Zhian Jia (贾治安)

Quantum information theory in a nutshell

– From qubit to the quantum universe –

Contents

Part I
Quantum states and quantum operations

Chapter 1
Quantum states as density operators

We must be clear that when it comes to atoms, language can be used only as in poetry.

By Niels Bohr
In his first meeting with Werner Heisenberg in early summer 1920, quoted in "Defense Implications of International Indeterminacy" (1972) by Robert J. Pranger

For realistic application of quantum technologies, a quantum system is hardly isolated from its environments perfectly, the system exchange energy, particle and information all the time with the environments. This motivates us to study the *open quantum system*. An open quantum system is defined as a system which can exchange energy, particle and information with its environments and we are not able to observe the environments.

A quantum process includes state preparation, state transformation (or, if the transmission is in time rather than space, this is time evolution) and measurement. We will try to figure out what is the mathematical model to describe an open quantum system, how to describe the quantum states, evolution and measurement. The approach we take to study the open quantum system is to regard the system $\mathcal{H}_S$ combined with its environment $\mathcal{H}_E$ as a closed system $\mathcal{H}_{SE}$, then we will discuss how the states, evolution and measurement of the closed system $\mathcal{H}_{SE}$ (whose description we already know) behave if we only have access to system $\mathcal{H}_S$.

In this chapter, we will focus on the description of quantum states of the open system $\mathcal{H}_S$, present basic concepts and fix notations. In the next chapter the evolution and measurement will be discussed.

§ 1.1 Qubit and Bloch sphere

In classical information theory, *bit* is used as the unit of classical information, it is realized by a two-classical-state (0 and 1 states) device physically. Correspondingly, *qubit* (abbreviation of *quantum bit*) is the unit of quantum information, it is physically realized by two-level quantum system, for example, spin-1/2 system.

Mathematically, we can regard qubit as the smallest nontrivial Hilbert space $\mathcal{H} = \mathbb{C}^2$ together with its corresponding quantum states. The basis of the qubit system is usually chosen as standard basis

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \ |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1.1}$$

Physically speaking, this bases is the spin- up and down states along the z-direction, since they are very useful in quantum information and quantum computation theory, they have a special name *computational basis*. A general qubit state is of the form $|\psi\rangle = a|0\rangle + b|1\rangle$, which is the superposition of $|0\rangle, |1\rangle$, and a, b are complex numbers with $|a|^2 + |b|^2 = 1$.

A d-dimensional quantum system is sometimes called *qudit* system in the similar way as nomenclature of qubit. Whenever qudit is used, we will denote the standard basis as $|0\rangle, \cdots, |d-1\rangle$.

Now consider the N-qubit case, the corresponding Hilbert space becomes tensor product of single qubit spaces $\mathcal{H} = \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$. Tensor product of two qubit $|\psi\rangle = (a, b)$ and $|\phi\rangle = (c, d)$ can be defined as

$$|\psi\rangle \otimes |\phi\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a \begin{pmatrix} c \\ d \end{pmatrix} \\ b \begin{pmatrix} c \\ d \end{pmatrix} \end{pmatrix} = \begin{pmatrix} ac \\ ad \\ bc \\ bd \end{pmatrix}. \tag{1.2}$$

The tensor product of operators can be defined accordingly, for $A = (a_{ij})$ and $B = (B_{kl})$, their tensor product is

$$A \otimes B = \begin{pmatrix} a_{00}B \ a_{01}B \\ a_{10}B \ a_{11}B \end{pmatrix}. \tag{1.3}$$

It's obvious that $(A \otimes B)|\phi\rangle \otimes |\phi\rangle = (A|\psi\rangle) \otimes (B|\phi\rangle)$. For ease of notation, we will also omit the tensor product symbol and denote

$$|ij\rangle := |i\rangle \otimes |j\rangle, \tag{1.4}$$

and $AB = A \otimes B$ whenever there is no risk of ambiguities. At the begining, you may feel uncomfortable with these notations, but when you are facing a large number of qubits, this will be extremely convenient for us to do calculations. So just do more practice and get familiar with this.

1.1.1 Pauli operators

The best way to familiarize ourselves with the notion of qubit is to see a realistic example, the spin-1/2 system, this is familiar to most of physics students. There are three spin operators S_x, S_y, S_z, the characteristic commutation relation they satisfy is $[S_i, S_j] = i\hbar\varepsilon_{ijk}S_k$. We can find a two-dimensional representation of there operators, which are the famous *Pauli operators*:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.5}$$

It is easily checked that, if we set $S_j = \hbar\sigma_j/2,\ j = x, y, z$ (we will also frequently used the subscripts $j = 1, 2, 3$), the commutation relation of the spin operator is satisfied. In some cases, we will refer Pauli operators as spin operators for simplicity. In quantum information and quantum computation theory, to ease the notation, one will also use X, Y, Z to represent three Pauli matrices (another notation we will use later is $\tilde{Y} = -iY$, which is unitary but not Hermitian).

Pauli matrices σ_μ with $\sigma_0 = I, \sigma_1 = \sigma_x, \sigma_2 = \sigma_y, \sigma_3 = \sigma_z$ are so important, thus they are worthy a thorough exploration. Firstly, they are both unitary and Hermitian, thus they can be regarded both as time evolution and physical observable, this properties make them extremely useful. There are some crucial properties of Pauli matrices which you might have been familiar from quantum mechanics, here we list some

Exercise 1.1. Check the following facts about the Pauli matrices:

1. Pauli operators satisfy an important and useful relation (we refer to as *Pauli relation*)

$$\sigma_i\sigma_j = \delta_{ij}I + i\varepsilon_{ijk}\sigma_k, \quad i, j, k = 1, 2, 3 \tag{1.6}$$

   where I is two-by-two identity matrix. Using this formula and the fact $S_j = \hbar\sigma^j/2$, the commutation relation is easily verified.
2. Dirac relation:

$$(\vec{a}\cdot\vec{\sigma})(\vec{b}\cdot\vec{\sigma}) = (\vec{a}\cdot\vec{b})I + i(\vec{a}\times\vec{b})\cdot\vec{\sigma}. \tag{1.7}$$

   This can be proved from the Pauli relation easily.
3. For a unit vector $\vec{n}$, the operator $\sigma_{\vec{n}} = \vec{n}\cdot\vec{\sigma}$ is the spin operator along $\vec{n}$ direction, its eigenvalues are ± 1 and we have the following useful formula

$$e^{i\theta\vec{n}\cdot\vec{\sigma}} = \cos\theta I + i\sin\theta\vec{n}\cdot\vec{\sigma}. \tag{1.8}$$

   This can be proved by using the Taylor expansion and noticing that $(\vec{n}\cdot\vec{\sigma})^2 = I$.

4. The Pauli matrices generate a group, which is known as one-qubit *Pauli group*
$$\mathbf{P}_1 =\{e^{i\theta}\sigma_i|i = 0, 1, 2, 3, \theta = 0, \pi/2, \pi, 3\pi/2\}. \tag{1.9}$$
Notice that here we must introduce the phase factor to ensure that the set is closed under multiplication. The order of $\mathbf{P}_1$ is thus 16, the Pauli group will show up again and again in the quantum information theory, when we discuss the stabilizer codes, the more general Pauli groups will be introduced and their properties and representations will be discussed in detail.
5. Another way to construct a group from Pauli matrices is to use $Y^{a,b} = X^aZ^b$ with $a, b = 0, 1$,
$$Y^{0,0} = I, \quad Y^{1,0} = X, \quad Y^{0,1} = Z, \quad Y^{1,1} = \tilde{Y}. \tag{1.10}$$
Notice that we have
$$Z^aX^b = (-1)^{ab}X^bZ^a, \tag{1.11}$$
which implies that
$$Y^{a,b}Y^{k,l} = (-1)^{bk}Y^{a+k,b+l}. \tag{1.12}$$
Thus, we have a group
$$\begin{aligned}\mathbf{P}(1; 2) =&\{\pm 1\} \times \{Y^{a,b}|a, b = 0, 1\}\\ =&\mathbb{Z}_2 \times \{I, X, \tilde{Y}, Z\}.\end{aligned} \tag{1.13}$$
The qudit generalization of this group will be discussed later in exercise 1.3 □

The eigenstates of σ_z are $|0\rangle$ and $|1\rangle$, more precisely, we have
$$\sigma_z|0\rangle = +1|0\rangle,\ \sigma_z|1\rangle = -1|1\rangle. \tag{1.14}$$
When the other two Pauli operators act on $|0\rangle, |1\rangle$, we have
$$\sigma_x|0\rangle = |1\rangle,\ \sigma_x|1\rangle = |0\rangle; \quad \sigma_y|0\rangle = i|1\rangle,\ \sigma_y|1\rangle = -i01\rangle. \tag{1.15}$$
In quantum information and computation community, the notations X, Y, Z are also used to represent Pauli matrices hereinafter.

In quantum-mechanics books, the z-axis spin up and spin down states are usually denoted as $|\uparrow\rangle$ and $|\downarrow\rangle$ respectively, and we are used to denote the spin down state, which is the ground state, as vacuum state $|0\rangle$ in the Fock representation (here '0' represents 'no particle'), but the sad thing is that during the development of quantum information theory, people take the convention that $|0\rangle = (1, 0)^T = |\uparrow\rangle$. To avoid the ambiguity, we will always denote the vacuum state (ground state) of a system as $|\Omega\rangle$ in this book.

Let us now introduce two other bases of qubit space, the ± 1 eigenstates of σ_x:

$$
\begin{aligned}
|+\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix},\\
|-\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.
\end{aligned}
\tag{1.16}
$$

And ± 1 eigenstates of σ_y:

$$
\begin{aligned}
|\circlearrowleft\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix},\\
|\circlearrowright\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}.
\end{aligned}
\tag{1.17}
$$

The basis transformation between $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$, known as Hadamard transformation, is of great importance in quantum information theory. The matrix is denoted as

$$
H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}. \tag{1.18}
$$

It is easily verified that $H^2 = I$, i.e., $H^{-1} = H$, more precisely, we have

$$
H|0\rangle = |+\rangle, H|1\rangle = |-\rangle, H|+\rangle = |0\rangle, H|-\rangle = |1\rangle. \tag{1.19}
$$

The general transformation of qubit state is characterized by unitrary transformation.

Notice that Pauli operators are simultaneously Hermitian and unitary, when generalizing them into higher dimensional space, then two conditions does no hold simultaneously anymore. In the following two exercises, we discuss two possible generalizations.

Exercise 1.2 (Generalized Gell-Mann matrices). In this exercise, we will consider the qudit generalization of Pauli matrices that preserves the Hermicity. In fact, there are many possible generalizations of Pauli matrices in a higher dimensional Hilbert space, here we just introduce one of them: the generalized Gell-Mann matrices.

1. (**Hilbert-Schmidt inner product**) For the space of all operators [1] $\mathbf{B}(\mathcal{H})$ over Hilbert space $\mathcal{H}$, the inner product of the space can be defined as *Hilber-Schmidt inner product*

$$
\langle A, B\rangle = \mathrm{Tr}(A^\dagger B). \tag{1.20}
$$

[1] The notation here we use is a standard notation in functional analysis, the captibtal '**B**' represents the term 'bounded operator'. In finite dimensional space, all linear operators are bounded. Thus hereinafter we will use this notation to represent the set of all operators.

Show that this is indeed a inner product. In fact, if we regard the matrices $A = (a_{ij})$ and $B = (b_{kl})$ as $d \times d$ dimensional vectors, this is just the the usual inner product of vectors $\langle A, B\rangle = \sum_{ij} a_{ij}^* b_{ij}$.

2. (**The real vector space of Hermitian operators**) Show that the set of all Hermitian operators over the Hilbert space $\mathcal{H}$, which we denote $\mathbf{H}(\mathcal{H}) = \{A \in \mathbf{B}(\mathcal{H}) | A^\dagger = A\}$ hereinafter, is a real sub-vector space of the complex vector space $\mathbf{B}(\mathcal{H})$. When $\dim \mathcal{H} = d$, we have $\dim \mathbf{H}(\mathcal{H}) = d^2 - 1$, in this situation, we will also use the notation $\mathbf{H}(d)$.
3. (**Hilber-Schmidt basis**) We can choose a basis which serves as a generalization of Pauli basis of $\mathbf{H}(2)$. It's convenient to use the Hilbert-Schmidt basis $\{\sigma_\mu | \mu = 0, \cdots, d^2 - 1\}$ which satisfies

   - The basis includes $\sigma_0 = I$;
   - $\mathrm{Tr}(\sigma_j) = 0$ for all $j \geq 1$;
   - These matrices are orthogonal $\mathrm{Tr}(\sigma_\mu \sigma_\nu) = d\delta\mu\nu$.

   A typical explicit matrix representation of such a basis is the generalized Gell-Mann (GGM) matrices which consists of

   (a) $\frac{d(d-1)}{2}$ symmetric GGM
$$\Lambda_s^{jk} = \sqrt{\frac{d}{2}} \left(|j\rangle\langle k| + |k\rangle\langle j|\right), \quad 0 \leq j < k \leq d-1; \tag{1.21}$$
   (b) $\frac{d(d-1)}{2}$ antisymmetric GGM
$$\Lambda_a^{jk} = \sqrt{\frac{d}{2}} \left(-i|j\rangle\langle k| + i|k\rangle\langle j|\right), \quad 0 \leq j < k \leq d-1; \tag{1.22}$$
   (c) $(d-1)$ diagonal GGM
$$\Lambda^l = \sqrt{\frac{d}{(l+1)(l+2)}} \left(\sum_{j=0}^{l} |j\rangle\langle j| - (l+1)|l+1\rangle\langle l+1| \right),$$
   with $\quad 0 \leq l \leq d-2$;

   (d) The identity matrix I.

   There are in total $\frac{d(d-1)}{2} + \frac{d(d-1)}{2} + (d-1) + 1 = d^2$ matrices. Show that GGM matrices satisfy the condition of Hilbert-Schmidt basis and using the Hilbert-Schmidt basis to show that Hilbert-Schmidt inner product is indeed an inner product, viz., it's real valued, postive definite, symmetric and bilinear. □

Exercise 1.3 (Weyl operators). Let's now consider another kind of generalization of Pauli operators which preserves the unitarity, known as Weyl operators. Let $\omega_d = e^{2\pi i/d}$ be the d-th root of unity, define

$$Z = \sum_{q=0}^{d-1} \omega_d^q |q\rangle\langle q|. \tag{1.23}$$

$$X = \sum_{q=0}^{d-1} |q+1\rangle\langle q|. \tag{1.24}$$

$$Y^{a,b} = X^a Z^b, \quad a, b \in \mathbb{Z}_d. \tag{1.25}$$

These operators are obviously unitary and they are called Weyl operators. Show the following statements:

1. We have the following commutation relation

$$ZX = \omega_d XZ. \tag{1.26}$$

$$Y^{a,b} Y^{k,l} = \omega_d^{bk-al} Y^{k,l} Y^{a,b}. \tag{1.27}$$

2. The Weyl operators generate a group (qudit Pauli group)

$$\mathbf{P}(1; d) = \mathbb{Z}_d \times \{Y^{a,b}, a, b \in \mathbb{Z}_d\}, \tag{1.28}$$

   where $\mathbb{Z}_d \simeq \langle \omega_d \rangle = \{\omega_d^0, \cdots, \omega_d^{d-1}\}$. When $d = 2$, this group coincide with the qubit Pauli group $\mathbf{P}(1; 2)$ which we have introduced before.
3. Show $Z^d = X^d = I$, this implies that $\mathbf{P}(1; d)$ contains the cyclic subgroup $\mathbb{Z}_d$.

$\square$

1.1.2 *Pure and mixed qubit state*

One of the characteristic feature of quantum mechanics is superposition principle. A general qubit state

$$|\psi\rangle = a|0\rangle + b|1\rangle,$$

when measuring in σ_z basis, it has probability $p = |a|^2$ (suppose that $p \neq 1, 0$) to be in $|0\rangle$ and probability $(1-p) = |b|^2$ in $|1\rangle$. But we can also build a classical statistical system with two states $|0\rangle, |1\rangle$ which be in 0 with probability p and 1 with probability $1-p$. So, what is the difference between a qubit and a bit in terms of probability distribution? To see this, let us consider an observable A, the expectation value of A upon $|\psi\rangle$ is

$$\langle A\rangle_Q = \langle\psi|A|\psi\rangle = p\langle 0|A|0\rangle + (1-p)\langle 1|A|1\rangle + a^* b\langle 0|A|1\rangle + b^* a\langle 1|A|0\rangle,$$

but upon classical mixture of $|0\rangle, |1\rangle$

$$\langle A\rangle_C = p\langle 0|A|0\rangle + (1-p)\langle 1|A|1\rangle.$$

Now if we take $A = \sigma_z$, we see that $\langle\sigma_z\rangle_Q = \langle\sigma_z\rangle_C$, two cases are of no difference, but if we take $A = \sigma_x$, we see that $\langle\sigma_x\rangle_Q \neq \langle\sigma_x\rangle_C$. The quantum mechanical coherence shows up here. To describe quantum superposition and classical mixture in a unified framework, we need introduce the concepts of *density operator* or *density matrix*. Let us rewrite the quantum superposition $|\psi\rangle = a|0\rangle + b|1\rangle$ of $|0\rangle, |1\rangle$ in a matrix form

$$\rho_Q = |\psi\rangle\langle\psi| = \begin{pmatrix} aa^* & ab^* \\ ba^* & bb^* \end{pmatrix} = \begin{pmatrix} p & ab^* \\ ba^* & 1-p \end{pmatrix}. \tag{1.29}$$

Similarly, we can write classical probabilistic mixture of $|0\rangle, |1\rangle$ in a matrix form

$$\rho_C = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}. \tag{1.30}$$

We see that the quantum and classical differs mainly in their off-diagonal entities, in classical case, all off-diagonal entities are equal to zero. The state ρ_Q here is called a pure qubit state, and ρ_C is called a mixed qubit state.

How to take the expectation value of observable A in this density operator representation ρ? The answer is to take the trace

$$\langle A\rangle = \mathrm{Tr}(A\rho). \tag{1.31}$$

It is easily verified that $\langle A\rangle_Q = \mathrm{Tr}(A\rho_Q)$ and $\langle A\rangle_C = \mathrm{Tr}(A\rho_C)$.

In general, we can take probabilistic mixture of several states $|\psi_1\rangle, \cdots, |\psi_n\rangle$ with probabilities $p_1, \cdots, p_n$, the corresponding density operator will be

$$\rho = \sum_{i=1}^{n} p_i|\psi_i\rangle\langle\psi_i|. \tag{1.32}$$

A pure state operator can be regarded as a probabilistic mixture of $|\psi\rangle$ with probability with $p = 1$ and all other states with probability 0.

We see that the qubit density operator is a Hermitian operator, and the expectation value $\langle\psi|\rho|\psi\rangle \geq 0$ for all qubit state ψ, this means that ρ is a positive semidefinite operator. Notice that we also have $\mathrm{Tr}\,\rho = 1$ since ρ is a probabilistic mixture of pure states. The set of all qubit density operators will be denoted as $\mathbf{D}(\mathbb{C}^2)$ hereinafter.

Exercise 1.4. Prove that for mixed qubit state $\mathrm{Tr}(\rho_C^2) < 1$ and for pure qubit state $\mathrm{Tr}(\rho_Q^2) = 1$.

1.1.3 Bloch sphere representation of qubit

For a given pure qubit state $|\psi\rangle = a|0\rangle + b|1\rangle$ which is a linear combination of qubit basis $|0\rangle$ and $|1\rangle$, a, b are complex numbers and $|a|^2 + |b|^2 = 1$. When measuring in the qubit basis, we get 0 with probability $p = |a|^2$ and 1 with probability $1 - p = |b|^2$. There is a useful geometric representation of the qubit state, known as *Bloch sphere representation.*

Since complex coefficients a and b of ψ satisify $|a|^2 + |b|^2 = 1$, it is sufficient to use only three independent real parameters to characterize the state. We can rewrite the state ψ in the following form

$$|\psi\rangle = e^{i\gamma}(\cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}e^{i\varphi}|1\rangle), \tag{1.33}$$

where γ, θ, φ are three independent real parameters. Since the overall factor of a quantum state has no physically observable effect, $e^{i\gamma}$ can be ignored, thus we can effectively rewrite the state as

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}e^{i\varphi}|1\rangle. \tag{1.34}$$

We can interpret θ, φ as spherical coordinates of a vector in unit sphere in $\mathbb{R}^3$, the state corresponds a vector in the unit sphere are shown in Fig. 1.1. For example, the qubit basis $|0\rangle$ and $|1\rangle$ corresponds to $\hat{\mathbf{z}}$ and $-\hat{\mathbf{z}}$ respectively; $|+\rangle$ and $|-\rangle$ corresponds to $\hat{\mathbf{x}}$ and $-\hat{\mathbf{x}}$ respectively; and $|\circlearrowleft\rangle$ and $|\circlearrowright\rangle$ corresponds to $\hat{\mathbf{y}}$ and $-\hat{\mathbf{y}}$ respectively.

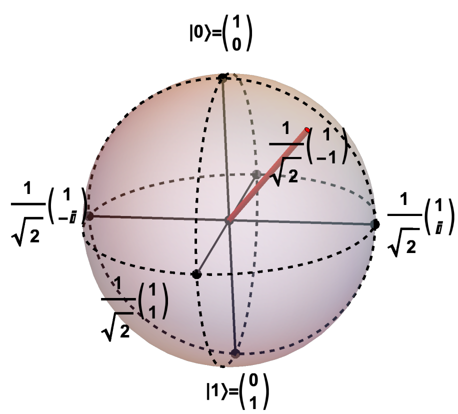


Fig. 1.1 Bloch sphere representation of a qubit $|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}e^{i\varphi}|1\rangle$.

The Bloch representation has its physical motivation, we see that vectors $\pm\hat{\mathbf{x}}$, $\pm\hat{\mathbf{y}}$,$\pm\hat{\mathbf{z}}$ in Bloch sphere corresponds to the eigenstates of spin operators σ_x, σ_y and σ_z for the eigenvalues ± 1. As we will show, this correspondence is universal. Let us now consider the spin operator $\sigma(\hat{\mathbf{n}})$ pointing to $\hat{\mathbf{n}} = \hat{\mathbf{n}}\cdot\boldsymbol{\sigma} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ which is a unit vector corresponds to (θ, ϕ) in unit sphere,

$$\sigma(\hat{\mathbf{n}}) = \hat{\mathbf{n}}\cdot\boldsymbol{\sigma} = n_x\sigma_x + n_y\sigma_y + n_z\sigma_z = \begin{pmatrix} \cos\theta & e^{-i\varphi}\sin\theta \\ e^{i\varphi}\sin\theta & -\cos\theta \end{pmatrix}. \tag{1.35}$$

The eigenvectors of $\sigma(\hat{\mathbf{n}})$ corresponds to eigenvalues ± 1 are

$$|\hat{\mathbf{n}}+\rangle = \begin{pmatrix} \cos\frac{\theta}{2} \\ e^{i\varphi}\sin\frac{\theta}{2} \end{pmatrix}, \quad |\hat{\mathbf{n}}-\rangle = \begin{pmatrix} \sin\frac{\theta}{2} \\ -e^{i\varphi}\cos\frac{\theta}{2} \end{pmatrix}, \tag{1.36}$$

which are exactly two state corresponds to $\pm\hat{\mathbf{n}}$ in Bloch sphere representation.

Now let's generalize the above analysis to the case of mixed qubit states. Intuitively, the vector corresponding to a mixed state

$$\rho = p_1|\psi(\mathbf{n}_1)\rangle\langle\psi(\mathbf{n}_1)| + p_2|\psi(\mathbf{n}_2)\rangle\langle\psi(\mathbf{n}_2)| \tag{1.37}$$

can be guessed as $p_1\mathbf{n}_1 + p_2\mathbf{n}_2$, which is a convex combination of two unit vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ for two pure states. In fact, this is true. To prove it, we must find the expansion of density matrix in Pauli matrices.

As we have discussed for pure states, the expansion of $\rho(\mathbf{n}_1) = |\psi(\mathbf{n}_1)\rangle\langle\psi(\mathbf{n}_1)|$ and $\rho(\mathbf{n}_1) = |\psi(\mathbf{n}_2)\rangle\langle\psi(\mathbf{n}_2)|$ in Pauli matrices are

$$\rho(\mathbf{n}_1) = \frac{1}{2}(I + \mathbf{n}_1\cdot\boldsymbol{\sigma}), \tag{1.38}$$

$$\rho(\mathbf{n}_2) = \frac{1}{2}(I + \mathbf{n}_2\cdot\boldsymbol{\sigma}). \tag{1.39}$$

Thus the mixed density matrix is

$$\begin{aligned} \rho =& p_1\rho(\mathbf{n}_1) + p_2\rho(\mathbf{n}_1) \\ =& p_1\frac{1}{2}(I + \mathbf{n}_1\cdot\boldsymbol{\sigma}) + p_2\frac{1}{2}(I + \mathbf{n}_1\cdot\boldsymbol{\sigma}) \\ =& \frac{1}{2}(I + \mathbf{n}\cdot\boldsymbol{\sigma}) \end{aligned} \tag{1.40}$$

with $\mathbf{n} = p_1\mathbf{n}_1 + p_2\mathbf{n}_2$. This can be generalized to N pure state mixture straightforwardly.

It's worth mentioning that, the vector $\mathbf{n}$ inside the Bloch sphere can be decomposed into probabilistic mixture of vectors on the Bloch sphere in infinitely many ways. For example, a vector inside the Bloch sphere can be written as convex combination of any two vectors corresponds to the crossing points of a line through endpoint of $\mathbf{n}$ with the Bloch sphere. This reflect in the fact that, a mixed matrix can be decomposed into mixture of pure states in infinitely many ways.

Definition 1.1 (Bloch representation). A general qubit state

$$\rho = \frac{1}{2}(I + \mathbf{a} \cdot \boldsymbol{\sigma}) \tag{1.41}$$

is represented as a vector $\mathbf{a}$ in $\mathbb{R}^3$ (know as Bloch vector). When $|\mathbf{a}| = 1$, ρ is a pure state, its Bloch vector set in the unit sphere, when $|\mathbf{a}| < 1$, ρ is mixed state, its Bloch vector lies inside the Bloch sphere.

The Bloch representation can be understood from a different aspect, recall that the Hermitian operator space $\mathbf{H}(\mathbb{C}^2)$ is a real vector space and Pauli matrices $\{I, \sigma_x, \sigma_y, \sigma_z\}$ forms a Hilber-Schmidt basis of $\mathbf{H}(\mathbb{C}^2)$. Thus by mapping σ_μ to standard basisi $\vec{e}_\mu$ of $\mathbb{R}^4$, we see that $\mathbf{H}(\mathbb{C}^2)$ is isomorphic to $\mathbb{R}^4$. A qubit state ρ can be represented by

$$\rho = \frac{1}{2}(x_0 I + \sum_{i=1}^{3} x_i \sigma_i) \tag{1.42}$$

where $x_i = \mathrm{Tr}(\sigma_\mu \rho)$ for $\mu = 0, \cdots, 3$. Since $\mathrm{Tr}(\rho) = 1$, we see that $x_0 = 1$. Thus there are three freee variables x_1, x_2, x_3, they lie in $\mathbb{R}^3$. We know tht $\mathrm{Tr}(\rho^2) \leq 1$, then from expression (1.42) we see that

$$\mathrm{Tr}(\rho^2) = \frac{1}{2}(1 + \|\vec{x}\|^2) \leq 1, \tag{1.43}$$

this implies that $\|\vec{x}\| \leq 1$. Notice that the positivity of ρ in equation (1.42) is equivalent to

$$\det \rho = \frac{1}{4}(1 - \|\vec{x}\|^2) \geq 0, \tag{1.44}$$

this is now automatically satisfied for $\|\vec{x}\| \leq 1$. These results means that for a given (mixed or pure) qubit state, we can find a 3-dimensional vector representation with norm no more than one. And for a given 3-dimensional vector with norm no more than one, we obtain a correponding qubit state. In summary, the Bloch representation provides a one-to-one correpondence between the set of all qubit states and a 3-ball (known as Bloch sphere in this content).

Exercise 1.5 (Bloch representation for qudit state). If we choose the bisis of Hermitian operator space $\mathbf{H}(d)$ as Hilbert-Schmidt basis as in exercise 1.2, $\{\sigma_\mu | \mu = 0, \cdots, d^2 - 1\}$. The Bloch representation of qudit density operator is

$$\rho = \frac{1}{d}\sum_{\mu=0}^{d-1} a_\mu \sigma_\mu = \frac{1}{d}(I + \vec{a}\cdot\vec{\sigma}). \tag{1.45}$$

Show that

(a) We must have $a_0 = 1$, since all σ_μ are traceless except $\sigma_0 = I$ and the density operator is trace-one.
(b) From the condition that purity $\mathrm{Tr}(\rho^2) \leq 1$, we have $\|\vec{a}\|^2 \leq d-1$. □

1.1.4 Appendix: mathematical preliminaries

For the convenience of the later discussion, in this part, we collect some mathematical results mainly from linear algebra and functional analysis that will be useful in quantum information theory.

1.1.4.1 Unitary transformation

Here we introduce the rigorous definition of unitary transformation first.

Definition 1.2 (Unitary transformation). Let $U : \mathcal{H} \to \mathcal{X}$ be a linear map, it's called unitary if and only if it satisfy one of the following equivalent conditions:

1. it preserves inner product, $(Ux, Uy)_{\mathcal{X}} = (x, y)_{\mathcal{H}}$ for all $x, y \in \mathcal{H}$;
2. it preserves norm, $\|Ux\|_{\mathcal{H}} = \|x\|_{\mathcal{K}}$ for all $x \in \mathcal{H}$;
3. it satisfies $U^\dagger U = I = UU^\dagger$.

A unitary transformation with the same domain and codomain is called a unitary operator.

Exercise 1.6. Prove that the conditions in the definition 1.2 of unitary transformation are equivalent. It would be helpful to use the *polarization formula*

$$(x, y) = \frac{1}{4}\sum_k i^k \|x + i^k y\|^2 \tag{1.46}$$

where the sum extends for $k = 0, 2$ if the scalars are real and extends for $k = 0, 1, 2, 3$ if the scalars are complex. □

Now let us consider the set of all 2×2 unitary operators of qubit states, known as unitary group

$$U(2) = \{U | U^\dagger U = I\}, \tag{1.47}$$

and its subset, whose elements have determinant one, called special unitary group

$$SU(2) = \{\det U = 1 | U^\dagger U = I\}. \tag{1.48}$$

As you may have known, $SU(2)$ group is extremely important in quantum mechanics, since it is homomorphic to special three dimensional rotation group $SO(3)$ which is defined as the set of all real 3×3 matrices O with determinant one and $OO^T = O^T O = I$. More precisely, we have

$$SO(3) \simeq SU(2)/\pm I. \tag{1.49}$$

This correspondence make it clear how to see spins intuitively in $\mathbb{R}^3$ space and will be a pertinent for us to give a 3-dimensional vector representation of qubit state, called Bloch representation.

Exercise 1.7 (Pauli operators form a basis of $SU(2)$ group). Show that every $U \in SU(2)$ can be decomposed into a linear combination of Pauli operators $\{I, \sigma_x, \sigma_y, \sigma_z\}$.

Proof. Let us take a closer look at the group $SU(2)$ here, since it will play an important role in the quantum information and quantum computation theory. By definition, an element $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SU(2)$, where a, b, c, d are complex numbers (eight real parameters), satisfies two conditions

$$U^\dagger U = I, \ \ \det U = 1.$$

From $U^\dagger U = I$, we have

$$a^* a + c^* c = 1, \tag{1.50}$$
$$b^* b + d^* d = 1, \tag{1.51}$$
$$a^* b + c^* d = 0, \tag{1.52}$$
$$b^* a + d^* c = 0, \tag{1.53}$$

where the last two equalities are equivalent, the first two are real constraints, thus they give four real constraints for real parameters of the matrix entries.

From $\det U = 1$, we obtain

$$ad - bc = 1, \tag{1.54}$$

which are two real constraints. However, these two real constraints are not independent from Eqs. (1.50)-(1.53). Using Eqs. (1.52) and (1.51), we have

$$c = -\frac{ab^*}{d^*}, \tag{1.55}$$
$$b = \frac{1 - d^* d}{b^*}, \tag{1.56}$$

then by substituting these two equalities to Eq. (1.54), we see that

$$a = d^*. \tag{1.57}$$

Similarly, by calculating a, d from Eqs. (1.52) and (1.51) and substituting the results to Eq. (1.54), we obtain

$$c = -b^*. \tag{1.58}$$

Therefore, an element in $SU(2)$ is of the form

$$U = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}, \ a, b \in \mathbb{C}, \ aa^* + bb^* = 1. \tag{1.59}$$

Actually, there is a more concise way to derive the above result. Since $U^{-1} = U^\dagger$, for which we must use the formula of inverse matrix ($A^{-1} = A^{\mathrm{ad}} / \det A$, where A^{ad} is the adjugate matrix of A)

$$U^{-1} = \frac{1}{\det U} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

From condition that $\det U = 1$ and by comparing it with $U^\dagger$, we obtain what is required.

From the general expression 1.59 and by setting $a = t + iz$ and $b = y + iz$, we see that

$$U = tI + ix\sigma_x + iy\sigma_y + iz\sigma_z, \quad t^2 + x^2 + y^2 + z^2 = 1. \tag{1.60}$$

Every $U \in SU(2)$ can be decomposed as a linear combination of Pauli operators. $\square$

Quantum circuit notation

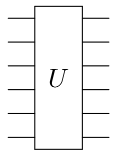
Since we have give a terse introduction to the unitary operators, it's now a good postisition to introduce the quantum circuit notation. Quantum circuit notation is a graphical way to represent quantum states and their time evolutions, measurements, it can helps us to understand these processes more intuitively. This representation play a crucial role in quantum computation theory, which we will discuss in the second volume of this book.

The state of a single qubit is denoted as a wire, called *quantum wire* (to distinguish it from classical bit, we denote the classical bit as a double wire). A unitary operator U is denoted as box with label U, with input state left and output state right. Graphically, the equation $U|\psi\rangle = |\varphi\rangle$ for single qubit can be represented as

$$|\psi\rangle \; - \boxed{U} - \; |\varphi\rangle \tag{1.61}$$

Similarly, n-qubit states are drawn as n quantum wires, a unitary operator acting on n-qubit state is represented as a box with n input wires and n

output wires,

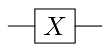


(1.62)

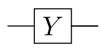
Since in quantum computation theory, these unitary operators play the role of gates as in classical computation, they will also be called *quantum gates*.

There are several frequently used unitary operator that deserve special notations:

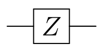
- Three Pauli operators, spin flip gate $X = \sigma_x$, phase flip gate $Z = \sigma_z$, and their combination $Y = \sigma_y$,



(1.63)

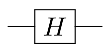


(1.64)



(1.65)

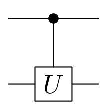
- The Hadamard gate corresponding to Hadamard operator,



(1.66)

- There are a kind of crucial two qubit unitary operators, called controlled-U, which we denote $C(U)$, the definition is

$$C(U) = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes U \tag{1.67}$$

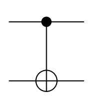


(1.68)

- A special case of controlled-U is controlled-X, which is ubiquitous in quantum information and quantum computation. It's the quantum counterpart of classical controlled-not operation, we will called it the controlled-not gate and denote it as CNOT.

$$\mathrm{CNOT} = C(X) = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X. \tag{1.69}$$

In quantum circuit representation, we have



(1.70)

1.1.4.2 Theorems about linear operators

It's a good place for us to recall the polar decomposition and singular value decomposition. Hereinafter, we will use the following notations

- The Hilbert spaces are denoted as $\mathcal{H}$, $\mathcal{K}$, $\mathcal{X}$, $\mathcal{Y}$, etc.
- The set of all linear operations between two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ is denoted as $\mathcal{L}(\mathcal{H},\mathcal{K})$, and we have $\mathcal{L}(\mathcal{H}) := \mathcal{L}(\mathcal{H},\mathcal{H})$. Similarly, we have the sets of all bounded linear operators $\mathcal{B}(\mathcal{H},\mathcal{K})$ and $\mathcal{B}(\mathcal{H})$.
- The set of all density operators, i.e., positive semidefinite trace-one operators, over Hilbert space $\mathcal{H}$ is denoted as $\mathcal{S}(\mathcal{H})$. In mathematical literatures, the set of all positive semidefinite operators is usually denoted as $\mathsf{Pos}(\mathcal{H})$.

It's well known that a nonzero complex number can decompose as $c = e^{i\theta}r$. This can be generalized to arbitrary nonzero linear operators.

Theorem 1.1 (Polar decomposition). *Let $A \in \mathcal{L}(\mathcal{H})$ be a nonzero linear operator. Then there exist a unitary U and positive operators R and L such that*

$$A = UL_A = R_A U, \tag{1.71}$$

where L_A and R_A are unique and $L_A = \sqrt{A^\dagger A}$ and $R_A = \sqrt{AA^\dagger}$. If A is invertible, then U is also unique. $A = UL_A$ and $A = R_A U$ are called left and right polar decomposition respectively.

Since this is a standard result in linear algebra, we won't give the proof here. The generalization of polar decomposition of linear operators (which is square matrix in a given basis) to linear transformations (which may be $n \times m$ matrices in the given basis) is the singular value decomposition.

Exercise 1.8. Show that $A^\dagger A$ and $AA^\dagger$ have the same eigenvalues for any $A \in \mathbf{B}(\mathcal{X},\mathcal{Y})$.

Proof. Suppose that λ_k is an eigenvalue of $A^\dagger A$ with eigenvector ψ_k. Then

$$AA^\dagger A\psi_k = \lambda_k A\psi_k, \tag{1.72}$$

i.e., λ_k is also an eigenvalue of $AA^\dagger$. Similarly, all eigenvalues of $AA^\dagger$ are also eigenvalues of $A^\dagger A$. This completes the proof. □

Theorem 1.2 (Singular value decomposition). *Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite dimensional complex Hilbert space, let $A \in \mathcal{L}(\mathcal{H},\mathcal{K})$ with rank n. There exist positive real numbers $s_i, \cdots, s_n$ and orthonormal vectors $\{|x_i\rangle \in \mathcal{H}\}$ and $\{|y_i\rangle \in \mathcal{K}\}$ such that*

$$A = \sum_{i=1}^{n} s_i |y_i\rangle\langle x_i|, \tag{1.73}$$

where $s_i(A) = \sqrt{\lambda_i(AA^\dagger)} = \sqrt{\lambda_i(A^\dagger A)}$ *are known as singular value of* A, $\{|x_i\rangle\}$ *and* $\{|y_i\rangle\}$ *are the eigenvectors of* $A^\dagger A$ *and* $AA^\dagger$ *respectively.*

In the matrix form, we have

$$A = U\Lambda_A V, \quad \Lambda_A = U^\dagger A V^\dagger, \tag{1.74}$$

where U, V *are unitary operators and* Λ_A *is diagonal matrix with diagonal elements* s_i*.*

Exercise 1.9. Let A be a linear operator and U unitary, show that

$$|\operatorname{Tr}(AU)| \leq \operatorname{Tr}|A|_R = \operatorname{Tr}\sqrt{AA^\dagger}, \tag{1.75}$$

$$|\operatorname{Tr}(UA)| \leq \operatorname{Tr}|A|_L = \operatorname{Tr}\sqrt{A^\dagger A}. \tag{1.76}$$

The equality is obtained by choosing U to be the unitary in polar decomposition of A, i.e., $A = |A|_R U = U|A|_L$.

Proof. dd □

1.1.5 Properties of Bloch sphere representation

Let us now analyze the properties of Bloch sphere representation of qubit state. This will help us to translate the abstract operations over qubit states into the geometric transformations in the Bloch sphere, which are much more intuitive to work with.

Unitary transformation.—

Using the isomorphism $SO(3) \simeq SU(2)/\{\pm 1\}$

Time reversal operation and spin-flip

From quantum mechanics we know that time reversal operation T is an antiunitary operator, that is $T : \mathcal{H} \to \mathcal{H}$ is a bijective operator which is *antilinear*, i.e.,

$$\mathsf{T}(\alpha|\psi\rangle + \beta|\varphi\rangle) = \alpha^*|\psi\rangle + \beta^*|\varphi\rangle, \quad \forall \alpha, \beta \in \mathbb{C}, |\psi\rangle, |\varphi\rangle \in \mathcal{H}, \tag{1.77}$$

and $\langle \mathsf{T}\psi, \mathsf{T}\varphi\rangle = \langle \psi, \varphi\rangle^*$. Notice that for antiunitary operator T, the definition of its adjoint $\mathsf{T}^\dagger$ becomes

$$\langle \mathsf{T}\psi, \varphi\rangle = \langle \psi, \mathsf{T}^\dagger\varphi\rangle^*, \quad \forall \psi, \varphi \in \mathcal{H}. \tag{1.78}$$

The adjoint of antiunitary operator is still antiunitary and we have $\mathsf{T}\mathsf{T}^\dagger = \mathsf{T}^\dagger\mathsf{T} = I$.

Time reversal operation keeps space coordinates invariant, but since momentum and angular momentum all involves first-order derivative of time, we must have

$$\mathsf{T}x_j\mathsf{T}^\dagger = x_j, \quad \mathsf{T}p_j\mathsf{T}^\dagger = -p^j, \quad \mathsf{T}L_j\mathsf{T}^\dagger = -L_j \quad \mathsf{T}\sigma_j\mathsf{T}^\dagger = -\sigma_i. \tag{1.79}$$

In many cases, the time reversal operator can be written as

$$\mathsf{T} = UK, \tag{1.80}$$

where U is a unitary operator and K is operator which takes complex conjugation of quantum state in a given basis. Notice that $K^{-1} = K^\dagger = K$, it's easy to check that it's antiunitary.

Here let's focus on the spin momentum operator constraint $\mathsf{T}\sigma_j\mathsf{T}^\dagger = -\sigma_i$. If we assume that $\mathsf{T} = UK$, Notice that in $|0\rangle, |1\rangle$ basis

$$K\sigma_x K^\dagger = \sigma_x, \quad K\sigma_y K^\dagger = -\sigma_y, \quad K\sigma_z K^\dagger = \sigma_z, \tag{1.81}$$

we thus have

$$U\sigma_x U^\dagger = -\sigma_x, \quad U\sigma_y U^\dagger = \sigma_y, \quad U\sigma_z U^\dagger = -\sigma_z. \tag{1.82}$$

Thus we can choose $U = \sigma_y$, this implies an expression for time reversal operation for spin 1/2 particle

$$\mathsf{T} = \sigma_y K. \tag{1.83}$$

Similarly, for many qubit case, we have

$$\mathsf{T} = (\sigma_y \otimes \cdots \otimes \sigma_y)K. \tag{1.84}$$

Classically, when we do time reversal operation, the spin is flipped, in quantum case, this means that T plays the same role as spin-flip operator for any direction of Bloch sphere. Actually, for a Bloch vector $\mathbf{n}$ and the corresponding state $|\mathbf{n}+\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$, you can easily verified that

$$\mathsf{T}|\mathbf{n}+\rangle = e^{if(\varphi)}|\mathbf{n}-\rangle. \tag{1.85}$$

§ 1.2 Density operator

In this section, let's pin down a little more precise what it means for the density operator (also called density matrix for finite dimensional case). We have discussed the concept for qubit case, now, let's analyze it from a more general perspective. Roughly speaking, there are two approaches to considering a density operator:

- We can regard it as a description of the state of an ensemble, which leads to the ensemble interpretation of density operator;

- We can also regard it as a description for the open part S of a larger system SE, where SE is a closed system. This leads to the open system interpretation of the density operator.

Two viewpoints are closely related, we will use them interchangeably.

1.2.1 Density operator: ensemble approach

As you may have learned from statistical mechanics, an ensemble is the set of N ($N \to \infty$) independent hypothetical copies of a system. Suppose that there are N_1 systems in state $|\psi_1\rangle$, N_2 systems in state $|\psi_2\rangle$, and so on. Thus, each time when we want to measure the system, the probability that we choose the state $|\psi_i\rangle$ to measure is $p_i = N_i/N$ ($N \to \infty$). This kind of ensemble is called mixed ensemble, to describe the state of the ensemble, we introduce the density operator

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi|, \tag{1.86}$$

here ρ is an operator over the Hilbert space $\mathcal{H}$.

In contrast, for pure ensemble, the state is described by a *ray* $|\psi\rangle$ in Hilbert space $\mathcal{H}$. Here we use the term ray to mean the equivalence class in Hilbert space,

$$|\psi\rangle = \{\phi \in \mathcal{H} | \exists \lambda \in \mathbb{C} \setminus \{0\}, \phi = \lambda\psi\}. \tag{1.87}$$

Whenever there is no risk to make ambiguity, we won't distinguish terms vectors and rays in a Hilbert space. We can consider the superposition of states $|\psi_i\rangle$

$$|\psi\rangle = \sum_i c_i |\psi_i\rangle, \tag{1.88}$$

where coefficients c_i satisfy $|c_i|^2 = p_i$.

What is the difference between state ρ in equation (1.86) and ψ in equation (1.87)? From Born's rule, we know that quantum superposed state ψ has property that the particle lies in the state $\psi_i(x)$ with probability $p_i = |c_i|^2$ if $|\psi_i\rangle$ are a set of orthonormal states. This looks very similar as the interpretation of ρ. To distinguish the two cases, we must consider the measurement of the ensemble. For the mixed ensemble, each time when we want to measure an observable A, we must choose a state $|\psi_i\rangle$ from the states of the system, the probability for it is p_i. Therefore the expectation value for A is

$$\langle A \rangle_\rho = \sum_i p_i \langle\psi_i|A|\psi_i\rangle. \tag{1.89}$$

For the pure ensemble, the expectation of an operator A over the state $|\psi\rangle$ is

$$\begin{aligned}\langle\psi|A|\psi\rangle &= \sum_i |c_i|^2\langle\psi_i|A|\psi_i\rangle + \sum_{i\neq j} c_i^* c_j\langle\psi_i|A|\psi_j\rangle \\ &= \sum_i p_i\langle\psi_i|A|\psi_i\rangle + \sum_{i\neq j} c_i^* c_j\langle\psi_i|A|\psi_j\rangle.\end{aligned} \tag{1.90}$$

There are some cross terms appear in the superposition state, which is the result of the quantum coherence of the states $|\psi_i\rangle$.

Now, let us take a closer look at the expression of expectation value of an observable for the mixed ensemble. It's easily checked that

$$\langle\psi|A|\psi_i\rangle = \mathrm{Tr}(A|\psi_i\rangle\langle\psi_i|), \tag{1.91}$$

from which and using the linearity of trace operation, we can rewrite the expression in equation (1.89) as

$$\langle A\rangle_\rho = \mathrm{Tr}(A\rho). \tag{1.92}$$

This is the mixed state generalization of the pure state expectation value of an operator.

From the above discussion, we arrive at the result that, the sate of a mixed ensemble is described by a density operator, and since pure state $|\psi\rangle$ can be written as

$$\rho_\psi = |\psi\rangle\langle\psi|, \tag{1.93}$$

it can be regarded as probabilistic mixture one just one ingredient $|\psi\rangle$, pure ensemble can then be treated as a special case of mixed ensemble.

Defining properties of density operator

We have seen that for a given set of states $|\psi_i\rangle$ and a probability distribution p_i, we have a corresponding density operator ρ. It's natural to ask, for a given operator $\rho \in \mathbf{B}(\mathcal{H})$ ($\mathbf{B}(\mathcal{H})$ is the set of all bounded linear operators acting on $\mathcal{H}$), under what conditions it becomes a density operator. To this end, let us analyze what properties the state

$$\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i| \tag{1.94}$$

satisfy. Note that here ψ_i are not necessarily orthogonal to each other.

Firstly, we observe that ρ is Hermitian, this is because p_i are real number and $(|\psi_i\rangle\langle\psi_i|)^\dagger = |\psi_i\rangle\langle\psi_i|$. Secondly, for arbitrary state $|\phi\rangle \in \mathcal{H}$, we can take the expectation value of ρ over it,

$$\langle\phi|\rho|\phi\rangle = \sum_i p_i\langle\phi|\psi_i\rangle\langle\psi_i|\phi\rangle = \sum_i p_i|\langle\phi|\psi_i\rangle|^2 \geq 0. \tag{1.95}$$

The expectation value is always real (which reflects the fact that ρ is Hermitian[2]) and the value is nonnegative. From linear algebra, we known that this means that ρ is positive semidefinite. Finally, when we take the trace of ρ, we find that $\mathrm{Tr}(\rho) = 1$. From these observation, we have the following definition of density operator:

Definition 1.3 (density operator). For a quantum system with Hilbert space $\mathcal{H}$, the density operator of the system is an operator $\rho \in \mathbf{B}(\mathcal{H})$ which satisfy the following conditions:

1. The operator ρ is a Hermitian;
2. All eigenvalues of ρ are nonnegative, or equivalently $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$;
3. The trace of ρ is 1, $\mathrm{Tr}(\rho) = 1$.

The first two conditions are known as semidefinite condition. In short, a density operator is a semidefinite trace-one operator. The set of all density operators over the Hilbert space $\mathcal{H}$ will be denoted as $\mathbf{D}(\mathcal{H})$ hereinafter.

From the above definition, we see that it's crucial for us to determine if a given operator is positive semidefinite or not. It's worthy to take a close look at positive semidefinite operators.

Exercise 1.10 (Positive semidefinite operators). Show that the following statements are equivalent for finite dimensional Hilbert space $\mathcal{H}$:

(a) The operator $\rho \in \mathbf{B}(\mathcal{H})$ is positive semidefinite, viz., ρ is Hermitian and for any $\psi \in \mathcal{H}$, the expectation value $\langle\psi|\rho|\psi\rangle$ is nonnegative.
(b) There exist linear operator $A : \mathcal{H} \to \mathcal{X}$ such that $\rho = A^\dagger A$.
(c) All eigenvalues of ρ are nonnegative, $\lambda_i(\rho) \geq 0$ for all i.
(d) For any positive semidefinite operator σ over $\mathcal{H}$, the Hilbert-Schmidt inner product $(\sigma, \rho) = \mathrm{Tr}(\sigma^\dagger \rho) = \mathrm{Tr}(\sigma\rho)$ is a nonnegative real number.

Hint: For the last statement, to show that (σ, ρ) is real valued we need to show tha Hilbert-Schmidt inner product is real valued for $\mathbf{H}(\mathcal{H})$, see exercise 1.2.

(b) Here A can be chosen as positve semidefinite square root $\sqrt{\rho} = \sum_i \sqrt{\lambda_i} \mathit{\Pi}_i$ of $\rho = \sum_i \lambda_i \mathit{\Pi}_i$.
□

With this definition, we can ask whether a given density operator ρ is a pure state which is a description of the status of pure ensemble or mixed state

[2] It's worth metioning that this is not a rigorous statement, since the eigenvalues of Hermitian operators are always real, but the operators which have only real eigenvalues are not necessarily Hermitian.

which is a description of the status of mixed ensemble. This is defined from the the observation that for pure state $\rho_\psi = |\psi\rangle\langle\psi|$, the trace of $\rho_\psi^2 = \rho_\psi$ is one, i.e., $\mathrm{Tr}(\rho_\psi^2) = 1$. Meanwhile, for (non-trivial) mixed state $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ with ψ_i orthogonal to each other, the trace of ρ^2 is $\mathrm{Tr}(\rho^2) = \sum_i p_i^2 < 1$.

Definition 1.4 (pure and mixed state). The density operator ρ is called pure state if $\mathrm{Tr}(\rho^2) = 1$ and it's called mixed if $\mathrm{Tr}(\rho^2) < 1$. The quantity $\mathrm{Tr}(\rho^2)$ will be called purity of the state ρ.

Notice that for $n \times n$ density operator, the purity satisfies $1/n^2 \leq \mathrm{Tr}(\rho^2) \leq 1$. The state which this the minimal purity $\mathrm{Tr}(\rho^2) = 1/n$ is called *maximally mixed state*. In this case $p_1 = \cdots = p_n = 1/n$ and ψ_is are orthonormal . The state is of the form

$$\rho = \frac{1}{n}\sum_{i=1}^{n} |\psi_i\rangle\langle\psi_i| = \frac{1}{n} I. \tag{1.96}$$

Notice that the matrix representation of the state is independent of the basis choice, since I is independent of basis choice.

Example 1.1 (Maximally mixed qubit state). For the qubit case, in computational basis, the maximally mixed state is

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \tag{1.97}$$

In Bloch representation, we have

$$\rho = \frac{1}{2}(I + \vec{0} \cdot \vec{\sigma}), \tag{1.98}$$

it's obvious that the Bloch vector is the zero vector, which lies at the center of the Bloch sphere. □

Example 1.2 (Maximally mixed qudit state). Similar as the qubit case, if we choose the bisis of the Hermitian operator space as Hilbert-Schmidt basis $\{\sigma_0 = I, \sigma_1, \cdots, \sigma_{d^2-1}\}$, the Bloch vector corresponds to the maximally mixed state $\rho = \frac{1}{d}\sum_{i=0}^{d-1} |i\rangle\langle i|$ is the zero vector which lies in the center of Bloch sphere. See exercise 1.5 for details of the Bloch representation of qudit state. □

Let us now reexamine the difference between classical probabilistic mixture of quantum states and quantum superposition of quantum states. Consider the state

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \tag{1.99}$$

which is a quantum superposition of $|0\rangle$ and $|1\rangle$ in σ_z basis, but when we look at it in the σ_x basis, it's a basis state, thus there is no quantum coherence. This suggests that quantum coherence must be defined in given basis.

Definition 1.5 (coherent state). A quantum state ρ is said to possess quantum coherence in the measurement basis $\{|\psi_i\rangle, i = 1, \cdots, n\}$ if the matrix representation of ρ in this basis have non-vanishing non-diagonal entries. If the matrix representation ρ is diagonal in the basis basis $\{|\psi_i\rangle, i = 1, \cdots, n\}$, it is called non-coherent state in this basis.

A typical example of non-coherent state is

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|0\rangle\langle 0| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}. \tag{1.100}$$

Note that this density operator is diagonal in any measurement basis, thus it is non-coherent in any basis. Because of this reason, the state is call a *maximally mixed state*. This reflects the fact that this state is maximally entangled with its environment.

Properties of ensemble interpretation of density operator

Let's now discuss some crucial properties of density operator.

Convexity.—From the definition of the density operator, it is clear that the convex combination $\rho = \alpha\rho_1 + (1-\alpha)\rho_2$ (where $0 \leq \alpha \leq 1$) of two density operators ρ_1 and ρ_2 is still a density operator. Therefore the set of all density operators, denoted as $\mathbf{D}(\mathcal{H})$, forms a convex subset of $\mathbf{H}(\mathcal{H})$.

The convexity play an important role in studying the properties of the density operator, the topic will be discussed later in this book.

Exercise 1.11. Let's introduce the following notations:

- The set of all linear operators $\mathbf{L}(\mathcal{H})$ which is a complex vector space, when equipped with Hilbert-Schmidt inner product

$$(A, B) := \mathrm{Tr}(A^\dagger B) = \sum_{i,j} A_{ij}^* B_{ij}, \tag{1.101}$$

  it becomes a complex inner product space;
- The set of all bounded linear operators $\mathbf{B}(\mathcal{H})$ which is a complex vector space. In general $\mathbf{B}(\mathcal{H}) \subsetneq \mathbf{L}(\mathcal{H})$; for finite dimensional Hilbert space, $\mathbf{B}(\mathcal{H}) = \mathbf{L}(\mathcal{H})$;

- The set of all Hermitian operators $\mathbf{H}(\mathcal{H})$ is a real vector space (not complex vector space), thus the convex analysis, which is a powerful tool for real vector space, works very well in this space;
- The set of all positive semidefinite operators $\mathbf{Pos}(\mathcal{H})$ forms a convex subset of $\mathbf{H}(\mathcal{H})$;
- The set of all density operators $\mathbf{D}(\mathcal{H}) := \{\rho \in \mathbf{Pos}(\mathcal{H}) | \operatorname{Tr}(\rho) = 1\}$ is a convex subset of $\mathbf{Pos}(\mathcal{H})$.

Check the above statements. Recall that a subet X of a real vector space is called *convex* if and only if for any $x, y \in X$ and $p \in [0, 1]$, we have $px + (1 - p)y \in X$. A point x of a convex set X is called an *extreme point* if and only if it's not a proper convex combination of other points, viz., if there exist $p \in (0, 1)$ and $y, z \in X$ such that $x = py + (1 - p)z$, we must have $x = y = z$. $\square$

Ensemble realization of a density operator.—Another important property of density operators is that the ensemble realization of the density operator is not unique. This can most easily be seen from the fact that $\rho = I/2 = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -|$, the density operator can be realized as the equally probable classical mixture of spin up and down states along z-axis or spin left and right state along x-axis. Actually, according to the Bloch sphere representation of qubit state, any mixed state lie inside the unit sphere can be realized by the mixture of two pure qubit states lie in the Bloch sphere for which the point representing the mixed state lies in the segment connected two point corresponding two the two pure states.

For convenience, let's introduce the following definition

Definition 1.6 (ρ-ensemble). Given a density operator $\rho \in \mathbf{D}(\mathcal{H})$, a ρ-ensemble of order d (with $d \geq \operatorname{rank}(\rho)$) is a collection of states $\{|\psi_i\rangle\}_{i=1}^d$ together with a probability distribution p_i such that

$$\rho = \sum_{i=1}^{d} p_i |\psi\rangle\langle\psi_i|. \tag{1.102}$$

The ρ-ensemble is called linearly independent if the states $\{|\psi_i\rangle\}_{i=1}^d$ are linearly independent.

It's natural to ask that what is the relationship between two ρ-ensembles, this will be answered in the next section by Schrödinger-HJW theorem: they are connected by unitary transformations.

§ 1.3 Composed system and reduced states

After having understood the single-particle quantum states, let us now turn to the discussion of the system with two and more particles, this kind of system is known as *composed system*. As you will see, many of crucial quantum phenomenon, like quantum entanglement, Bell nonlocality and so on, which differs quantum mechanics essentially from classical mechanics appear in two or more particle quantum systems. We first discuss the two-particle state, and the multipartite sate will be discussed later in this chapter. For a two-particle system AB, the Hilbert space is the tensor product $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ of two respective Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ with respective orthonormal basis $|i_A\rangle, i = 0, \cdots, d_A - 1$ and $|j_B\rangle, j = 0, \cdots, d_B - 1$. A general pure state of the composed system AB is of the form

$$|\psi_{AB}\rangle = \sum_{ij} c_{ij} |i_A\rangle \otimes |j_B\rangle. \tag{1.103}$$

To avoid cluttering of equations, we will sometime omit the tensor product symbol and use the abbreviation $|i_A\rangle|j_B\rangle$ or $|i_A j_B\rangle$ to mean $|i_A\rangle \otimes |j_B\rangle$. This is a common convention in quantum information community.

Now let us first see how the density operator can be interpreted as the state of the part of a larger system as we promised.

1.3.1 Density operator: open system approach

We start with an example $|\psi_{AB}\rangle = a|0_A\rangle|0_B\rangle + b|1_A\rangle|1_B\rangle$ with $a, b \neq 1, 0$, what is the state of subsystem A if we have no knowledge of the system B? To answer this question, let us consider the measurement over system A, since physical information of the system is revealed by quantum measurement. If we want to perform a measurement O_A upon system A and do nothing upon system B, we naturally have the measurement

$$O_A \otimes I_B, \tag{1.104}$$

for the composed system, where I_B is the identity operator of system B. The expectation value of the observable over state $|\psi_{AB}\rangle$ is

$$\langle O_A\rangle = \langle\psi_{AB}|O_A \otimes I_B|\psi_{AB}\rangle = |a|^2\langle 0_A|O_A|0_A\rangle + |b|^2\langle 1_A|O_A|1_A\rangle. \tag{1.105}$$

If we set the state of A as

$$\rho_A = |a|^2|0_A\rangle\langle 0_A| + |b|^2|1_B\rangle\langle 1_B|, \tag{1.106}$$

it's easily checked that

$$\langle O_A \rangle = \mathrm{Tr}(O_A \rho_A). \tag{1.107}$$

How to explain the match of the calculated result? Suppose that an experimenter is measuring the system B in basis $|0_B\rangle, |1_B\rangle$, He obtain the measurement result is 0 with probability $|a|^2$, the state of the system becomes $|0_a\rangle|0_B\rangle$, thus the state of A becomes $|0_A\rangle$ with probability $|a|^2$; similarly, if he obtain 1 with probability $|b|^2$, the state of A is on state $|1_A\rangle$ with probabilty $|b|^2$. Since we have no knowledge about B, the state of A should be a probabilistic mixture of $|0_A\rangle$ and $|1_A\rangle$ with probabilities $|a|^2$ and $|b|^2$ respectively.

The above reasoning sounds good, let's try to apply it to the general case. For a general bipartite quantum state

$$|\psi_{AB}\rangle = \sum_{ij} c_{ij} |i_A\rangle \otimes |j_B\rangle = \sum_j (\sum_i c_{ij} |i_A\rangle)|j_B\rangle. \tag{1.108}$$

If we set $a_j = \sqrt{\sum_i |c_{ij}|^2}$ and

$$|\phi_j^A\rangle = \frac{1}{a_j} \sum_i c_{ij} |i_A\rangle, \tag{1.109}$$

which is a state of system A, we have

$$|\psi_{AB}\rangle = \sum_j a_j |\phi_j^A\rangle |j_B\rangle. \tag{1.110}$$

If an experimenter choose to measure system B in the orthonormal basis $|j_B\rangle, j = 0, \cdots, d_B - 1$, he obtain the result j with probability $|a_j|^2$, the state of A becomes $|\phi_j^A\rangle$ with probability $|a_j|^2$. Since we have no knowledge of system B, the state of A is a probabilistic mixture of $|\phi_j^A\rangle$ with probability $|a_j|^2$:

$$\rho_A = \sum_j |a_j|^2 |\phi_j^A\rangle\langle\phi_j^A| = \sum_{i,i'} (\sum_j c_{ij} c^*_{i'j}) |i_A\rangle\langle i'_A|. \tag{1.111}$$

It can be checked that for the observable O_A, when applied to the composed system AB, we have

$$\langle O_A \rangle = \langle \psi_{AB} | O_A \otimes I_B | \psi_{AB} \rangle = \mathrm{Tr}(O_A \rho_A). \tag{1.112}$$

For state $|\psi_{AB}\rangle = \sum_j a_j |\phi_j^A\rangle|j_B\rangle$, we see that

$$\langle k_B | \psi_{AB} \rangle = \sum_j a_j |\phi_j^A\rangle \langle k_B | j_B \rangle = a_k |\phi_k^A\rangle, \tag{1.113}$$

from which we have

$$\sum_k \langle k_B|\psi_{AB}\rangle\langle\psi_{AB}|k_B\rangle = \sum_k |a_k|^2 |\phi_k^A\rangle\langle\phi_k^A| = \rho_A. \tag{1.114}$$

Therefore, the density operator of A is obtained from $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ by take *partial trace*,

$$\mathrm{Tr}_B(\rho_{AB}) = \sum_k \langle k_B|\rho_{AB}|k_B\rangle. \tag{1.115}$$

Notice that partial trace does not depend on the choice of the basis for B, thus we get a well-defined definition of density operator: a density operator is obtained from a pure state of larger system by taking partial trace.

From this definition, we see that the matrix entries of the density operator ρ_A in the basis $|i_A\rangle$ are $\sum_j c_{ij}c^*_{i'j}$, from which we can prove that ρ_A is a positive semidefinite trace-one operator.

Exercise 1.12. Prove that for any bipartite state $|\psi_{AB}\rangle$, the density operator

$$\rho_A = \mathrm{Tr}_B(|\psi_{AB}\rangle\langle\psi_{AB}|) \tag{1.116}$$

is Hermitian, trace-one, and for arbitrary state $|\phi_A\rangle \in \mathcal{H}_A$ of system A, we have $\langle\phi_A|\rho_A|\phi_A\rangle \geq 0$.

In summary, we have the following definition

Definition 1.7 (reduced state). Since partial trace operation is linear, it can be extended to arbitrary density operator ρ_{AB} of composed system AB, for which $\rho_A = \mathrm{Tr}_B(\rho_{AB})$ and $\rho_B = \mathrm{Tr}_A(\rho_{AB})$ are called *reduced states*.

Notice that the definition of partial trace Tr_B and reduced state $\rho_A = \mathrm{Tr}\,\rho_{AB}$ is a direct result of requirement that disregarding subsystem B should have no influence on the outcomes of any measurement performed on A alone. If the pure state $|\psi_{AB}\rangle$ is not correlated, i.e., $|\psi_{AB}\rangle = |\phi_A\rangle \otimes |\chi_B\rangle$, then the reduced state of A and B must also be pure states, otherwise, they are mixed states.

1.3.2 Purification of mixed states

In the above discussion, we first have a composed system AB, then by taking the reduction of the pure states of AB, density operators of A and B are obtained respectively. We can also go in the other direction. If a state ρ_A of system A is given, if there exist a system B and a pure state $|\psi_{AB}\rangle$ of composed system AB such that ρ_A is the reduced state of $|\psi_{AB}\rangle$. The answer is yes, there always exist such a state, and it's named are the purification of ρ_A.

Theorem 1.3. *For a given density operator $\rho \in \mathbf{D}(\mathcal{H}_A)$, there exist a purification of ρ in space $\mathcal{H}_A \otimes \mathcal{H}_B$ if and only if $\dim \mathcal{H}_B \geq \mathrm{rank}(\rho)$.*

Proof. If ρ has a purification $|\psi_{AB}\rangle$, then the reduced states $\mathrm{rank}(\rho) = \mathrm{rank}(\rho_A) = \mathrm{rank}(\rho_B) \leq \dim \mathcal{H}_B$. This will be more clear using Schmidt decomposition, which we will discussed later.

If $\dim \mathcal{H}_B \geq \mathrm{rank}(\rho) = d$. Suppose that ρ has the following spectral decomposition

$$\rho = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i|. \tag{1.117}$$

Then choosing d orthonormal states ϕ_i in $\mathcal{H}_B$ (this can be done only if $\dim \mathcal{H}_B \geq \mathrm{rank}(\rho) = d$), we can check that

$$|\psi_{AB}\rangle = \sum_{i=1}^{d} \sqrt{p_i} |\psi_i\rangle \otimes |\phi_i\rangle \tag{1.118}$$

is the purification of ρ. □

The above theorem give us a sufficient and necessary condition for the existence of the purification. For a given state ρ, the purification is not unique. In fact, there are infinite purifications. Let's now consider the relationship between different purifications.

Proposition 1.1. *For a given density operator $\rho \in \mathbf{D}(\mathcal{H}_A)$, if $|\Psi_{AB}\rangle$ and $|\Phi_{AB}\rangle$ are two purifications of ρ over the space $\mathcal{H}_A \otimes \mathcal{H}_B$, then there exists a unitrary operator U such that $|\Phi_{AB}\rangle = (I \otimes U)|\Psi_{AB}\rangle$.*

Proof. Suppse that the rank of ρ is $d \leq d_A$ and the eigenvalues (in decreasing order) and eigenstates are p_i and $|\psi_i\rangle$, $i = 1, \cdots, d$; and $p_i = 0$ for $d < i \leq d_A$, the corresponding orthonormal eigenstates can be chosen as $|\phi_j\rangle$, then

$$\rho = \sum_{i=1}^{d_A} p_i |\psi_i\rangle\langle\psi_i| = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i|. \tag{1.119}$$

Choose an orthonormal basis for $\mathcal{H}_B$ as $|u_j\rangle, j = 1, \cdots, d_B$, expanding the purification $|\Psi_{AB}\rangle$ as

$$|\Psi_{AB}\rangle = \sum_{i=1}^{d_A} \sum_{j=1}^{d_B} c_{ij} |\phi_i\rangle \otimes |u_j\rangle. \tag{1.120}$$

Setting $|\mu_j\rangle = \sum_{j=1}^{d_B} c_{ij}|u_j\rangle$, we see that

$$|\Psi_{AB}\rangle = \sum_{i=1}^{d_A} |\phi_i\rangle \otimes |\mu_i\rangle. \tag{1.121}$$

By taking partial trace over B, we see that $\rho_A = \sum_{i=1}^{d_A} \sum_{j=1}^{d_B} |\phi_i\rangle\langle\phi_j|\langle\mu_j|\mu_i\rangle = \sum_{i=1}^{d} |\phi_i\rangle\langle\phi_i|$, this implies that $\langle\mu_j|\mu_i\rangle = \delta_{ji}p_i$ for $i,j \leq d$ and $|\mu_j\rangle = \vec{0}$ for $j > d$. By renormalizing $|\mu_i\rangle$ as $|x_i\rangle = |\mu_i\rangle/\sqrt{p_i}$ for $i \leq q$ and expanding them into an orthonormal basis of $\mathcal{H}_B$, we obtain (this is in fact the Schmidt decomposition of $|\Psi_{AB}\rangle$ which will be discussed later in this chapter)

$$|\Psi_{AB}\rangle = \sum_{i=1}^{d} \sqrt{p_i}|\phi_i\rangle \otimes |x_i\rangle. \tag{1.122}$$

Similarly, for $|\Phi_{AB}\rangle$, we can find an orthonormal basis $|y_j\rangle$ for $\mathcal{H}_B$ such that

$$|\Phi_{AB}\rangle = \sum_{i=1}^{d} \sqrt{p_i}|\phi_i\rangle \otimes |y_i\rangle. \tag{1.123}$$

We can construct the unitary operator U correspinding to the basis transformation from $|x_i\rangle$ to $|y_i\rangle$, which satisfy our requirement. □

Remark 1.1. Notice that here two purifications are require to be in the same space, but for purifications in diffrent spaces $\mathcal{H}_A \otimes \mathcal{H}_B$ and $\mathcal{H}_A \otimes \mathcal{H}_{B'}$, similar result holds. A slightly tricky case is when $\dim \mathcal{H}_B \geq \dim \mathcal{H}_{B'}$, in this situation, there does not exist any unitary operator. Nevertheless, we can construct the unitary operator in a subspace which the purifications lie in.

Exercise 1.13 (Some tricks based on maximally entangled states). In this exercise, we explore some interesting and useful properties of a special state, which is known as Greenberger–Horne–Zeilinger (GHZ) state:

$$|GHZ\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle \otimes |i\rangle, \tag{1.124}$$

$$|\Omega\rangle = \sum_{i=0}^{d-1} |i\rangle \otimes |i\rangle. \tag{1.125}$$

They live in the space $\mathcal{X} \otimes \mathcal{Y}$ with $\dim \mathcal{Y} \geq \dim \mathcal{X}$, $|\Omega\rangle$ and $|GHZ\rangle$ differ by an overall factor, both forms are of great importance.

Show that the following results hold:

1. If $|\psi\rangle$ is the purification of $\rho \in \mathbf{D}(\mathcal{X})$, there exist some unitary operators U and V such that
$$|\psi\rangle = (\sqrt{\rho} U \otimes V)|\Omega\rangle. \tag{1.126}$$
2. For operator $\rho, \sigma \in \mathbf{B}(\mathcal{X})$, the Hilbert-Schmidt inner product satisfies
$$\langle \Omega|\rho \otimes \sigma|\Omega\rangle = \mathrm{Tr}(\rho^T \sigma) = \langle \rho^*, \sigma\rangle = \mathrm{Tr}(\sigma^T \rho) = \langle \sigma^*, \rho\rangle. \tag{1.127}$$

Exercise 1.14 (Quantum marginal problem). For two quantum systems

1.3.3 Schrödinger-GHJW theorem

We now know that a density operator has two interpretations: an ensemble of pure states and the state of an open quantum system.

§ 1.4 Distance between quantum states

There are several different ways to quantify the similarity and difference between density operators. For pure states, the square root of transition probability (which is called fidelity)
$$F(\psi, \varphi) = |\langle \psi|\varphi\rangle| \tag{1.128}$$
is a satisfying quantification of similarity. Another mathematically natural way to measure difference between two states is using various operator norms.

1.4.1 Operator norm distance

Before we start, let's recall some basic definition about operator norms. For a given vector space $\mathcal{X}$, the norm is a real valued function $\|\cdot\| : \mathcal{X} \to \mathbb{R}$ which satisfies the following three conditions:

1. Positive definiteness: $\|x\| \geq 0$ for all $x \in \mathcal{X}$ with $\|x\| = 0$ if and only if $x = 0$.
2. Absolutely homogeneous: $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{C}$ and $x \in \mathcal{X}$.
3. The triangle inequality (sub-additivity): $\|x + y\| \leq \|x\| + \|y\|$.

A vector space equipped with a norm is called a normed vector space, a complete normed vector space is called a Banach space. We will mainly concern the norm on the space of bounded operators, $\mathbf{B}(\mathcal{X}, \mathcal{Y})$ between two normed vector spaces $\mathcal{X}$ and $\mathcal{Y}$. This kind of norm will be called operator norm (or matrix norm).

From a given norm function, we can define a distance function

$$d(x, y) = \|x - y\| \tag{1.129}$$

which can be used to measure difference between two states. It's easy to verify from the definition of norm that the norm induced distance function satisfies the axioms of a distance (or metric):

1. Positive definiteness: $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;
2. Symmetric: $d(x, y) = d(y, x)$;
3. Triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$.

A crucial family of norms we will use later is the so-called *Schatten norm*, which is a generalization of the p-norm of vectors

$$\|\vec{x}\|_p := \left(\sum_i |x_i|^p\right)^{1/p}. \tag{1.130}$$

The notation $\|\cdot\|$ is usually preserved to denote the 2-norm for vectors (and also for operators later).

Definition 1.8 (Schatten norm). For a linear transformation $A : \mathcal{X} \to \mathcal{Y}$, the Schatten p-norm for any $p \geq 1$ is defined as

$$\|A\|_p := [\mathrm{Tr}((A^\dagger A)^{p/2})]^{1/p} \tag{1.131}$$

There are several important specials cases:

1. Trace norm $\|A\|_1 =$
2.

1.4.2 Quantum fidelity

Quantum fidelity is an extensively used quantity to quantify the similarity between quantum states in quantum information theory. For two density operators $\rho, \sigma \in \mathbf{D}(\mathcal{H})$, their fidelity is defined as

$$F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \mathrm{Tr}(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}). \tag{1.132}$$

It's easy to verify that for pure states

$$F(|\psi\rangle, |\varphi\rangle) = |\langle\psi|\varphi\rangle|. \tag{1.133}$$

For states $|\psi\rangle$ and ρ, we have

$$F(\psi, \rho) = \mathrm{Tr}(\sqrt{|\psi\rangle\langle\psi|\rho|\psi\rangle\langle\psi|}) = \sqrt{\langle\psi|\rho|\psi\rangle}. \tag{1.134}$$

If ρ and σ commute, there exists common eigenstates $|i\rangle$'s such that

$$\rho = \sum_i p_i|i\rangle\langle i|, \quad \sigma = \sum_i q_i|i\rangle\langle i|. \tag{1.135}$$

In this case

$$F(\rho, \sigma) = \sum_i \sqrt{p_i q_i} \tag{1.136}$$

This coincides with the fidelity between two probability vectors $\vec{p}$ and $\vec{q}$.

Properties of fidelity

Fidelity satisfies many fine properties, notice that the definition of fidelity function can be generalized to the set of all positive semidefinite operators $\mathbf{Pos}(\mathcal{X})$. This generalization of convenient for the later discussion.

Symmetric.—The fidelity is a symmetric function

$$F(\rho, \sigma) = F(\sigma, \rho), \quad \forall \rho, \sigma \in \mathbf{Pos}(\mathcal{X}). \tag{1.137}$$

This is because that $A^\dagger A$ and $AA^\dagger$ have the same eigenvalues (see Exercise 1.8). When taking $A = \sqrt{\rho}\sqrt{\sigma}$, from definition of fidelity $F(\rho, \sigma) = \|A\|_1$, $F(\sigma, \rho) = \|A^\dagger\|_1$, they are of the same value.

Scaling.—The definition of fidelity function can be generalized to the set of all positive semidefinite operators. For this fidelity function, we have

$$F(\lambda\rho, \sigma) = \sqrt{\lambda}F(\sigma, \rho) = F(\rho, \lambda\sigma), \quad \forall \lambda \geq 0, \forall \rho, \sigma \in \mathbf{Pos}(\mathcal{X}). \tag{1.138}$$

Invariant under unitary transformations.—The fidelity is invariant under simultaneous unitary transformations for both states,

$$F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho,\sigma), \quad \forall \rho,\sigma \in \mathbf{Pos}(\mathcal{X}). \tag{1.139}$$

To prove this, recall that for positive semidefinite operators $\sqrt{U\rho U^\dagger} = U\sqrt{\rho}U^\dagger$. Thus

$$F(U\rho U^\dagger, U\sigma U^\dagger) = \|U\sqrt{\rho}\sqrt{\sigma}U^\dagger\|_1 = \|\sqrt{\rho}\sqrt{\sigma}\|_1 = F(\rho,\sigma). \tag{1.140}$$

Joint multiplicative under tensor product.—Consider density operators ρ_1,ρ_2 and σ_1,σ_2, we have

$$F(\rho_1\otimes\rho_2, \sigma_1\otimes\sigma_2) = F(\rho_1,\sigma_1)F(\rho_2,\sigma_2). \tag{1.141}$$

This is because that

$$\begin{aligned} F(\rho_1\otimes\rho_2, \sigma_1\otimes\sigma_2) =&\|\sqrt{\rho_1\otimes\rho_2}\sqrt{\sigma_1\otimes\sigma_2}\|_1 \\ =&\|\sqrt{\rho_1}\sqrt{\sigma_1}\otimes\sqrt{\rho_2}\sqrt{\sigma_2}\|_1 \\ =&\|\sqrt{\rho_1}\sqrt{\sigma_1}\|_1\|\sqrt{\rho_2}\sqrt{\sigma_2}\|_1 \\ =&F(\rho_1,\sigma_1)F(\rho_2,\sigma_2). \end{aligned} \tag{1.142}$$

Uhlmann's theorem for fidelity.—A crucial property of fidelity is that fidelity of two density operators can be characterized by their purifications, this is found by Uhlmann and now called Uhlmann's theorem.

Theorem 1.4 (Uhlmann's theorem). *Let $\rho,\sigma\in\mathbf{D}(\mathcal{X})$ be two density operators, then*

$$F(\rho,\sigma) = \max_{\psi,\varphi}\{|\langle\psi|\varphi\rangle| = F(\psi,\varphi)|\, \mathrm{Tr}_{\mathcal{Y}}|\psi\rangle\langle\psi| = \rho, \mathrm{Tr}_{\mathcal{Y}}|\varphi\rangle\langle\varphi| = \sigma\}, \tag{1.143}$$

where the maximum is taken over all purifications $\psi,\varphi\in\mathcal{X}\otimes\mathcal{Y}$.

Proof. Let $|\Omega\rangle = \sum_{i=1}^{d_{\mathcal{X}}}|i\rangle\otimes|i\rangle\in\mathcal{Y}\otimes\mathcal{X}$, it's easy to verify that (see Exercise 1.13), any purifications of ρ,σ can be expressed as

$$|\psi\rangle = U_{\mathcal{Y}}\otimes\sqrt{\rho}U_{\mathcal{X}}|\Omega\rangle, \tag{1.144}$$

$$|\varphi\rangle = V_{\mathcal{Y}}\otimes\sqrt{\sigma}V_{\mathcal{X}}|\Omega\rangle. \tag{1.145}$$

by choosing appropriate unitary operators. Taking the inner product gives

$$\begin{aligned}|\langle\psi|\varphi\rangle| =& |\langle\Omega|U_{\mathcal{Y}}^{\dagger}V_{\mathcal{Y}}\otimes U_{\mathcal{X}}^{\dagger}\sqrt{\rho}\sqrt{\sigma}V_{\mathcal{X}}|\Omega\rangle| \\ =& \operatorname{Tr}((U_{\mathcal{Y}}^{\dagger}V_{\mathcal{Y}})^{T}U_{\mathcal{X}}^{\dagger}\sqrt{\rho}\sqrt{\sigma}V_{\mathcal{X}}) \\ =& \operatorname{Tr}(U\sqrt{\rho}\sqrt{\sigma})\end{aligned} \tag{1.146}$$

where we have used expression (1.127) and set $U = V_{\mathcal{X}}(U_{\mathcal{Y}}^{\dagger}V_{\mathcal{Y}})^{T}U_{\mathcal{X}}^{\dagger}$. Then, from the result of Exercise 1.9, we have

$$|\langle\psi|\varphi\rangle| = \operatorname{Tr}(U\sqrt{\rho}\sqrt{\sigma}) \le \operatorname{Tr}|\sqrt{\rho}\sqrt{\sigma}|_{L} = F(\rho,\sigma). \tag{1.147}$$

The equality is reached by choosing appropriate unitary operators such that U satisfies

$$\sqrt{\rho}\sqrt{\sigma} = U^{\dagger}|\sqrt{\rho}\sqrt{\sigma}|_{L}, \tag{1.148}$$

i.e., $U^{\dagger}$ is a unitary operator in polar decomposition of $\sqrt{\rho}\sqrt{\sigma}$. This completes the proof. □

Exercise 1.15. Show that the fidelity can also be expressed as

$$F(\rho,\sigma) = \max_{\varphi}\{|\langle\psi|\varphi\rangle| \,|\, \operatorname{Tr}_{\mathcal{Y}}|\psi\rangle\langle\psi| = \rho, \operatorname{Tr}_{\mathcal{Y}}|\varphi\rangle\langle\varphi| = \sigma\} \tag{1.149}$$

with ψ a fixed purification of ρ. □

Exercise 1.16. The Uhlmann's theorem works for all positive semidefinite operators. The main difference here is that the norm of purification ψ of a positive semidefinite operator ρ is not 1 anymore. In fact $\|\psi\| = \operatorname{Tr}(\rho)$. Try to give the explicit formulation of the Uhlmann's theorem and its proof in this situation. □

Fidelity is bounded.—From Uhlmann's theorem, it's clear that the fidelity function is bounded

$$0 \le F(\rho,\sigma) \le 1, \quad \forall \rho,\sigma \in \mathbf{D}(\mathcal{X}). \tag{1.150}$$

Concavity.—There are several different kinds of concavity of fidelity function that will be useful in application.

Theorem 1.5 (Strong joint concavity). *Let ρ_i and σ_i are two collection of positive semidefinite operators in $\mathbf{Pos}(\mathcal{X})$, then*

$$F(\sum_{i=1}^{k}\rho_i, \sum_{i=1}^{k}\sigma_i) \ge \sum_{i=1}^{k} F(\rho_i,\sigma_i). \tag{1.151}$$

Proof. Choose purifications ψ_i and φ_i of ρ_i and σ_i in space $\mathcal{X} \otimes \mathcal{Y}$ such that $F(\rho_i, \sigma_i) = \langle \psi_i, \varphi_i \rangle$ for all $i = 1, \cdots, k$, their existence is guaranteed by Uhlamnn's theorem (notice that by tuning the overall factor, the absolute value symbol can be dropped).

We can introduce a new space $\mathcal{Z}$ with dimension k and orthonormal basis e_i. Define two vectors

$$|\Psi\rangle = \sum_{i=1}^{k} |\psi_i\rangle \otimes |e_i\rangle, \quad |\Phi\rangle = \sum_{i=1}^{k} |\varphi_i\rangle \otimes |e_i\rangle. \tag{1.152}$$

We see that $|\langle \Psi | \Phi \rangle| = \sum_i F(\rho_i, \sigma_i)$. And it's clear that

$$\mathrm{Tr}_{\mathcal{Y}\otimes\mathcal{Z}} |\Psi\rangle\langle\Psi| = \sum_{i=1}^{k} \rho_i, \quad \mathrm{Tr}_{\mathcal{Y}\otimes\mathcal{Z}} |\Phi\rangle\langle\Phi| = \sum_{i=1}^{k} \sigma_i. \tag{1.153}$$

From Uhlmann's theorem, we obtain

$$F(\sum_{i=1}^{k} \rho_i, \sum_{i=1}^{k} \sigma_i) \geq |\langle \Psi | \Phi \rangle| = \sum_{i=1}^{k} F(\rho_i, \sigma_i). \tag{1.154}$$

This completes the proof. □

From the above theorem and scaling property of fidelity function, we directly have the following result.

Corollary 1.1. *Let p_i and q_i are two probability vectors and ρ_i and σ_i are two collection of density operators in $\mathbf{D}(\mathcal{X})$, then*

$$F(\sum_{i=1}^{k} p_i\rho_i, \sum_{i=1}^{k} q_i\sigma_i) \geq \sum_{i=1}^{k} \sqrt{p_i q_i} F(\rho_i, \sigma_i). \tag{1.155}$$

This result further implies that

$$F(\sum_{i=1}^{k} p_i\rho_i, \sum_{i=1}^{k} p_i\sigma_i) \geq \sum_{i=1}^{k} p_i F(\rho_i, \sigma_i). \tag{1.156}$$

If we set all $\sigma_i = \sigma$, we see that

$$F(\sum_{i=1}^{k} p_i\rho_i, \sigma) \geq \sum_{i=1}^{k} p_i F(\rho_i, \sigma). \tag{1.157}$$

The fidelity function is concave in the first entry, from the symmetry of fidelity function, it's also concave in the second entry.

Alberti's theorem.—There are several different kinds of concavity of fidelity function that will be useful in application.

§ 1.5 Entanglement I: pure state case

Consider a bipartite quantum system $\mathcal{H}_{AB}$, a pure quantum state $|\Psi\rangle_{AB}$ is called a product state[3] if there exist two quantum states $|\psi\rangle_A$ and $|\varphi\rangle_B$ for A, B respectively such that $|\Psi\rangle_{AB} = |\psi\rangle_A \otimes |\varphi\rangle_B$, otherwise the state is called *entangled*.

Typical examples of entangled states are four *Bell states* (also known as *EPR pairs*):

$$|\phi^+\rangle = \tfrac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \tag{1.158}$$

$$|\phi^-\rangle = \tfrac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \tag{1.159}$$

$$|\psi^+\rangle = \tfrac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \tag{1.160}$$

$$|\psi^-\rangle = \tfrac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \tag{1.161}$$

The *entanglement* is a core concept and is ubiquitous in quantum information theory. We will discuss it in different depth in these lecture notes, here we only comments that entanglement cannot be created with local operations $U_A \otimes U_B$ and classical communications. To create a product state, $|\psi\rangle_A \otimes |\varphi\rangle_B$, a referee can send Alice and Bob messages about what state they should prepare. However, to create an entangled state, some nonlocal joint operations between Alice and Bob must be made.

In this section, we discuss how to characterized pure state entanglement. In the next section, the mixed state case will be discussed.

1.5.1 Schmidt decomposition

It's useful to write a pure entangled state in a standard form, known as *Schmidt decomposition*. The Schmidt decomposition is in fact the singular value decomposition of matrix given by the coefficients of the bipartite state in the given basis.

[3] A product state is a special case of the more general notion of separable states, entangled state is defined as the non-separable state generally, this will be discussed later.

Theorem 1.6 (Schmidt decomposition). *For every bipartite pure state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ with $d := \min\{\dim \mathcal{H}_A, \dim \mathcal{H}_B\}$, there exist orthonormal bases $\{|u_j\rangle \in \mathcal{H}_A\}$ and $\{|v_j\rangle \in \mathcal{H}_B\}$ such that*

$$|\psi\rangle_{AB} = \sum_{j=1}^{d} \sqrt{p_j}|u_j\rangle \otimes |v_j\rangle, \tag{1.162}$$

with $p_j \geq 0$ and $\sum_{j=1}^{d} p_j = 1$. The coefficients $\{\lambda_j := \sqrt{p_i}\}$ are called Schmidt coefficients and the number of nonzero λ_j is called the Schmidt rank of $|\psi\rangle_{AB}$.

Proof. Suppose that in the given bases $\{|i\rangle_A \in \mathcal{H}_A\rangle$ and $\{|j\rangle_A \in \mathcal{H}_A\rangle$, the state is of the form

$$|\psi\rangle_{AB} = \sum_{i,j} c_{ij}|i\rangle_A \otimes |j\rangle_B, \tag{1.163}$$

for the $d_A \times d_B$ matrix $C = (c_{ij})$, using the singular value decomposition, $\Lambda = U^\dagger C V^\dagger$, where U, V are unitary operators and $\Lambda = (\lambda_i \delta_{ij})$ is diagonal. Define

$$|u_k\rangle = \sum_{i=1}^{d_A}(U^*)_{ki}|i\rangle_A, \quad |v_l\rangle = \sum_{j=1}^{d_B}(V^\dagger)_{lj}|j\rangle_B, \tag{1.164}$$

they are two orthonormal bases of system A and B respectively, since $U^*, V^\dagger$ are unitary operators. We thus have that $|i\rangle_A = \sum_{k=1}^{d_A}(U^T)_{ik}|u_k\rangle$ and $|j\rangle_B = \sum_{l=1}^{d_B} V_{jl}|v_l\rangle$, substituting them into the expression (1.163) of $|\psi\rangle$, we arrive at

$$\begin{aligned}
|\psi\rangle &= \sum_{i,j} c_{ij}\left(\sum_{k=1}^{d_A}(U^T)_{ik}|u_k\rangle\right) \otimes \left(\sum_{l=1}^{d_B} V_{jl}|v_l\rangle\right) \\
&= \sum_{k=1}^{d_A}\sum_{l=1}^{d_B}\left(\sum_{i,j} U_{ki}c_{ij}V_{jl}\right)|u_k\rangle \otimes |v_l\rangle \\
&= \sum_{k=1}^{d_A}\sum_{l=1}^{d_B} \lambda_k \delta_{kl}|u_k\rangle \otimes |v_l\rangle \\
&= \sum_{k=1}^{d} \lambda_k |u_k\rangle \otimes |v_k\rangle.
\end{aligned} \tag{1.165}$$

Since the norm of $|\psi\rangle$ is one, the obtain $\sum_k \lambda_k^2 = 1$. $\square$

Schmidt decomposition provides us a very convenient sufficient and necessary criterion for pure state entanglement: we say a pure state is entangled if and only if the Schmidt rank of the state is equal or greater than two,

Exercise 1.17 (Entanglement spectrum).

For a state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$, if its Schmidt coefficients are $\lambda_j = 1/\sqrt{d}$ for all $j = 1, \cdots, d$, then it's called a *maximally entangled state*. The name is justified by the fact that every other states of the same dimension can be obtained with unit probability from a maximally entangled state by means of *local operations and classical communications* (LOCC). This will be illustrated later in this book. Bell states are 2-dimensional expamples. Another typical example is the Greenberger–Horne–Zeilinger (GHZ) state

$$|GHZ\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |j\rangle_A \otimes |j\rangle_B. \tag{1.166}$$

We will see that the maximally entangled sates is crucial in quantum information theory, there are many tricks, like channel-state duality, entanglement distillation and concentration, etc., based on maximally entangled states.

1.5.2 Superdense coding

Let's now consider an interesting application of quantum entanglement called *superdense coding* or *dense coding*, where by using pre-shared entangled quantum states, Alice can send two classical bits to Bob by sending just one qubit. It can be thought of as the opposite of *quantum teleportation* (which we will discuss in the next section), in which one transfers one qubit from Alice to Bob by communicating two classical bits, with Alice and Bob having a pre-shared Bell pair.

The superdense coding is a kind of secure quantum communication. If an eavesdropper intercept the Alice's transmitted qubit in the route to Bob, the state he obtain is just $\rho_A = I_A/2$ which carries no information at all. All the information is encoded in the correlations between particles A and B, this information is inaccessible unless the eavesdropper is able to obtain both particles of the entangled pair.

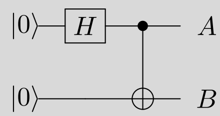
The protocol works in four steps: entangled-state preparation and sharing, encoding, qubit sending, and decoding.

Entangled-state preparation and sharing

Suppose that Charlie prepares the Bell state

$$|\phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B) \tag{1.167}$$

and she sends the two particles to Alice and Bob respectively. The preparation circuit is like



(1.168)

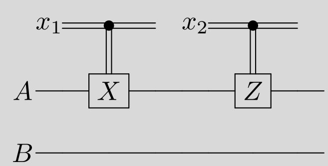
Or we can suppose that Alice prepare the state $|\phi^+\rangle_{AB}$ and send the second half to Bob. The preparation process is completed long before Alice try to communicate with Bob. This kind of viewpoint can help us to understand that the superdense coding is not contrary with Holevo's theorem as will be remarked later.

Encoding

Now Alice and Bob share the Bell pair $|\phi^+\rangle_{AB}$. Alice encodes two classical information as

- $x_1x_2 = 00$ as do nothing on her state, i.e. operates I_A, the resulting states is $|\phi^+\rangle_{AB}$;
- $x_1x_2 = 01$ as bit-flip, i.e., operates σ_x^A, the resulting state is $|\psi^+\rangle_{AB}$;
- $x_1x_2 = 10$ as phase-flip, i.e., operates σ_z^A, the resulting state is $|\phi^-\rangle_{AB}$;
- $x_1x_2 = 11$ as both bit-flip and phase-flip, i.e., operates $\sigma_z^A\sigma_x^A$, the resulting state is $|\psi^-\rangle_{AB}$.

The circuit of encoding process is like



(1.169)

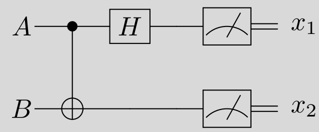
Qubit sending

After encoding, Alice sends her half of qubit to Bob, there is only one-qubit communication.

Decoding

When Bob receives the qubit, he performs measurements in four Bell state basis, the measurement outcome unambiguously distinguishes the four possible actions that Alice could have performed. Thus Bob obtain two classical bit of information.

Another way Bob can decode two classical bits of information works as follows.



(1.170)

The superdense coding seems to be contrary to Holevo's theorem (the details of the theorem will be discussed later in this book) at first glimpse. A special case of Holevo's theorem states that, if Alice sends one qubit at a time, no matter how she prepares qubit state and no matter how Bob measures it, no more than one classical bit can be carried by each qubit. In superdense coding protocol, we see that Alice send one qubit but transmit two classical bits, it seems that there is a contradict. The reason behind this is that, Alice really need to transmit two qubit to complete the protocol, the first one qubit transmitted in the preparation and sharing state. Thus a two qubit state contains at most two classical bit of information, there is no contradiction.

1.5.3 Quantum teleportation

§ 1.6 Entanglement II: mixed state case

1.6.1 Positive partial transpose criterion

1.6.2 Entanglement purification

1.6.3 Entanglement concurrence

Exercise 1.18. Let $\tilde{\rho} = (\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)$, show that operators $\sqrt{\rho}\tilde{\rho}\sqrt{\rho}$ and $\rho\tilde{\rho}$ have the same spectrum, viz., they have the same eigenvalues.

1.6.4 Examples of entangled states

Let us now see some crucial examples of quantum states which is entangled.

Bell states

GHZ states

W states

Bell diagonal states

Werner states

In 1989 R. Werner introduced a class of states when he studied the mixed state entanglement, now these states are known as *Werner states*.

Definition 1.9. For a bipartite quantum system $\mathcal{H}_{AB} = \mathbb{C}^d \otimes \mathbb{C}^d$, a state $\rho \in B(\mathbb{C}^d \otimes \mathbb{C}^d)$ is called a Werner state if for any unitary transformation $U \in U(\mathbb{C}^d)$, ρ is invariant under $U \otimes U$, namely

$$(U \otimes U)\rho(U \otimes U)^\dagger = \rho. \tag{1.171}$$

It turns out that the state is of the form

$$\rho_\alpha = (1-\alpha)\frac{I_{d^2}}{d^2} + \alpha\frac{2P_{as}}{d(d-1)}, \tag{1.172}$$

where P_{as} denotes the projector on antisymmetric subspace.

Before discussing the entanglement properties of the states, let's give a quick proof of the explicit form of the Werner states. Recall that there is a direct sum decomposition[4] of bipartite system $\mathbb{C}^d \otimes \mathbb{C}^d$,

$$\mathbb{C}^d \otimes \mathbb{C}^d = \mathrm{Sym}^2(\mathbb{C}^d) \oplus \mathrm{Alt}^2(\mathbb{C}^d), \tag{1.173}$$

where $\mathrm{Sym}(\mathbb{C}^d)$ and $\mathrm{Alt}(\mathbb{C}^d)$ is the symmetric and antisymmetric subspace of dimension $d(d+1)/2$ and $d(d-1)/2$. For a given basis $|i\rangle$, $i = 0, \cdots, d-1$, the bases for $\mathrm{Sym}^2(\mathbb{C}^d)$ are

$$\begin{array}{cccccc}
|00\rangle, & \frac{|01\rangle+|10\rangle}{\sqrt{2}}, & \frac{|02\rangle+|20\rangle}{\sqrt{2}}, & \cdots, & \frac{|0(d-1)\rangle+|(d-1)1\rangle}{\sqrt{2}}, \\
 & |11\rangle & \frac{|12\rangle+|21\rangle}{\sqrt{2}}, & \cdots, & \frac{|1(d-1)\rangle+|(d-1)1\rangle}{\sqrt{2}}, \\
 & & |22\rangle, & \cdots, & \frac{|2(d-1)\rangle+|(d-1)2\rangle}{\sqrt{2}}, \\
 & & & \ddots & \vdots \\
 & & & & |(d-1)(d-1)\rangle,
\end{array} \tag{1.174}$$

and the bases for $\mathrm{Alt}(\mathbb{C}^d)$ are

[4] This is a special property for bipartite system $\mathcal{H} \otimes \mathcal{H}$, for $\mathcal{H}^{\otimes n}$ $(n \geq 3)$, this is no similar result.

$$\begin{array}{cccc} \frac{|01\rangle-|10\rangle}{\sqrt{2}}, & \frac{|02\rangle-|20\rangle}{\sqrt{2}}, \cdots, & & \frac{|0(d-1)\rangle-|(d-1)1\rangle}{\sqrt{2}}, \\ & \frac{|12\rangle-|21\rangle}{\sqrt{2}}, \cdots, & & \frac{|1(d-1)\rangle-|(d-1)1\rangle}{\sqrt{2}}, \\ & & \ddots & \vdots \\ & & & \frac{|(d-2)(d-1)\rangle-|(d-1)(d-2)\rangle}{\sqrt{2}}. \end{array} \tag{1.175}$$

The corresponding projectors denote P_s and P_{as}.

Notice that the projectors to symmetric and antisymmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$ is of the form

$$P_{as} = \frac{1}{\sqrt{2}}(I - V_{AB}), \tag{1.176}$$

$$P_s = \frac{1}{\sqrt{2}}(I + V_{AB}), \tag{1.177}$$

where $V_{AB} = \sum_{ij} |ij\rangle\langle ji|$ is the swap operator for which $V_{AB}|\varphi_A\rangle|\psi\rangle = |\psi\rangle|\varphi\rangle$.

Exercise 1.19. Prove that the projectors onto symmetric and antisymmetric subspaces of $\mathbb{C}^d \otimes \mathbb{C}^d$ are of form in expressions (1.176) and (1.177).

Consider an operator $A \in B(\mathbb{C}^d \otimes \mathbb{C}^d)$, which is invariant under the action $U \otimes U$ for all $U \in U(\mathbb{C}^d)$, or euivalently $[A, U \otimes U] = 0$, for a basis $|i\rangle$, $i = 0, \cdots, d-1$ of $\mathbb{C}^d$, the matrix element of A is then

$$A_{ij,kl} = \langle ij|A|kl\rangle. \tag{1.178}$$

Consider the unitary transformations U_r ($r = 0, \cdots, d-1$) which maps $|r\rangle \to -|r\rangle$ but leaves all other basis elements unchanged, $A(U_r \otimes U_r) = (U_r \otimes U_r)A$ implies that matrix elements of $A_{ij,kl} \neq 0$ only when (i) $i = j = k = l$, or (ii) $i = k \neq j = l$, or (iii) $i = l \neq j = k$, or (iv) $i = j \neq k = l$. Since the permutation of basis is also unitary, acting permutation U_σ ($\sigma \in S_d$) implies that $A_{\sigma(i)\sigma(j),\sigma(k)\sigma(l)} = A_{ij,kl}$,

Isotropic states

Graph state.—

§ 1.7 Entanglement III: Bell inequality

1.7.1 Local hidden variable model

1.7.2 Bell nonlocality

1.7.3 Quantum steering

1.7.4 Quantum discord

1.7.5 Hierarchy

§ 1.8 Multipartite quantum state

1.8.1 Graph state

§ 1.9 Reading materials

§ 1.10 Problems

Problem 1.1. dd

Chapter 2
Measurement as positive operator-valued measure

Beautiful mathematics eventually tends to be useful, and useful mathematics eventually tends to be beautiful.

From Meyer, Carl (2000) Matrix analysis and applied linear algebra

In the last chapter we discussed the states of a quantum open system, and demonstrated that they are mathematically described by density operators, which are trace-one positive semidefinite operators. In this and the next chapters, we will develop the theory of measurement and time evolutions from the quantum open system perspective. As we will see, the quantum measurements are characterized by *generalized measurements*, which are mathematically described by positive operator-valued measure (POVM); the time evolutions are characterized by *quantum channels*, which are mathematically described by completely positive trace-preserving (CPTP) maps. Both of the generalized measurements and quantum channels can be regarded as *quantum operations*, which are completely positive (CP) maps.

Before we start to discuss the details of generalized measurements and quantum channels, let's first recall some mathematical concepts which play a crucial role in this and the next chapter. The states of system is described by the operators over a Hilbert space, the quantum operations transforms quantum states to quantum states, thus they are maps between the sets of operators over Hilbert spaces. Consider Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ for system A, B, the set of linear operators over them are denoted as $\mathcal{L}(\mathcal{H}_A)$ and $\mathcal{L}(\mathcal{H}_B)$ respectively, they are both vector spaces. A quantum transformation between system A and B is defined as a linear map $\mathcal{M} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$, since $\mathcal{M}$ maps operator into operator, it's also called a *superoperator*. The set of all superoperator is denoted as $\mathbf{T}(\mathcal{H}_A, \mathcal{H}_B) := \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$. Generalized measurements and quantum channels are all quantum transformation between quantum systems, thus are elements of $\mathbf{T}(\mathcal{H}_A, \mathcal{H}_B)$.

Consider a toy model for which a quantum process can be regarded as a composition of three basic ingredients: state preparation, state transformation and state measurement. These are all special cases of quantum operations. The state preparation is given by a transformation in $\mathbf{T}(\mathbb{C}, \mathcal{H}_A)$; the state transformation is described by elements in $\mathbf{T}(\mathcal{H}_A, \mathcal{H}_B)$, the state measurements is thus given by transformations in $\mathbf{T}(\mathcal{H}_B, \mathcal{H}_B)$.

§ 2.1 von Neumann's projective measurement

From the Copenhagen axiomatic formulation of quantum mechanics, we know that a quantum measurement may be described as an orthogonal projection operator. To measure an observable F with outcomes labeled as $a = 0, 1, \cdots, N-1$, we need to choose a corresponding apparatus for which we can read out the *pointer states* $|a\rangle_A$, these pointer states correspond to different outcomes of obsevable F and are correlated with some macroscopic classical variables. By tuning on the coupling between system and measurment apparatus, we will modify the Hamiltonian of our world such that there is an interaction of system and meausement apparatus. After a period of time evolution, the resulting state is

$$|\Psi\rangle_{SA} = U(|\psi\rangle_S \otimes |0\rangle_A) = \sum_{a=0}^{N-1} c_a |a\rangle_S \otimes |a\rangle_A \tag{2.1}$$

The probability of observing a of F upon state $|\psi\rangle$ is

$$p(a) = \|I \otimes (|a\rangle\langle a|)|\Psi\rangle_{SA}\|^2 = |c_a|^2. \tag{2.2}$$

This is a well-known result from textbook quantum mechanics.

Thinking more abstractly, for an observable F, suppose that $\{E_a, a = 0, 1, \cdots, N\}$ is a complete set of orthogonal projectors cooresponding the different outcomes of F, they satisfy

$$E_a E_b = \delta_{ab} E_a, E_a^\dagger = E_a, \sum_{a=0}^{N-1} E_a = I. \tag{2.3}$$

To measure F, we introduce an N-dimensional apparatus space with pointer states $|a\rangle_A$, $a \in \mathbb{Z}_N$. The coupling of system and apparatus is characterized by the unitary operator

$$U = \sum_{a,b=0}^{N-1} E_a \otimes |b + a\rangle\langle b|. \tag{2.4}$$

Exercise 2.1. Prove that the operator $U = \sum_{a,b=0}^{N-1} E_a \otimes |b+a\rangle\langle b|$ is unitary.

Suppose that the initial state of system and apparatus are $|\psi\rangle_S$ and $|0\rangle_A$ respectively, the resultant state after coupling is

$$|\Psi\rangle_{SA} = U|\psi\rangle_S \otimes |0\rangle_A = \sum_{a=0}^{N-1} E_a|\psi\rangle_S \otimes |a\rangle_A. \tag{2.5}$$

The outcome a occurs with probability

$$p(a) = \|I \otimes (|a\rangle\langle a|)_A|\Psi\rangle\|^2 = \langle\Psi|I \otimes (|a\rangle\langle a|)_A|\Psi\rangle = \langle\psi|E_a|\psi\rangle. \tag{2.6}$$

After we read out value a, the post-measurement state of system is

$$|\psi_a\rangle = \frac{E_a|\psi\rangle}{\|E_a|\psi\rangle\|}. \tag{2.7}$$

Alternatively, we can express it in density matrix form

$$p(a) = \mathrm{Tr}(E_a|\psi\rangle\langle\psi|E_a^\dagger), \quad |\psi_a\rangle = \frac{E_a|\psi\rangle\langle\psi|E_a^\dagger}{\mathrm{Tr}(E_a|\psi\rangle\langle\psi|E_a^\dagger)}. \tag{2.8}$$

If the measurement is performed but the outcome value is not read out, the output state is a mixed state

$$\sum_a p(a)|\psi_a\rangle\langle\psi_a| = \sum_a E_a|\psi\rangle\langle\psi|E_a^\dagger. \tag{2.9}$$

The above discussion is for pure state, if the initial state is a mixed state ρ, we can express it a an ensemble of pure states, then similar results will be obtained. After the measurement is perfomed and outcome a is read out, the post-mesurement state is

$$\rho_a = \frac{E_a\rho E_a^\dagger}{\mathrm{Tr}(E_a\rho E_a^\dagger)}. \tag{2.10}$$

If the measurement is performed but the outcomes are not read out, the output state is

$$\sum_a E_a\rho E_a^\dagger. \tag{2.11}$$

2.1.1 Gleason's theorem

§ 2.2 Positive operator-valued measure

2.2.1 Naimark's theorem

We now know that POVMs of a system $\mathcal{H}_S$ can arise when applying the projective measurements on a larger system $\mathcal{H}'$, it is natural to ask if all POVMs, i.e., an arbitrary set of positive operators which satisfy the completeness condition, can be realized in this way. The answer, as we will see, is *yes*, this is guaranteed by the Naimark's theorem [1].

Theorem 2.1 (Naimark's theorem).

2.2.2 Postive superoperators

§ 2.3 Quantum instrument

[1] The theorem is also named as Neumark's theorem by some authors, but the two names both refer to the same Soviet mathematician, Mark Aronovich Naimark, whose name has been translated in these two ways.

Chapter 3
Time evolution as quantum channels

§ 3.1 Unitary evolution of closed quantum system

For a closed quantum system, the time evolution is controlled by Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle. \tag{3.1}$$

§ 3.2 Quantum operations and CPTP map

3.2.1 Completely positive maps

3.2.2 Trace-preserving maps

§ 3.3 Quantum channels

We have seen that quantum state of an open quantum system is described by density operator ρ which is a positive semidefinite operator with trace 1. The quantum evolution of state ρ can intuitively be regarded as a transform $\mathcal{E}$ acting on ρ, which maps density operator to density operators, viz., $\mathcal{E}(\rho)$ is also a density operator for arbitrary density operator ρ.

3.3.1 Kraus operator-sum representation

From the open-system viewpoint, the evolution of an open system S can be understood as the reduced part of a closed system SE where

The completeness property of Kraus operators read $\sum_a K_a^\dagger K_a = I$

§ 3.4 Channel state duality

We have see that a quantum channel is a CPTP superoperator, and a quantum state is a positive semidefinite trace-one operator, they seem to be very different. However, as we will show now, they are equivalent in the sense which will be clarified later.

§ 3.5 Natural representation of quantum channel

The first representation we will discuss is the so-called natural representation. From the mathematical point of view, superoperator is nothing but a special kind of operators, thus we can treat them in the same way as for operators. Although natural representation provide us with a straightforward way to represent a superoperator, this representation has the shortcoming in characterizing the properties of superoperators, like positivity, trace-preserving, because it essentially throw away the operator structure of input and output operators. This will be remedied by other representations that we will discuss later in this chapter.

3.5.1 Operator-vector correspondence

Before we discuss the natural representation of quantum channel, we first introduce a useful tool, *operator-vector correspondence*. This correspondence says that for any operator there exist a corresponding bipartite vector, and conversely, for every bipartite vector, there is a corresponding operator.

This can be shown by defining a linear isomorphism, which we call *vector mapping*,

$$\|\bullet\rangle\!\rangle : \mathbf{B}(\mathcal{H}_A, \mathcal{H}_B) \to \mathcal{H}_B \otimes \mathcal{H}_A \tag{3.2}$$

In the given bases of $\{|i_B\rangle\}$ and $\{|j_A\rangle\}$ of $\mathcal{H}_B$ and $\mathcal{H}_A$, it's as

$$\|I_{i,j}\rangle\!\rangle = \|(|i_B\rangle\langle j_A|)\rangle\!\rangle = |i_B\rangle \otimes |j_A\rangle. \tag{3.3}$$

Hereinafter we use "double-ket" notation to denote the vector map. For a given operator $A \in \mathbf{B}(\mathcal{H}_A, \mathcal{H}_B)$ expressed in a given basis

$$A = \sum_{i,j} A_{ij} |i_B\rangle\langle j_A| \tag{3.4}$$

with $A_{ij} = \langle i_B|A|j_A\rangle$, the vector mapping sends A to a bipartite vector

$$\|A\rangle\!\rangle = \sum_{i,j} A_{ij}|i_B\rangle \otimes |j_A\rangle = \sum_j A|j_A\rangle \otimes |j_A\rangle = (A \otimes I)\sum_j |j_A\rangle \otimes |j_A\rangle. \quad (3.5)$$

Notice that the vector map is independent of the the basis choice of $\mathcal{H}_B$ but dependent of the basis choice of $\mathcal{H}_A$. This is clear from the expression (3.5) and

$$\sum_\alpha |\alpha\rangle \otimes |\alpha\rangle = (U \otimes U)\sum_j |j_A\rangle \otimes |j_A\rangle. \quad (3.6)$$

The vector map is a bijection and also an isometry

$$\langle A, B\rangle_{\mathrm{HS}} = \mathrm{Tr}(A^\dagger B) = \langle\!\langle A\|B\rangle\!\rangle, \quad (3.7)$$

where we have adopted the notation $\langle\!\langle A\| = (\|A\rangle\!\rangle)^\dagger$.

For $A, B \in \mathbf{B}(\mathcal{X}, \mathcal{Y})$, from direct calculation we have

$$\mathrm{Tr}_{\mathcal{X}}\, \|A\rangle\!\rangle\langle\!\langle B\| = AB^\dagger, \quad (3.8)$$

$$\mathrm{Tr}_{\mathcal{Y}}\, \|A\rangle\!\rangle\langle\!\langle B\| = (B^\dagger A)^T. \quad (3.9)$$

For $A \in \mathbf{B}(\mathcal{X}_1, \mathcal{Y}_1)$, $B \in \mathbf{B}(\mathcal{X}_2, \mathcal{Y}_2)$ and $\rho \in \mathbf{B}(\mathcal{X}_2, \mathcal{X}_1)$, we have

$$(A \otimes B)\|\rho\rangle\!\rangle = \|A\rho B^T\rangle\!\rangle. \quad (3.10)$$

Conversely, for every bipartite pure state $|\psi\rangle_{AB} = \sum_{i,j} c_{ij}|i_B\rangle \otimes |j_A\rangle \in \mathcal{H}_B \otimes \mathcal{H}_A$, we can associate it with an operator $A_{|\psi\rangle_{AB}} : \mathcal{H}_A \to \mathcal{H}_B$ with $A = \sum_{ij} c_{ij}|i_B\rangle\langle j_A|$. This trick is useful in the study of quantum correlations.

3.5.2 The natural representation

From the previous discussion we known that quantum operations, including quantum channels and quantum measurements, are linear superoperators. Using the operator-vector correspondence, for any superoperator $\mathcal{E} : \mathbf{B}(\mathcal{X}) \to \mathbf{B}(\mathcal{Y})$, we have a corresponding operator $N(\mathcal{E}) : \mathcal{X} \otimes \mathcal{X} \to \mathcal{Y} \otimes \mathcal{Y}$:

$$\|\rho\rangle\!\rangle = \|\mathcal{E}(\rho)\rangle\!\rangle = N(\mathcal{E})\|\rho\rangle\!\rangle, \quad \forall \rho \in \mathbf{B}(\mathcal{X}). \quad (3.11)$$

In the computational basis $I_{a,b} = |a\rangle\langle b|$, it's easy to check that

$$\begin{aligned} N(\mathcal{E}) &= \sum_{a,b\in\Gamma_{\mathcal{X}}} \sum_{c,d\in\Gamma_{\mathcal{Y}}} \langle I_{c,d}, \mathcal{E}(I_{a,b})\rangle \|I_{c,d}\rangle\!\rangle\langle\!\langle I_{a,b}\| \\ &= \sum_{a,b\in\Gamma_{\mathcal{X}}} \sum_{c,d\in\Gamma_{\mathcal{Y}}} \langle I_{c,d}, \mathcal{E}(I_{a,b})\rangle I_{c,a} \otimes I_{d,b}. \end{aligned} \quad (3.12)$$

Since vector map $\|\bullet\rangle\rangle$ and $\mathcal{E}$ are both linear, $N(\mathcal{E})$ is also a linear operator. The natural representation is itself linear,

$$N(\alpha\mathcal{E}_1 + \beta\mathcal{E}_2) = \alpha N(\mathcal{E}_1) + \beta N(\mathcal{E}_2). \tag{3.13}$$

Notice that natural representation preserves that compostion

$$N(\mathcal{E} \circ \mathcal{F}) = N(\mathcal{E})N(\mathcal{F}) \tag{3.14}$$

This can be verified using expression (3.12). When focusing on the evolution of a given quantum sytems, this means that natural representation is a semigroup homomorphism. The above property implies that natural representation $N(\mathcal{E}^{-1})$ of left (resp. right) inverse $\mathcal{E}$ is the left (resp. right) inverse of $N(\mathcal{E})^{-1}$, i.e.,

$$N(\mathcal{E}^{-1}) = N(\mathcal{E})^{-1}. \tag{3.15}$$

The natural representation also respects the notion of Hermitian adjoint,

$$N(\mathcal{E}^{\dagger}) = N(\mathcal{E})^{\dagger}. \tag{3.16}$$

Recall that here the Hermitian adjoint is defined under Hilbert-Schmidt innner product $\langle \rho, \mathcal{E}(\sigma)\rangle = \langle \mathcal{E}^{\dagger}(\rho), \sigma\rangle$.

Exercise 3.1. Check the above properties of natural representation.

§ 3.6 Choi-Jamiołkowski representation

To remedy the shortcoming of natural representation in characterizing positive semidefiniteness and trace-preserving. We now introduce a powerful representation called Choi-Jamiokowski representation , also known as *channel-state correspondence*.

3.6.1 Choi-Jamiokowski representation

Theorem 3.1 (Choi-Jamiokowski isomorphism). *Consider two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, and let $|\Omega\rangle = \sum_{i=1}^{d_A} |ii\rangle \in \mathcal{H}_A \otimes \mathcal{H}_A$ and $E_{ij} = |i\rangle\langle j|$. The Choi-Jamiokowski map $J : \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B)) \to \mathcal{L}(\mathcal{H}_B \otimes \mathcal{H}_A)$ defined as*

$$J(\mathcal{E}) = \mathcal{E} \otimes \mathcal{I}(|\Omega\rangle\langle\Omega|) = \mathcal{E}(E_{ij}) \otimes E_{ij} \tag{3.17}$$

is a linear isomorphism. Its inverse map is given by $J^{-1}(\rho_{BA})(\sigma_A) = \mathrm{Tr}_A[(I_B \otimes \sigma_A^T)\rho_{BA}]$

Proof. □

Exercise 3.2. Prove that for two superoperators $\mathcal{M} \in \mathbf{T}(\mathcal{H}_A, \mathcal{H}_B)$ and $\mathcal{N} \in \mathbf{T}(\mathcal{H}_B, \mathcal{H}_C)$ we have

$$J(\mathcal{N} \circ \mathcal{M}) = \mathrm{Tr}_B[(I_C \otimes J(\mathcal{M})^{T_B})(J(\mathcal{N}) \otimes I_A)] \tag{3.18}$$

§ 3.7 Equivalence of three representations

We have provided three different wats to represent quantum operations and quantum channels. In this section, let's consider how to translate these representation to each other.

§ 3.8 Lindblad equation

§ 3.9 Examples of quantum channels

3.9.1 Depolarizing channel

Theorem 3.2 (polar decomposition). *If $A : \mathcal{V} \to \mathcal{V}$ is a linear map on a finite-dimensional inner-product space $\mathcal{V}$, then there exist positive semidefinite operator L, R and unitary U, such that A decomposes as*

$$A = L_A U = U R_A, \tag{3.19}$$

where $L_A = \sqrt{AA^\dagger}$ and $R_A = \sqrt{A^\dagger A}$ are uniquely determined by A, and U is unique if A is invertible.

Proof.

Definition 3.1. The fidelity between two states ρ and σ is defined as

$$F(\rho, \sigma) := \mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \tag{3.20}$$

Part II
Quantum Shannon Theory

Chapter 4
Classical Shannon Theory

In this chapter, we will discuss in some depth the *classical Shannon theory* or *Classical information theory*, which is established by Shannon and is one of the greatest discovery of 20th century. The basic concepts and theorems which useful for us to understand the their quantum analogues will be discussed.

The basic model for communication is Shannon-Weaver model which consists of three parts: the *information source*, the *information channel* and the *information receiver*. The basic process of the communication is like

$$\begin{aligned}&\text{information source} \to \text{encoding} \to \text{information channel}\\ &\to \text{information receiver } y = x + \varepsilon \to \text{decoding},\end{aligned} \tag{4.1}$$

where the information source wants to send a message to the receiver, he first encodes the information as a string of letters $x \in \mathcal{H}_C$ chosen from an alphabeta Γ, then he sends the string of letters using an information channel which may introduce the errors ε in this process, the receiver then obtain the $y = x + \varepsilon$, finally, he tries to decode from the received string and recover the message $x = y - \varepsilon$. In this chapter, we consider even simpler model which omits the encoding and decoding process, since the they form an independent theme, classical and quantum *error-correcting codes* which we will discuss in subsequent chapters. So now, we need first mathematically model the information source and information channel. Shannon's theory is largely based on probability theory, as we will see later, the information source is characterized as a random variable and the information channel is characterized a matrix whose entries and conditional probabilities.

There are some main thrusts of the Shannon theory, (i) how to quantify, characterize and transform information; (ii) how redundant a message is or how to compress the information; (iii) how to transmit information reliably using the noise channel. We will find the Shannon entropy play a crucial role. In this chapter, these topics will be discussed in its modest level and we will mainly focus on asymptotic case.

§ 4.1 Mathematical model for information source

An information source can send messages which consists of strings of letters chosen from a given alphabet Γ. Each letter $x \in \Gamma$ appears with a corresponding probability $p(x)$. This means that we can regard an information source as a random variable X which takes values in Γ.

Definition 4.1 (Information source). An (discrete) information source is a random variable X which takes values from a given alphabet $\Gamma = \{0, 1, \dots, d-1\}$.

Each letter $x \in \Gamma$ contains information, we can ask how much information a it contains. Shannon notice that we can quantify the information contained in x by the uncertainty before we know the exact value of x. For example, suppose that someone is playing dice, if a he told you that he get 2, before you receive this message, your uncertainty of his result is $1/6$, after you obtain the message, you are certain with his result. The information contained in this message is thus roughly $1-1/6 = 5/6$. The larger $p(x)$ is, the less information it contains. If you obtain a message x="tomorrow the sun will rise in the east", before or after you obtain the message, you are both certain with this fact, thus the message contains no information. With these observations, we can quantify the information contained in a message x as

$$I(X = x) = \log_2 \frac{1}{p(x)} = -\log_2 p(x). \tag{4.2}$$

The logarithm base does not matter, we can choose it as any positive value. Hereinafter, we will work in bit case, so we choose it as 2.

For an information source $X := \{x, p(x)\}$, the information contained in each letter is $I(X = x) = -\log_2 p(x)$, we can naturally regard the information contained in the source is probabilistic average of each letters $I(X) = \sum_x p(x) I(X = x) = -\sum_x p(x) \log_2 p(x)$, this quantity if noting but the famous *Shannon entropy*

$$H(X) = -\sum_x p(x) \log_2 p(x). \tag{4.3}$$

Consider the special case where alphabet $\Gamma = \{0, 1\}$ with $p(X = 0) = p$ and $p(X = 1) = 1 - p$. The corresponding Shannon entropy is so crucial thus has special name *binary Shannon entropy* and denotes $h(p)$,

$$h(p) = -p \log_2 p - (1-p) \log_2 (1-p). \tag{4.4}$$

See Figure 4.1 for its graph. It's symmetric along $p = 0.5$ and when $p = 0.5$ it takes the maximum value $h(0.5) = 1$.

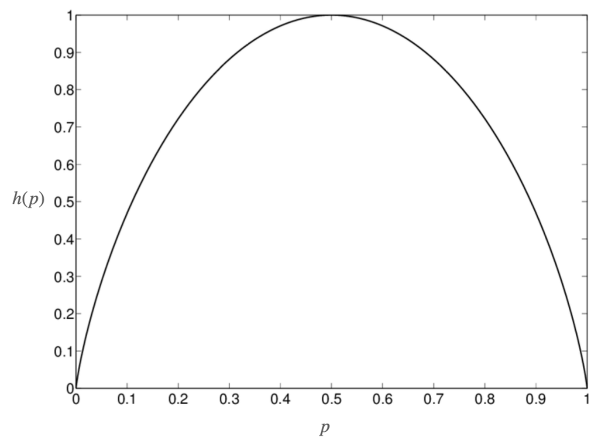


Fig. 4.1 The graph of binary Shannon entropy function $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$.

The above argument that $H(p)$ quantifies the average information contained in each letter of information source can be made rigorous. Let's do it now.

4.1.1 Shannon entropy and data compression

For a given information source $X = \{x, p(x)\}$ with X taking values in an alphabet $\Gamma = \{0, 1, \cdots, d-1\}$.

Theorem 4.1 (The law of large numbers).

4.1.2 Properties of Shannon entropy

It's a good place to introduce some other crucial entropy functions

Definition 4.2 (Entropy functions).

1. Shannon entropy: $H(p_i) = -\sum_i p_i \log_2 p_i$;
2. Rényi entropy: $H_\alpha(p_i) = \frac{1}{1-\alpha} \log_2 \sum_i p_i^\alpha$;
3. Tsallis entropy: $S_q(p_i) = \frac{1}{q-1}(1 - \sum_i p_i^q)$;
4. Min-entropy: $H_\infty = -\log_2(\max\{p_i\})$;
5. Collision entropy: $H_2(p_i) = -\log_2 \sum_i p_i^2$

§ 4.2 Data compression

§ 4.3 Channels

Chapter 5
Quantum Shannon Theory

The basic model for communication is

Entropy is thus a measure of uncertainty or 'ignorance' about a probabilistic system.

§ 5.1 basics of quantum error correction

Chapter 6
Classical error-correcting codes

Chapter 7
Stabilizer code

Stabilizer code is the quantum analogue of the classical additive code, thus it is sometimes called quantum additive code. The philosophy of stabilizer code is that instead of studying the code space $\mathcal{C}$, we focus on the stabilizer operators T_i of the code space, the code space is invariant under the stabilizer operators $T_i\mathcal{C} \subseteq \mathcal{C}$. This is similar as what we have done for classical linear code, where we focus on the encoding map and check matrices instead of code space itself. The stabilizer formalism turns out to be very convenient.

§ 7.1 Pauli group and stabilizer group

The advantage of stabilizer formalism originated from the clever use of group theory of the n-qubit unitary group $\mathbf{U}(2^n)$, and its subgroups—Pauli group $\mathbf{P}_n$ and Clifford group $\mathbf{C}_n$. As some of the readers may not be familiar with these notions, we briefly recall the definition and properties of the mathematical terms we will use here.

7.1.1 Pauli groups

Since we are working in a n-qubit Hilbert space $(\mathbb{C}^2)^{\otimes n}$, the *$n$-qubit unitary group* $\mathbf{U}(2^n)$ consists of all unitary operators over the Hilbert space. In the computational basis, $U(2^n)$ consists of all $2^n \times 2^n$ unitary matrices. The n-qubit Pauli group is a finite subgroup of $U(2^n)$, which is generated by Pauli matrices (the group operation is matrix mutiplication). The rigorous definition is as follows:

Definition 7.1 (Pauli group). The n-qubit Pauli group, denoted as $\mathbf{P}_n$, is defined as

$$\mathbf{P}_n = \{e^{i\theta}\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} | i_k = 0, 1, 2, 3, \text{ and } \theta = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}. \quad (7.1)$$

Here, $\sigma_0, \cdots, \sigma_3$ are Pauli matrices. The order, viz., the number of elements, of $\mathbf{P}_n$ is 4^{n+1}.

As we have mentioned in chapter 1, the one-qubit Pauli group is

$$\mathbf{P}_1 = \{\pm I, \pm X, \pm Y, \pm Z, \pm iI, \pm iX, \pm iY, \pm iZ\}. \quad (7.2)$$

Note that, for convenience, we will use the notations I, X, Y, Z and $\sigma_0, \sigma_1, \sigma_2, \sigma_3$ for Pauli matrices interchangeably.

Exercise 7.1. Using the relation $\sigma_i\sigma_j = \delta_{ij} + i\varepsilon_{ijk}\sigma_k$ to prove that for any pair of elements $g, g' \in \mathbf{P}_n$, they can only be commutative $gg' = g'g$ or anticommutative $gg' = -g'g$.

The cyclic group $\langle w_4 = e^{2i\pi/4}\rangle = \{e^0, e^{i\pi/2}, e^{i\pi}, e^{i3\pi/2}\}$ is a normal subgroup of $\mathbf{P}_n$ is the sense that $e^\theta = e^\theta I$. Then we can construct a quotient group $\mathbf{P}_n^* = \mathbf{P}_n/\langle w_4 = e^{2\pi/4}\rangle$. You can regard the group $\mathbf{P}_n^*$ as the group which consists of $\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}$, for example,

$$\mathbf{P}_1^* = \{[I], [X], [Y], [Z]\} = \{I, X, Y, Z\} \quad (7.3)$$

the phase factor do not appear in this group. Here is a comment for the readers who care more about mathematical strinency: we use the rigorous notation $[X]$ to mean the equivalence class of X in $\mathbf{P}_1$,

$$[X] := \{\pm X, \pm iX\}, \quad (7.4)$$

but for convenience, in the following discussion, we will just use the representative element X to represent the equivalence class $[X]$ whenever there is no risk to lead ambiguity.

It's obvious that any element $g \in \mathbf{P}_n^*$ is idempotent, that is, $g^2 = I$. And $\mathbf{P}_n^*$ is an Abelian group, namely, for any $g, g' \in \mathbf{P}_n^*$, we have $gh = hg$ (since for any two element $g, g' \in \mathbb{P}_n$, we either have $gg' = g'g$ or $gg' = -g'g$, but the factor in $\mathbf{P}_n^*$ is suppressed).

The $\mathbb{Z}_2$-vector representation of Paul group.—There is an important representation of $\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}$ with $2n$-dimensional $\mathbb{Z}_2$ vectors, i.e., with 0,1 sequences of length $2n$. In mathematical language, we can construct a $2n$-dimensional $\mathbb{Z}_2$ an isomorphism between $\mathbf{P}_n^*$ and additive group $\mathbb{Z}_2^{2n}$,

$$\varphi : \mathbf{P}_n^* \to \mathbb{Z}_2^{2n} \tag{7.5}$$

To see how it works, let us consider the simplest case $\varphi : \mathbf{P}_1^* \to \mathbb{Z}_2^2$, where

$$\varphi(I) = (0,0), \varphi(X) = (1,0), \varphi(Z) = (0,1), \varphi(Y) = (1,1) \tag{7.6}$$

The multiplication of Pauli matrices coincides with the addition of vectors, e.g., $\varphi(Y) = \varphi(XZ) = \varphi(X) + \varphi(Z)$. For the two-qubit case, we can set

$$\begin{aligned}
&\varphi(I_1 \otimes I_2) = (0_1, 0_2|0_1, 0_2), \quad \varphi(I_1 \otimes X_2) = (0_1, 1_2|0_1, 0_2), \\
&\varphi(X_1 \otimes I_2) = (1_1, 0_2|0_1, 0_2), \quad \varphi(I_1 \otimes Z_2) = (0_1, 0_2|0_1, 1_2), \\
&\varphi(Z_1 \otimes I_2) = (0_1, 0_2|1_1, 0_2), \quad \varphi(X_1 \otimes X_2) = (1_1, 1_2|0_1, 0_2), \\
&\varphi(Z_1 \otimes Z_2) = (0_1, 0_2|1_1, 1_2), \quad \cdots
\end{aligned} \tag{7.7}$$

Here the subscripts are used to indicate the label of qubit and the Y term can be obtained from X and Z term by adding the corresponding vectors, so we omit them.

For general n-qubit Pauli matrices $\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}$ which is represented as a $2n$ vector, σ_{i_1} is represent by the first and the $(n+1)$–th components of the vector, σ_{i_2} is represented by the second and the $(n+2)$-th components of the vector, etc. For example, $X \otimes Z \otimes I$ can be represented as $(1,0,0|0,1,0)$. It's obvious that for a $\mathbb{Z}_2$-vector

$$v = (a_1, \cdots, a_n | b_1, \cdots, b_n), \tag{7.8}$$

the value a_j indicates if there is a X operator in j-th qubit, and the value of b_j indicates if there is a Z operator in j-th qubit. When there are both X and Z operators in j-th qubit, there should be a Y operator there.

Now we are at a position to use this $\mathbb{Z}_2$-vector representation to explore the properties of the Pauli group. We start by discussing a very special transformation of the Pauli group using the Hadamard matrix

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{7.9}$$

We assign each Pauli matrix σ_i to its Hadamard conjugation $H\sigma_i H^\dagger = H\sigma_i H$, for example, for one qubit case, we have

$$I \mapsto HIH = I, X \mapsto HXH = Z, Z \mapsto HZH = X, Y \mapsto HYH = -Y. \tag{7.10}$$

Translating them into the $\mathbb{Z}_2$-vector representation, we obtain

$$(0,0) \mapsto (0,0), (1,0) \mapsto (0,1), (1,0) \mapsto (0,1), (1,1) \mapsto (1,1). \tag{7.11}$$

Thus the transformation can be represented by a matrix

$$\Lambda_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{7.12}$$

Similarly, we can analyze the n-qubit Hadamard conjugation by applying $H \otimes \cdots \otimes H$. A few moments thinking lead the result that the corresponding matrix is

$$\Lambda_n = \begin{pmatrix} 0^* & I_n \\ I_n & 0^* \end{pmatrix}, \tag{7.13}$$

which is written in block form, 0^* is a $n \times n$ matrix will all entries zeroes and I_n is the $n \times n$ identity matrix.

Exercise 7.2. Prove that the the Hadamard conjugation, in $\mathbb{Z}_2$-vector representation, is represented by Λ_n.

Exercise 7.3. Give the matrix in $\mathbb{Z}_2$-vector representation for the conjugation operations corresponding to X, Y, Z and control not $\Lambda(X)$.

The matrix Λ_n turn out to be useful for analyzing the Pauli group $\mathbf{P}_n^*$.

Proposition 7.1. *Two elements $g = \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}$ $g' = \sigma_{j_1} \otimes \cdots \otimes \sigma_{j_n}$ in $\mathbf{P}_n^*$ are commutative if and only if*

$$\varphi(g)\Lambda_n \varphi(g')^T = 0. \tag{7.14}$$

Proof. We only need to count the number of qubit where $\sigma_{i_k} = \sigma_{j_k}$ for g and g',

$$d(g, g') = \#\{k = 1, \cdots, n | \sigma_{i_k} = \sigma_{j_k}\}. \tag{7.15}$$

If the number $d(g, g')$ is odd, g and g' are anticommutative; if the number is even, g and g' are anticommutative. Then using the matrix Λ_n, it's easily to see that $\varphi(g)\Lambda_n\varphi(g')^T = 1$ if $d(g, g')$ is odd and $\varphi(g)\Lambda_n\varphi(g')^T = 0$ if $d(g, g')$ is even. □

Consider several elements $g_1, \cdots g_l \in \mathbf{P}_n^*$, they are called independent if any one of them can not be represented as a product of the other elements. As $\mathbf{P}_n^*$ is Abelian and each element in it is idempotent, the group generated by $g_1, \cdots g_l$ is

$$\langle g_1, g_2, \cdots g_l \rangle = \{g = g_1^{\alpha_1} g_2^{\alpha_2} \cdots g_n^{\alpha_n} | \alpha_i = 0, 1\}. \tag{7.16}$$

Therefore $g_1, \cdots g_l$ are independent if, any group generated by g_1, $\cdots$, g_{i-1}, g_{i+1}, $\cdots$, g_l where some g_i is removed is smaller than $\langle g_1, g_2, \cdots g_l \rangle$, i.e.,

$$\langle g_1, \cdots, g_{i-1}, g_{i+1}, \cdots, g_l \rangle < \langle g_1, g_2, \cdots g_l \rangle. \tag{7.17}$$

To check if a given set of elements are independent or not is usually very time consuming using the current methodology. We will see that this can be done very easily using the $\mathbb{Z}_2$-vector representation.

Proposition 7.2. *A set of elements $g_1, \cdots g_l \in \mathbf{P}_n^*$ are independent if and only if the vectors $\varphi(g_1), \cdots, \varphi(g_1)$ are linearly independent.*

Proof. The key feature we are to use is that φ is a group homomorphism, i.e., $\varphi(gg') = \varphi(g) + \varphi(g')$ We can use this to prove the contrapositive: $g_1, \cdots g_l \in \mathbf{P}_n^*$ are not independent if and only if vectors $\varphi(g_1), \cdots, \varphi(g_1)$ are linearly dependent, i.e., there exist a set of $a_i = 0, 1$ (not all zero) such that $\sum_i a_i \varphi(g_i) = 0$.

If $g_1, \cdots, g_l$ are not independent, we must have $g_1^{\alpha_1} \cdots g_l^{\alpha_l} = I$ for some $\alpha_i = 0, 1$ (not all zero). This equivalent to

$$\varphi(g_1^{\alpha_1}) + \cdots \varphi(g_l^{\alpha_l}) = 0 \Leftrightarrow \alpha_1 \varphi(g_1) + \cdots + \alpha_l \varphi(g_l) = 0. \tag{7.18}$$

Since $\alpha_i = 0, 1$ are not all zero, thus $\varphi(g_1), \cdots, \varphi(g_1)$ are linearly dependent.

7.1.2 Stabilizer group

§ 7.2 Clifford group

Consider a group G and its subgroup H, the normalizer $N_G(H)$ of H in G is defined as the smallest subgroup of G which contains H as a normal subgroups. Equivalently the normalizer of subgroup H is defined as:

Definition 7.2 (Normalizer). For a give group G and its subgroup H, the normalizer of H in G is defined as

$$N_G(H) = \{g \in G | gHg^{-1} = H\}. \tag{7.19}$$

The normalizer $N_G(H)$ is also a subgroup of G and contains H as a normal subgroup.

Definition 7.3 (Clifford group). The n-qubit Clifford group is defined as the quotient group of the normalizer $N(\mathbf{P}_n)$ of Pauli group in $U(2^n)$ with $U(1) = \{e^{i\theta} | \theta \in [0, 2\pi)\}$. More precisely, the n-qubit Clifford group, denoted as $\mathbf{C}_n$, is defined as

$$\mathbf{C}_n = \{V \in U(2^n) | V\mathbf{P}_n V^\dagger = \mathbf{P}_n\}/U(1). \tag{7.20}$$

Note that for $V \in U(2^n)$ if for all $\sigma \in \mathbf{P}_n$ we have $V\sigma V^\dagger$, then we also have $V' = e^{i\theta}V$ satisfying $V'\sigma V'^\dagger$for all $\sigma \in \mathbf{P}_n$, this is the reason that the phase factor do not appear in Clifford group.

Notice that conjugation by U is a automorphism of $\mathbf{P}_n$, it must preserve the group operations. Since $V\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} V^\dagger = e^{i\theta}\sigma_{i'_1} \otimes \cdots \otimes \sigma_{i'_n}$, we see that the square of the left had side equals to I, thus the square of the right hand side must also be I, which impose the conditions on θ that $\theta = 0, \pi$. The condition of the definition of Clifford group can thus be simplified as

$$V\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} V^\dagger = \pm\sigma_{i'_1} \otimes \cdots \otimes \sigma_{i'_n},$$

for all possible Pauli matrices $\sigma_{i_1} \cdots \sigma_{i_n}$. Again since $\sigma_2 = -i\sigma_2\sigma_1$, we actually only need to set the constraint to the X, Z Pauli matrices. In summary, the Pauli group is equivalently defined as

$$\{V \in U(2^n) | V\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n} V^\dagger = \pm\sigma_{i'_1} \otimes \cdots \otimes \sigma_{i'_n} \text{for all}\, \sigma_{i_k} = X, Z\}/U(1),$$

note that here $\sigma_{i'_k}$ can be taken as Y. With this property, the number of the elements in Clifford group $\mathbf{C}_n$ can be determined:

$$|\mathbf{C}_n| = \prod_{i=1}^{n} 2 \times 4^j(4^j - 1) = 2^{n^2+2n} \prod_{j=1}^{n} (4^j - 1).$$

Theorem 7.1 (Gottesman). *The normalizer* $N(\mathbf{P}_n) = \{V \in U(2^n) | V\mathbf{P}_n V^\dagger = \mathbf{P}_n\}$ *of Pauli group* $\mathbf{P}_n$ *in unitary group* $U(2^n)$ *is generated from* $\{H_i, S_j, \Lambda_{ij}(X)\}$, *where* H_i *is Hadamard gate,* S *is phase gate, and* $\Lambda_{ij}(X)$ *is the CNOT gate.*

Proof. We now prove the theorem in several steps.

Step 1: $N(\mathbf{P}_1)$ is generated from H and S. Suppose that $U \in N(\mathbf{P}_1)$, then the map $Ue^{i\theta}\sigma U^\dagger \mapsto e^{i\theta'}\sigma'$ defines a group automorphism of $\mathbf{P}_1$, thus it must preserve the group structure of $\mathbf{P}_1$. Then the action of U on $\mathbf{P}_1$ is captured by the the action of U on X and Z.

$$UXU^\dagger = e^{i\theta_X}\sigma(X);\ UZU^\dagger = e^{i\theta_Z}\sigma(Z).$$

Taking squares for both sides of the two equations, it's easy to see that $\theta_X, \theta_Z = 0, \pi$, viz.,

$$UXU^\dagger = \pm\sigma(X);\ UZU^\dagger = \pm\sigma(Z).$$

Exercise 7.4. Suppose that $U, V \in U(2^n)$ are unitary operators on n qubits which transform $Z_1, \cdots, Z_n,\ X_1, \cdots, X_n$ by conjugation in the same way, i.e., $U(\cdot)U^\dagger = V(\cdot)V^\dagger$. Show that $U = e^{i\theta}V$ for some real number θ.
Hint: First, notice that the relation $U(\cdot)U^\dagger = V(\cdot)V^\dagger$ holds for all X and

Z operators implies that it holds for all Pauli matrices $\sigma = \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}$, since X and Z operators can generate all other Pauli operators. Then, from the fact that n-qubit Pauli matrices form a basis of the complex vector space $M_{2^n}(\mathbb{C})$, which is the space of all complex $2^n \times 2^n$ matrices, we know that $U(\cdot)U^\dagger = V(\cdot)V^\dagger$ holds for all $2^n \times 2^n$ matrices. This further implies that $V^\dagger U A = A V^\dagger U$ for all $A \in M_{2^n}(\mathbb{C})$.

Secondly, we claim that if a linear operator T commute will all other linear operators, then it must be a multiple of identity, i.e., $T = cI$ for some $c \in \mathbb{C}$.

Aut $\cong$

§ 7.3 Stabilizer state

§ 7.4 Stabilizer group

§ 7.5 Stabilizer quantum code

§ 7.6 Calderbank-Shor-Steane code

Chapter 8
Topological error-correcting code

Chapter 9
Appendix: mathematical preliminaries

§ 9.1 d

Index