangle - b|1\rangle,\end{aligned}\quad (1.11)$$

and σ_2 represents a phase and bit flip (ignoring overall phase):

$$\begin{aligned}\sigma_2(a|0\rangle + b|1\rangle) &= a \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= a \begin{pmatrix} 0 \\ i \end{pmatrix} - b \begin{pmatrix} i \\ 0 \end{pmatrix} = b|0\rangle - a|1\rangle.\end{aligned}\quad (1.12)$$

In conclusion to this subsection on the spin- $\frac{1}{2}$ representation, we will look at the consequence of relative phase on physical observation (measurement). The spin up and spin down states along x-axis (see eq. (1.8)) when measured along the z-axis will both yield spin up or spin down with equal probability, consider the superposition of spin up and spin down along x-axis $\frac{1}{\sqrt{2}}(|\uparrow_x\rangle + |\downarrow_x\rangle)$, what will be the result when we measure along z-axis? Classical probability tells us that we get spin up or spin down with equal probability, since spin up or down along x-axis is determined with equal probability. Yet in reality we would always get spin up along z-axis, because the relative phase cancel out in the $|1\rangle$ term when expanded in the basis $\{|0\rangle, |1\rangle\}$. So we see that qubits are very different from classical probabilistic bits in that they have relative phases, which cause **quantum interference**, making probabilities add up in unexpected ways. This property is used extensively in quantum algorithms, to cancel out all incorrect answers (states) and retain the correct ones in the superposition forming the result of a particular quantum computation.

1.3.2 Photon Polarizations

Another two state system that can represent a qubit is photon polarizations. Photons are massless spin-1 particles; they can have two independent polarizations, transverse to the direction of propagation. Under a rotation about the axis of propagation, the two linear polarization states $|x\rangle$ and $|y\rangle$ (representing horizontal and vertical polarizations respectively) transform as

$$\begin{aligned}|x\rangle &\rightarrow \cos\theta|x\rangle + \sin\theta|y\rangle \\ |y\rangle &\rightarrow -\sin\theta|x\rangle + \cos\theta|y\rangle.\end{aligned}\quad (1.13)$$

The 2-d matrix representation of this transform is $\begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$, and has the eigenstates $|R\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$ and $|L\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}$ with eigenvalues $e^{i\theta}$ and $e^{-i\theta}$, the states of right and left circular polarizations.

Suppose we have polarization analyzers that allow only one of the two linear photon polarizations to pass through. Then an x or y polarized photon has 0.5 probability of getting through a 45° rotated analyzer, and vice versa. Quantum interference in photon polarizations occurs when a 45° rotated analyzer is placed between an x and y analyzer. Before the third 45° rotated analyzer is placed, half of the photons that passed through the x or y analyzer is completely blocked at the other one, but after the third analyzer is inserted half of the photons pass through each of the three analyzers, and $\frac{1}{8}$ comes out at the end.

Considering 2-d Hilbert space unitary transformations, the rotation in eq. (1.13) is not the most general possible. But if we have a device for changing the relative phase between the horizontal and vertical polarizations, such as

$$\begin{aligned} |x\rangle &\rightarrow e^{-\frac{i\omega}{2}} |x\rangle \\ |y\rangle &\rightarrow e^{\frac{i\omega}{2}} |y\rangle, \end{aligned} \quad (1.14)$$

then the two transformations (1.13) and (1.14) can be applied together to achieve any unitary transformation with determinant 1 on the photon polarization state in 2-d Hilbert space.

1.4 The Density Operator

The previous formulations about one qubit all assume isolation from the environment, that is, the qubit is the whole system. Yet in practice all observations we make are inevitably limited to a small part of a much larger quantum system. We will consider the simplest example of such a situation: a 2-qubit system in which we only have access to one of the qubits.

Our goal is to characterize the observations that we can make on qubit A when qubit B is unavailable to us. Denote the orthonormal basis in qubit A 's and qubit B 's 2-d Hilbert space by $\{|0\rangle_A, |1\rangle_A\}$ and $\{|0\rangle_B, |1\rangle_B\}$ respectively. Then the state of the two qubits system is a vector in $2 \times 2 = 4$ -d Hilbert space, and can be expressed as the combination of 2-d Hilbert space vectors:

$$(a_A|0\rangle_A + b_A|1\rangle_A) \otimes (a_B|0\rangle_B + b_B|1\rangle_B) \equiv \begin{pmatrix} a_A \\ b_A \end{pmatrix} \otimes \begin{pmatrix} a_B \\ b_B \end{pmatrix} \equiv$$

$$\begin{pmatrix} a_A a_B \\ a_A b_B \\ b_A a_B \\ b_A b_B \end{pmatrix} \equiv a_A a_B |0\rangle_A |0\rangle_B + a_A b_B |0\rangle_A |1\rangle_B \\ + b_A a_B |1\rangle_A |0\rangle_B + b_A b_B |1\rangle_A |1\rangle_B, \quad (1.15)$$

where \otimes denotes the matrix direct product where appropriate, and the matrix direct product is implied in the far right of eq. (1.15).

Consider the state $|\psi\rangle_{AB} = a|0\rangle_A|0\rangle_B + b|1\rangle_A|1\rangle_B$, the qubits A and B are correlated. When we measure qubit A and obtain spin up (down), then qubit B (which we have no access) will also be in the state spin up (down), similarly for measurements on qubit B . Another way to look at it is that in this particular correlation between the two qubits, that is, in this particular state of the whole system, any preparation of only one of the qubits will cause the other qubit to have the same value. In this case the qubits A and B are **entangled**. But first let's characterize the measurements on qubit A more clearly.

An observable acting on A only can be expressed as $\mathbf{M}_A \otimes \mathbf{1}_B$, where \mathbf{M}_A is a self-adjoint operator acting on A , and $\mathbf{1}_B$ is the identity operator on B . Then the expectation value of the observable in $|\psi\rangle_{AB}$ is

$$\begin{aligned} & {}_{AB}\langle\psi|\mathbf{M}_A \otimes \mathbf{1}_B|\psi\rangle_{AB} \\ &= (a^*_A\langle 0|_B\langle 0| + b^*_A\langle 1|_B\langle 1|) (\mathbf{M}_A \otimes \mathbf{1}_B) (a|0\rangle_A|0\rangle_B + b|1\rangle_A|1\rangle_B) \\ &= |a|^2{}_A\langle 0|\mathbf{M}_A|0\rangle_{AB}\langle 0|\mathbf{1}_B|0\rangle_B + a^*b{}_A\langle 0|\mathbf{M}_A|1\rangle_{AB}\langle 0|\mathbf{1}_B|1\rangle_B \\ &\quad + ab^*_A\langle 1|\mathbf{M}_A|0\rangle_{AB}\langle 1|\mathbf{1}_B|0\rangle_B + |b|^2{}_A\langle 1|\mathbf{M}_A|1\rangle_{AB}\langle 1|\mathbf{1}_B|1\rangle_B \\ &= |a|^2{}_A\langle 0|\mathbf{M}_A|0\rangle_A + |b|^2{}_A\langle 1|\mathbf{M}_A|1\rangle_A \\ &= \text{tr}(\mathbf{M}_A\rho_A), \end{aligned} \quad (1.16)$$

where $\rho_A = |a|^2|0\rangle_{AA}\langle 0| + |b|^2|1\rangle_{AA}\langle 1|$. The operator ρ_A is called the **density operator (matrix)** for qubit A . It is self-adjoint, positive (its eigenvalues are nonnegative) and has unit trace. From the form of ρ_A and the fact that eq. (1.16) holds for any observable \mathbf{M}_A , the density operator can be interpreted as an ensemble of possible quantum states, each component in ρ_A is a projection onto a particular quantum state with the probability of that state as coefficient. That is, if we assume that qubit A is in the state $|0\rangle_A$ with probability $|a|^2$, and $|1\rangle_A$ with probability $|b|^2$, then the expected value of measurement \mathbf{M}_A would be

$$\langle\mathbf{M}_A\rangle = |a|^2{}_A\langle 0|\mathbf{M}_A|0\rangle_A + |b|^2{}_A\langle 1|\mathbf{M}_A|1\rangle_A, \quad (1.17)$$

exactly the same result as (1.16).

In the density operator representation of a qubit state, there is no information on relative phases, only the individual probabilities of the quantum states in the ensemble are known. This is a consequence of observing only part of a quantum system (local observations). When the relative phases of the state of a quantum system are known, the system can be represented as a **coherent superposition** (as in (1.2)), called a **pure state**; when the system establishes some correlation with another system (or the environment) in which we do not have access, then we are forced to make local observations, and the resulting state becomes (from our point of view) an ensemble of quantum states, the relative phase information is lost, and the system (from our perspective) has undergone **decoherence** (becoming a **mixed state**). Note that the discussions in sections 1.2 and 1.3 only involve the pure state of a qubit, mixed state only occurs when other unobserved systems are involved.

Consider an incoherent ensemble of spin up and spin down states along z-axis with equal probability $\rho_A = \frac{1}{2}(|0\rangle_{AA}\langle 0| + |1\rangle_{AA}\langle 1|) = \frac{1}{2}\mathbf{1}_A$ (the last equality due to the fact that the projections form an orthonormal basis), if we were to measure the spin along axis \hat{n} , then $\langle |\uparrow_{\hat{n}}\rangle_{AA}\langle \uparrow_{\hat{n}}| \rangle = \text{tr}(|\uparrow_{\hat{n}}\rangle_{AA}\langle \uparrow_{\hat{n}}| \rho_A) = \frac{1}{2}\text{tr}(|\uparrow_{\hat{n}}\rangle_{AA}\langle \uparrow_{\hat{n}}|) = \frac{1}{2}$, since this holds for any axis, the result of any measurement of the qubit is completely random, such a phenomenon would be impossible if qubit A is coherent (can be represented in the form of (1.2)).

Now let's generalize the discussion to an arbitrary state of any bipartite quantum system, where system A resides in Hilbert space \mathcal{H}_A , and system B in \mathcal{H}_B . The Hilbert space of the whole system is $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, the vector space tensor product of the constituting Hilbert spaces. Denote $\{|i\rangle_A\}$ and $\{|\mu\rangle_B\}$ as orthonormal basis for \mathcal{H}_A and \mathcal{H}_B respectively, then $\{|i\rangle_A|\mu\rangle_B\}$ is an orthonormal basis for \mathcal{H}_{AB} (matrix direct product implied). Thus an arbitrary pure state of the bipartite system can be expressed as

$$|\psi\rangle_{AB} = \sum_{i,\mu} a_{i\mu} |i\rangle_A |\mu\rangle_B, \quad (1.18)$$

where $\sum_{i,\mu} |a_{i\mu}|^2 = 1$. The expectation value of \mathbf{M}_A is

$$\begin{aligned} \langle \mathbf{M}_A \rangle &= {}_{AB} \langle \psi | \mathbf{M}_A \otimes \mathbf{1}_B | \psi \rangle_{AB} \\ &= \left(\sum_{j,\nu} a_{j\nu}^* \langle j | \nu \rangle \right) (\mathbf{M}_A \otimes \mathbf{1}_B) \left(\sum_{i,\mu} a_{i\mu} |i\rangle_A |\mu\rangle_B \right) \\ &= \sum_{i,j,\mu} a_{j\mu}^* a_{i\mu} \langle j | \mathbf{M}_A | i \rangle_A \\ &= \text{tr}(\mathbf{M}_A \rho_A), \end{aligned} \quad (1.19)$$

where the density operator for system A is

$$\rho_A = \text{tr}_B (|\psi\rangle_{ABAB}\langle\psi|) = \sum_{i,j,\mu} a_{j\mu}^* a_{i\mu} |i\rangle_{AA}\langle j|. \quad (1.20)$$

The density operator for subsystem A is obtained by performing a **partial trace** over subsystem B of the density matrix of the combined system (note though that the whole system is in a pure state). And from eq. (1.20) we can infer that ρ_A has the following properties:

1. Self-adjoint.
2. Positive semidefinite. For any $|\psi\rangle_A$, ${}_A\langle\psi|\rho_A|\psi\rangle_A = \sum_{\mu} |\sum_i a_{i\mu A} \langle\psi|i\rangle_A|^2 \geq 0$.
3. $\text{tr}(\rho_A) = 1$. Because $\sum_{i,\mu} |a_{i\mu}|^2 = 1$.

So ρ_A can be diagonalized, and all its eigenvalues are real and nonnegative, and they sum to one.

We see that the state of a subsystem of a larger quantum system may not be a ray in Hilbert space, and is generally represented as a density operator. So a quantum system with state as a ray is in a pure state, otherwise the state is mixed. In the case of a pure state, the density operator would be the projection onto the 1-d space spanned by the ray. So for a pure density operator $\rho^2 = \rho$.

A general density operator can be diagonalized by a suitable orthonormal basis, and results in the form

$$\rho = \sum_a p_a |\psi_a\rangle\langle\psi_a| \quad (1.21)$$

where $0 < p_a \leq 1$ and $\sum_a p_a = 1$. The state is pure if and only if there is only one term in the sum. The state is an incoherent superposition of the states $|\psi_a\rangle$, in that the relative phases are inaccessible.

For an observable \mathbf{M} , $\langle\mathbf{M}\rangle = \text{tr}(\mathbf{M}\rho) = \sum_a p_a \langle\psi_a|\mathbf{M}|\psi_a\rangle$. So the classical probabilistic interpretation mentioned earlier holds for general density operators and measurements. When system A becomes mixed due to interaction with system B , we say that the two systems are **entangled**, the entanglement destroys the coherence of system A , it is as if system A collapses from the initial superposition of states to one of the states each with a certain probability (like classical probabilistic bits).

Finally let's look at the independent evolution of density operators. For a quantum system consisting of subsystems A and B , without loss of generality,

we can assume the state of the whole system is pure; since we will only be interested in the “partial” evolution of A , all other quantum systems in contact with the whole system can be unioned with B . So the evolution of the whole system can be described by a Hamiltonian on $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Assume that its form is

$$\mathbf{H}_{AB} = \mathbf{H}_A \otimes \mathbf{1}_B + \mathbf{1}_A \otimes \mathbf{H}_B, \quad (1.22)$$

that is, assume the Hamiltonians of A and B are uncoupled, meaning A and B would evolve independently, then the evolution operator for the whole system can be separated into unitary evolution operators for each system

$$\mathbf{U}_{AB}(t) = \mathbf{U}_A(t) \otimes \mathbf{U}_B(t). \quad (1.23)$$

And the general bipartite state (1.18) evolves to

$$|\psi(t)\rangle_{AB} = \sum_{i,\mu} a_{i\mu} \mathbf{U}(t)_A |i(0)\rangle_A \mathbf{U}(t)_B |\mu(0)\rangle_B = \sum_{i,\mu} a_{i\mu} |i(t)\rangle_A |\mu(t)\rangle_B, \quad (1.24)$$

where $\{|i(t)\rangle_A\}$ and $\{|\mu(t)\rangle_B\}$ define new orthonormal basis for \mathcal{H}_A and \mathcal{H}_B respectively. So the density operator of subsystem A (found by taking the partial trace) is

$$\rho_A(t) = \sum_{i,j,\mu} a_{i\mu} a_{j\mu}^* |i(t)\rangle_{AA} \langle j(t)| = \mathbf{U}_A(t) \rho_A(0) \mathbf{U}_A(t)^\dagger. \quad (1.25)$$

In the basis in which $\rho_A(0)$ is diagonal, the last equation becomes

$$\rho_A(t) = \sum_a p_a \mathbf{U}_A(t) |\psi_a(0)\rangle_{AA} \langle \psi_a(0)| \mathbf{U}_A(t)^\dagger. \quad (1.26)$$

This can again be interpreted according to classical probabilities: since each of the $|\psi_a(0)\rangle_A$'s occurs with probability p_a at time 0, then at time t each $|\psi_a(t)\rangle_A$'s occurs with probability p_a . This again reflects the fact that in the density operator representation there is no information on relative phase (or that relative phase is unobservable), so no quantum interference can occur, leading to results that look like classical probabilities.

1.5 Bloch Sphere

From the discussions on density operators in section 1.4, it appears that the density operator representation of the state of a quantum system is more general than the ray in a Hilbert space representation in section 1.1. That is, whereas rays such as $|\psi\rangle = \sum_i a_i |i\rangle$ can only represent pure states, density

operators can represent both pure states *and* mixed states. In this section we will elaborate on these concepts in the case of a single qubit (entangled or isolated).

Recall that the most general density matrix is self-adjoint, positive semidefinite, and has unit trace. The most general form of a 2×2 self-adjoint matrix with unit trace is

$$\begin{pmatrix} \frac{1}{2} + a & b - ic \\ b + ic & \frac{1}{2} - a \end{pmatrix} = \frac{1}{2} \mathbf{1} + b\sigma_1 + c\sigma_2 + a\sigma_3, \quad (1.27)$$

where $a, b, c \in \mathfrak{R}$. A necessary and sufficient condition for a matrix to be positive semidefinite is that all its eigenvalues be nonnegative. That means the determinant of the matrix (equal to the product of its eigenvalues) must be nonnegative, and given a nonnegative trace (which is the sum of its eigenvalues), this condition is also sufficient. So we have

$$\det(\rho) = \frac{1}{4} - a^2 - b^2 - c^2 = \frac{1}{4} (1 - \vec{P}^2) \geq 0, \quad (1.28)$$

where $\vec{P} = (P_1, P_2, P_3) = (2b, 2c, 2a)$, and $\vec{P}^2 \leq 1$. So the most general density matrix of a qubit can be expressed as

$$\rho(\vec{P}) = \frac{1}{2} (\mathbf{1} + P_1\sigma_1 + P_2\sigma_2 + P_3\sigma_3) = \frac{1}{2} (\mathbf{1} + \vec{P} \cdot \vec{\sigma}), \quad (1.29)$$

where $0 \leq |\vec{P}| \leq 1$. Thus there is an one to one correspondence between the points within a unit ball in 3-d real vector space and the density matrix of a qubit. And we call this ball the **Bloch sphere** (though it's really a ball).

At the boundary of the Bloch sphere (where $|\vec{P}| = 1$), the determinant of the density matrix is zero, and since it has unit trace its two eigenvalues are 0 and 1. This means that the density matrices corresponding to points at the surface of the unit ball are one-dimensional projectors, and hence pure states. So the unit ball (Bloch sphere) representation is a natural extension of the 3-d unit vector representation discussed in section 1.3.1; pure states correspond to unit length vectors, and mixed states correspond to vectors with length less than 1. In fact, we will show that a pure state $|\psi(\theta, \varphi)\rangle$ pointing in the (θ, φ) direction has pure state density matrix $\rho(\hat{n}) = \frac{1}{2} (\mathbf{1} + \hat{n} \cdot \vec{\sigma})$, where $\hat{n} = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$.

The density matrix satisfies the property

$$\begin{aligned} (\hat{n} \cdot \vec{\sigma}) \rho(\hat{n}) &= \rho(\hat{n}) (\hat{n} \cdot \vec{\sigma}) = \frac{1}{2} ((\hat{n} \cdot \vec{\sigma}) + (\hat{n} \cdot \vec{\sigma})^2) \\ &= \frac{1}{2} ((\hat{n} \cdot \vec{\sigma}) + \mathbf{1}) = \rho(\hat{n}). \end{aligned} \quad (1.30)$$

Since the state $|\psi(\theta, \varphi)\rangle = |\psi(\hat{n})\rangle = |\uparrow_{\hat{n}}\rangle$ is an eigenstate of $\hat{n} \cdot \vec{\sigma}$ with eigenvalue 1 (see section 1.3.1), the density matrix is the projector

$$\rho(\hat{n}) = |\uparrow_{\hat{n}}\rangle\langle\uparrow_{\hat{n}}| = |\psi(\theta, \varphi)\rangle\langle\psi(\theta, \varphi)|, \quad (1.31)$$

hence the density matrix of state $|\psi(\theta, \varphi)\rangle$. Alternatively we can also acquire the same result using direct matrix calculation, the state is from (1.3)

$$|\psi(\theta, \varphi)\rangle = \begin{pmatrix} e^{-\frac{i\varphi}{2}} \cos \frac{\theta}{2} \\ e^{\frac{i\varphi}{2}} \sin \frac{\theta}{2} \end{pmatrix} \quad (1.32)$$

so the density matrix is

$$\begin{aligned} \rho(\theta, \varphi) &= |\psi(\theta, \varphi)\rangle\langle\psi(\theta, \varphi)| \\ &= \begin{pmatrix} \cos^2 \frac{\theta}{2} & \cos \frac{\theta}{2} \sin \frac{\theta}{2} e^{-i\varphi} \\ \cos \frac{\theta}{2} \sin \frac{\theta}{2} e^{i\varphi} & \sin^2 \frac{\theta}{2} \end{pmatrix} = \frac{1}{2} \mathbf{1} + \frac{1}{2} \begin{pmatrix} \cos \theta & \sin \theta e^{-i\varphi} \\ \sin \theta e^{i\varphi} & -\cos \theta \end{pmatrix} \\ &= \frac{1}{2} (\mathbf{1} + \hat{n} \cdot \vec{\sigma}) = \rho(\hat{n}). \end{aligned} \quad (1.33)$$

Also note that in the density matrix/Bloch sphere representation the overall phase of a pure state would be cancelled out, for $|\psi\rangle = e^{i\theta}|\varphi\rangle$, $\rho = |\psi\rangle\langle\psi| = e^{-i\theta}|\varphi\rangle\langle\varphi|e^{i\theta} = |\varphi\rangle\langle\varphi|$; thus all parameters in the density matrix have physical meaning.

For a qubit in the general Bloch state $\rho(\vec{P})$ defined in eq. (1.29), the expected value of a measurement of the observable $\hat{n} \cdot \vec{\sigma}$ is

$$\langle\hat{n} \cdot \vec{\sigma}\rangle_{\vec{P}} = \text{tr}(\hat{n} \cdot \vec{\sigma} \rho(\vec{P})) = \hat{n} \cdot \vec{P}; \quad (1.34)$$

this can be seen as the “amount” of spin component the state has on the axis \hat{n} , or the amount of **polarization** of the spin in that direction. With many identical preparations of $\rho(\vec{P})$ we can determine \vec{P} by measuring $\hat{n} \cdot \vec{\sigma}$ along each of three linearly independent axes. In the special case when \vec{P} has unit length the polarization of the spin in $\hat{n} = \vec{P}$ is maximum, and the state is just the pure state $|\psi(\vec{P})\rangle = |\uparrow_{\vec{P}}\rangle$. This implies that entanglement with other quantum systems decreases polarization, we will look at some explicit examples of such a phenomenon when we discuss general evolution in detail.

1.6 Schmidt Decomposition

A bipartite pure state can be expressed in a standard form called the **Schmidt decomposition** that is very useful.

An arbitrary vector in $\mathcal{H}_A \otimes \mathcal{H}_B$ (eq. (1.18)) can be expanded as

$$|\psi\rangle_{AB} = \sum_{i,\mu} a_{i\mu} |i\rangle_A |\mu\rangle_B \equiv \sum_i |i\rangle_A |\tilde{i}\rangle_B, \quad (1.35)$$

where we have defined

$$|\tilde{i}\rangle_B \equiv \sum_{\mu} a_{i\mu} |\mu\rangle_B. \quad (1.36)$$

Note that the $|\tilde{i}\rangle_B$'s need not be orthogonal or normalized.

Now suppose the $\{|i\rangle_A\}$ basis is chosen to be the basis in which ρ_A is diagonal, then

$$\rho_A = \sum_i p_i |i\rangle_{AA} \langle i|. \quad (1.37)$$

We can also compute ρ_A by performing a partial trace,

$$\begin{aligned} \rho_A &= \text{tr}_B (|\psi\rangle_{ABAB} \langle \psi|) \\ &= \text{tr}_B \left(\sum_{ij} |i\rangle_{AA} \langle j| \otimes |\tilde{i}\rangle_{BB} \langle \tilde{j}| \right) = \sum_{ij} {}_B \langle \tilde{j} | \tilde{i} \rangle_B (|i\rangle_{AA} \langle j|). \end{aligned} \quad (1.38)$$

The last equality is because

$$\text{tr}_B (|\tilde{i}\rangle_{BB} \langle \tilde{j}|) = \sum_k {}_B \langle k | \tilde{i} \rangle_{BB} \langle \tilde{j} | k \rangle_B = \sum_k {}_B \langle \tilde{j} | k \rangle_{BB} \langle k | \tilde{i} \rangle_B = {}_B \langle \tilde{j} | \tilde{i} \rangle_B, \quad (1.39)$$

where $\{|k\rangle_B\}$ is an orthonormal basis for \mathcal{H}_B . By comparing eq. (1.37) and eq. (1.38), we see that ${}_B \langle \tilde{j} | \tilde{i} \rangle_B = p_i \delta_{ij}$. So in (1.35) the $\{|\tilde{i}\rangle_B\}$ are orthogonal when the $\{|i\rangle_A\}$ basis diagonalizes ρ_A . Orthonormal vectors are obtained by rescaling, $|i'\rangle_B = p_i^{-\frac{1}{2}} |\tilde{i}\rangle_B$, so eq. (1.35) becomes

$$|\psi\rangle_{AB} = \sum_i \sqrt{p_i} |i\rangle_A |i'\rangle_B, \quad (1.40)$$

in terms of a *particular* orthonormal basis of \mathcal{H}_{AB} .

Eq. (1.40) is the Schmidt decomposition of the bipartite pure state $|\psi\rangle_{AB}$. Any bipartite pure state can be expressed in this form (since any density matrix $\rho_A = \text{tr}_B (|\psi\rangle_{ABAB} \langle \psi|)$ can be diagonalized), but in general we cannot simultaneously expand *both* $|\psi\rangle_{AB}$ and $|\varphi\rangle_{AB}$ in the Schmidt decomposition form with the same orthonormal bases for \mathcal{H}_A and \mathcal{H}_B (since in general ρ_A and $\rho'_A = \text{tr}_B (|\varphi\rangle_{ABAB} \langle \varphi|)$ cannot be simultaneously diagonalized with the same basis).

We can also evaluate the partial trace over \mathcal{H}_A (subsystem A) of eq. (1.40) to obtain

$$\rho_B = \text{tr}_A (|\psi\rangle_{ABAB} \langle \psi|) = \sum_i p_i |i'\rangle_{BB} \langle i'|. \quad (1.41)$$

So ρ_A and ρ_B have the same *nonzero* eigenvalues, but since the dimensions of \mathcal{H}_A and \mathcal{H}_B need not be equal, they generally do not have the same number of zero eigenvalues.

Now let's look at the uniqueness of the Schmidt decomposition. If ρ_A (and hence ρ_B) have no degenerate eigenvalues other than zero, then the Schmidt decomposition is unique. By pairing the eigenstates of ρ_A and ρ_B with the same eigenvalue, we can obtain the unique Schmidt decomposition. The basis states have been chosen to have zero phase (that is, with real coefficients); we can of course flip the phase of $|i\rangle_A$, but then for the eigenvalues to stay the same, the phase of $|i'\rangle_B$ also has to be flipped, so we get $e^{i\pi}|i\rangle_A e^{i\pi}|i'\rangle_B = |i\rangle_A |i'\rangle_B$, the Schmidt decomposition remains unchanged. If ρ_A has degenerate nonzero eigenvalues, then more information than that provided by the density matrices is needed to determine the Schmidt decomposition; that is, we need to know which degenerate eigenstates get paired together. The interested reader can explore the case when \mathcal{H}_A and \mathcal{H}_B have the same dimension and all eigenvalues are equal.

For any bipartite pure state its **Schmidt number** is the number of nonzero eigenvalues in ρ_A (or ρ_B) and hence the number of terms in the Schmidt decomposition of the pure state. Two systems forming a bipartite pure state are entangled when the Schmidt number is greater than one; otherwise they are separable (or unentangled). When systems A and B are entangled, using (1.40) we see that if we measure system A (local measurement) in the basis $\{|i\rangle_A\}$ and get $|j\rangle_A$, then measurement on system B will yield $|j'\rangle_B$ with probability one, we would get the same result if we measured system B first and got $|j'\rangle_B$. When systems A and B are separable, the Schmidt number is one, and the Schmidt decomposition is a direct product of pure states in A and B . Local manipulation of A or B cannot increase the Schmidt number of the bipartite system. That is, in order for two systems to be entangled, they must be brought together for direct interaction. Whereas unentangled states can be prepared by preparing pure states for A and B separately.

1.7 Ambiguity of the Ensemble Interpretation

1.7.1 Convexity of Density Operators

Recall that the most general density operator ρ acting on a Hilbert space \mathcal{H} satisfies

1. ρ is self-adjoint.
2. ρ is nonnegative (positive semidefinite).
3. $\text{tr}(\rho) = 1$.

So for two density operators ρ_1 and ρ_2 , we can construct another density operator as the convex linear combination of the two:

$$\rho(\lambda) = \lambda\rho_1 + (1 - \lambda)\rho_2, \quad (1.42)$$

where $0 \leq \lambda \leq 1$. The three requirements for density operators are easily proved. So in a Hilbert space \mathcal{H} of dimension N , the density matrices form a **convex subset** in the real vector space of $N \times N$ self-adjoint matrices. (A subset of a vector space is convex if linear combinations of members of the subset is in the subset.)

Most density operators can be expressed as a sum of other density operators in many different ways, but pure density operators cannot be expressed as a convex sum of two other density operators. Consider a pure state $\rho = |\psi\rangle\langle\psi|$, and let $|\psi_\perp\rangle$ denote a vector orthogonal to $|\psi\rangle$. Suppose the state can be expanded as in eq. (1.42), then

$$\langle\psi_\perp|\rho|\psi_\perp\rangle = 0 = \lambda\langle\psi_\perp|\rho_1|\psi_\perp\rangle + (1 - \lambda)\langle\psi_\perp|\rho_2|\psi_\perp\rangle. \quad (1.43)$$

Since both terms in the right hand side is nonnegative, they are both zero. When λ is not 0 or 1, we conclude that ρ_1 and ρ_2 are orthogonal to $|\psi_\perp\rangle$. And since $|\psi_\perp\rangle$ can be any vector orthogonal to $|\psi\rangle$, we conclude that $\rho_1 = \rho_2 = \rho$. So the pure state density matrices are **extremal points** of the convex subset in that they cannot be expressed as a linear combination of other matrices. Furthermore, *only* the pure states are extremal, since any mixed state is a convex sum of pure states.

In the Bloch sphere representation (the special case of 2×2 density matrices), the convexity property is easily seen, since any point in the unit ball (mixed states) can be expressed as the linear combination of two points on the surface of the ball (pure states), and any point on the surface cannot be so expressed. But the 2×2 case (qubit) is atypical in that if we extend the Bloch sphere representation to $N > 2$, then states at the boundary of the sphere ($\det \rho = \frac{1}{N^2} (1 - \vec{P}^2) = 0$ where \vec{P} is a N vector) are not necessarily pure, and hence not extremal. That is because $\det \rho = 0$ requires at least one zero eigenvalue, and when $N > 2$ there can be more than one nonzero eigenvalues, resulting in a mixed state; but when $N = 2$ there can only be one nonzero eigenvalue, since there's just two eigenvalues in total.

In conclusion, $N \times N$ density operators form a convex subset of the real vector space of $N \times N$ self-adjoint matrices, and a state is pure if and only if its density matrix is an extremal point of the subset. And when $N = 2$, a state is pure if and only if the determinant of its density matrix vanishes.

1.7.2 Ensemble Preparations

The convexity of density matrices has a simple and enlightening physical interpretation. Suppose a preparer agrees to prepare one of two possible states ρ_1 and ρ_2 with probabilities λ and $1 - \lambda$ respectively, then the expectation value of measurement \mathbf{M} performed on the state (by the receiver of the state) is

$$\begin{aligned} \langle M \rangle &= \lambda \langle M \rangle_1 + (1 - \lambda) \langle M \rangle_2 \\ &= \lambda \text{tr}(M\rho_1) + (1 - \lambda) \text{tr}(M\rho_2) \\ &= \text{tr}(M\rho(\lambda)), \end{aligned} \tag{1.44}$$

averaging over both possible preparations *and* possible measurement outcomes. So there is no observable difference if the preparer had prepared the state $\rho(\lambda)$.

In fact, for any mixed state ρ , there are an infinite variety of ways to express ρ as a linear combination of other density matrices, and hence an infinite variety of ways to prepare the state. Thus the preparation of a mixed state is always ambiguous, whereas the preparation of a pure state is unambiguous, since it can be determined uniquely if we are given enough copies to experiment with.

How ambiguous is the preparation of a mixed state? Since any density matrix can be expressed as a sum of pure states, we can instead ask how many combinations of pure states are possible for a given mixed state? Let's first consider the "maximally mixed" qubit $\rho = \frac{1}{2}\mathbf{1}$ (in that $|\vec{P}| = 0$), we can see that preparing spin up and spin down states along any axis \hat{n} with equal probabilities will yield the density matrix:

$$\rho = \frac{1}{2} |\uparrow_{\hat{n}}\rangle\langle\uparrow_{\hat{n}}| + \frac{1}{2} |\downarrow_{\hat{n}}\rangle\langle\downarrow_{\hat{n}}| = \frac{1}{2}\mathbf{1}, \tag{1.45}$$

since any such pair forms an orthonormal basis. It is clear that only when the density matrix has degenerate nonzero eigenvalues can there be distinct (in fact, infinite) ways of preparing the matrix from an ensemble of *mutually orthogonal* pure states. But orthogonality is not required, consider a point in the interior of the Bloch sphere

$$\rho(\vec{P}) = \frac{1}{2}(\mathbf{1} + \vec{P} \cdot \vec{\sigma}), \tag{1.46}$$

where $0 < |\vec{P}| < 1$. If $\vec{P} = \lambda\hat{n}_1 + (1 - \lambda)\hat{n}_2$, it can be expressed as

$$\rho(\vec{P}) = \lambda\rho(\hat{n}_1) + (1 - \lambda)\rho(\hat{n}_2), \quad (1.47)$$

where \hat{n}_1 and \hat{n}_2 are unit vectors. This illustrates that the preparation of a mixed state is indeed *very* ambiguous.

1.7.3 The GHJW Theorem and Quantum Erasure

Any density matrix can be realized as an ensemble of pure states, for a density matrix ρ_A , consider one such realization:

$$\rho_A = \sum_i p_i |\varphi_i\rangle_{AA} \langle \varphi_i|, \quad (1.48)$$

where $\sum_i p_i = 1$. The states $\{|\varphi_i\rangle_A\}$ are all normalized, but *not necessarily* orthogonal. We can construct a “purification” of ρ_A , that is a bipartite state $|\Phi\rangle_{AB}$ that yields ρ_A when we perform a partial trace over \mathcal{H}_B . Assume it’s of the form

$$|\Phi\rangle_{AB} = \sum_i \sqrt{p_i} |\varphi_i\rangle_A |\alpha_i\rangle_B, \quad (1.49)$$

where $\{|\alpha_i\rangle_B\}$ is an orthonormal basis of \mathcal{H}_B , clearly then

$$\text{tr}_B (|\Phi\rangle_{AB} \langle \Phi|) = \rho_A. \quad (1.50)$$

So given the purification we can realize the $\{|\varphi_i\rangle_A\}$ ensemble interpretation of ρ_A by a measurement in system B that projects onto the basis $\{|\alpha_i\rangle_B\}$. If the measurement result is $|\alpha_j\rangle_B$, then we know that system A is now in the state $|\varphi_j\rangle_{AA} \langle \varphi_j|$.

Now consider another ensemble interpretation of the same ρ_A

$$\rho_A = \sum_\mu q_\mu |\psi_\mu\rangle_{AA} \langle \psi_\mu|, \quad (1.51)$$

there is a corresponding purification

$$|\Psi\rangle_{AB} = \sum_\mu \sqrt{q_\mu} |\psi_\mu\rangle_A |\beta_\mu\rangle_B, \quad (1.52)$$

where again $\{|\beta_\mu\rangle_B\}$ is an orthonormal basis for \mathcal{H}_B . So as before the $\{|\psi_\mu\rangle_A\}$ ensemble interpretation can be realized by a measurement in \mathcal{H}_B that projects onto the $\{|\beta_\mu\rangle_B\}$ basis.

How are $|\Phi\rangle_{AB}$ and $|\Psi\rangle_{AB}$ related? They both yield ρ_A when we perform partial trace over \mathcal{H}_B , so their Schmidt decompositions (see 1.6) are

$$\begin{aligned} |\Phi\rangle_{AB} &= \sum_k \sqrt{\lambda_k} |k\rangle_A |k'_1\rangle_B, \quad \text{and} \\ |\Psi\rangle_{AB} &= \sum_k \sqrt{\lambda_k} |k\rangle_A |k'_2\rangle_B, \end{aligned} \quad (1.53)$$

where the λ_k 's are the eigenvalues of ρ_A and the $|k\rangle_A$'s are the corresponding eigenvectors. Since $\{|k'_1\rangle_B\}$ and $\{|k'_2\rangle_B\}$ are both orthonormal bases for \mathcal{H}_B , there is a unitary transformation \mathbf{U}_B such that $|k'_1\rangle_B = \mathbf{U}_B|k'_2\rangle_B$, and so

$$|\Phi\rangle_{AB} = (\mathbf{1}_A \otimes \mathbf{U}_B)|\Psi\rangle_{AB}. \quad (1.54)$$

Alternatively,

$$|\Phi\rangle_{AB} = \sum_i \sqrt{p_i} |\varphi_i\rangle_A |\alpha_i\rangle_B = \sum_\mu \sqrt{q_\mu} |\psi_\mu\rangle_A |\gamma_\mu\rangle_B \quad (1.55)$$

where $|\gamma_\mu\rangle_B = \mathbf{U}_B|\beta_\mu\rangle_B$ is yet another basis for \mathcal{H}_B . So there is a single purification of ρ_A such that both the $\{|\varphi_i\rangle_A\}$ ensemble and the $\{|\psi_\mu\rangle_A\}$ ensemble can be realized by choosing the appropriate measurement to perform in \mathcal{H}_B .

If there are many ensembles that realize ρ_A , where the maximum number of pure states in any ensemble is n , then we can choose a Hilbert space \mathcal{H}_B of dimension n , and construct a purification $|\Psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ such that any one of the ensembles can be realized by measuring a suitable observable of B . This is the **GHJW theorem**. It leads to a phenomenon called **quantum erasure**, which is illustrated below.

We will first consider the nature of coherence. The density matrix $\rho = \frac{1}{2}\mathbf{1}$ describes an incoherent superposition of pure states $|\uparrow_z\rangle$ and $|\downarrow_z\rangle$, whereas states such as

$$|\uparrow_x, \downarrow_x\rangle = \frac{1}{2}(|\uparrow_z\rangle \pm |\downarrow_z\rangle) \quad (1.56)$$

describes coherent superpositions. The difference between these two lies in whether or not relative phases are observable. Relative phases can be “observed” by quantum interference, that is, if we can detect quantum interference, then relative phase is observable; conversely, if the relative phase is unobservable, there can be no quantum interference. This is not much of an explanation, but we can look at coherence in this way: if the state is a coherent superposition of $|\uparrow_z\rangle$ and $|\downarrow_z\rangle$, we can not obtain the state with any conceivable measurement, that is, no measurement can tell us how the two states are superposed; but when the state is an incoherent superposition, we are still not sure of its state, yet we *know* that it is either $|\uparrow_z\rangle$ or $|\downarrow_z\rangle$, each with a given prior probability, so that it is essentially a probabilistic classical bit, and cannot possibly create interference. Yet another way to look at it is that if there is in principle no way to know what the state is, then it is coherent, and interferes; but if we can in principle find out what state it is in, then it cannot be coherent. For example, consider entangling spin A with spin B , spin A will decohere because we can, in principle, measure spin B on the z -axis to find out whether the state of A is $|\uparrow_z\rangle$ or $|\downarrow_z\rangle$,

as illustrated more generally at the beginning of this section, so the state of A is an incoherent superposition of $|\uparrow_z\rangle$ and $|\downarrow_z\rangle$ after entanglement with spin B .

Yet decoherence is not irreversible, that is, we can always create a coherent superposition of *some* ensemble of states by projecting the state to any pure state. For example, measuring the spin on x-axis yields

$$|\uparrow_x, \downarrow_x\rangle = \frac{1}{2}(|\uparrow_z\rangle \pm |\downarrow_z\rangle), \quad (1.57)$$

which are coherent superpositions of spin up and spin down along z-axis. That is, we have erased the information about whether the state is spin up or spin down along z-axis (or “erased” the possibility of knowing it) and thus created a coherent superposition of these two states. This is the quantum erasure phenomenon. We can generalize using the situation described at the beginning, if we realized the ensemble $\{|\varphi_i\rangle_A\}$ in ρ_A by a suitable measurement in \mathcal{H}_B , then the state of A is an incoherent superposition of the $|\varphi_i\rangle_A$ ’s, since we know that the state is $|\varphi_j\rangle_A$ for some unknown j . Yet if we now perform a measurement on \mathcal{H}_B that projects onto the $\{|\gamma_\mu\rangle_B\}$ basis, we would obtain state $|\psi_\nu\rangle_A$ for some ν , which is a coherent superposition of the ensemble $\{|\varphi_i\rangle_A\}$. We have restored coherence by quantum erasure. The information erased is called the “welcher weg” information, which is whether the state is $|\varphi_i\rangle_A$ or $|\varphi_j\rangle_A$.

1.8 Summary

Axioms. Quantum mechanics operate in a Hilbert space \mathcal{H} , the fundamental assumptions are:

1. A **state** is a **ray** in \mathcal{H} .
2. An **observable** is a **self-adjoint operator** on \mathcal{H} .
3. A **measurement** is an orthogonal **projection**.
4. **Evolution** in time is **unitary**.

Density operator. If we observe only part of a larger quantum system, the axioms need not hold. A quantum state is described by a **density operator** ρ , which is a nonnegative (positive semidefinite) operator with unit trace. It is **pure** (the state can be described by a ray) if and only if $\rho^2 = \rho$; otherwise the state is **mixed**. The expectation value of an observable \mathbf{M} is $\text{tr}(\mathbf{M}\rho)$.

Qubit. A quantum system with 2-d Hilbert space is called a **qubit**. The general density matrix of a qubit is

$$\rho(\vec{P}) = \frac{1}{2}(\mathbf{1} + \vec{P} \cdot \vec{\sigma}),$$

where $|\vec{P}| \leq 1$. Pure states have $|\vec{P}| = 1$, and the state $\vec{P} = 0$ is considered **maximally entangled**.

Schmidt decomposition. The Hilbert space of a system divided in two parts (**bipartite** systems) \mathcal{H}_A and \mathcal{H}_B is their vector space tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$. For any pure state $|\psi\rangle_{AB}$ of a bipartite system, there are orthonormal bases $\{|i\rangle_A\}$ and $\{|i'\rangle_B\}$ such that

$$|\psi\rangle_{AB} = \sum_i \sqrt{p_i} |i\rangle_A |i'\rangle_B,$$

which is the **Schmidt decomposition** of $|\psi\rangle_{AB}$. The number of non-vanishing eigenvalues of ρ_A (which is equal to ρ_B 's) is the **Schmidt number**. The bipartite state is **entangled** if the Schmidt number is greater than one.

Ensembles. A mixed state of a system A can be prepared as an **ensemble** of pure states in infinite ways, all of which are indistinguishable experimentally. Given a mixed state of system A , there is a pure state of the bipartite system composed of system A and system B (the system with which A is entangled) such that any preparation as an ensemble of pure states can be realized by a suitable measurement in system B (the **GHJW theorem**).