Lecture 3 – Basics of quantum information

Peter Brown

October 17, 2022

So far we have seen the following quantum formalism:

- States: Unit norm vectors in a Hilbert space $|\psi\rangle \in \mathcal{H}$.
- Transformations: Unitary operators (map unit norm vectors to unit norm vectors).
- Measurements: Projective / POVMs

However, this does not capture everything that one can actually do as we shall soon see!

Before we move on let us note some notation. We will often use the system name (e.g. A) and the associated Hilbert space $\mathcal{H}_A$ interchangeably. The set of linear transformations from system A to itself (square matrices acting on A) is denoted by $\mathcal{L}(A)$. The set of linear transformations from system A to system B (possibly non-square matrices mapping from A to B) is denoted by $\mathcal{L}(A, B)$. The notation $X \geq 0$ for a Hermitian matrix X denotes that the matrix is positive semidefinite, i.e., X is Hermitian and $\langle v|X|v\rangle \geq 0$ for all vectors $|v\rangle$. We say $X \geq Y$ if $X - Y \geq 0$. An *isometry* is a linear map $V \in \mathcal{L}(A, B)$ such that $V^\dagger V = I_A$ (all unitaries are isometries). The trace is a square matrix X is defined as $\mathrm{Tr}[X] := \sum_i \langle i|X|i\rangle = \sum_i X_i i$ (the sum of it's diagonal entries).

1 Quantum states (density matrices)

Suppose we have a quantum system with an associated Hilbert space $\mathcal{H}$. Suppose further that we prepare the state of the system probabilistically. That is, we first sample an index $k \in \{1, \ldots, n\}$ according to a probability distribution $p(k)$ and then we prepare the system in the state $|\psi_k\rangle$. Suppose we now measure a POVM $\{M_j\}_j$, what is the probability that we obtain outcome j for the measurement? Well, by the rules of conditional probability

$$\begin{aligned} \mathbb{P}(\text{Outcome } j) &= \sum_k \mathbb{P}(\text{Prepared state } |\psi_k\rangle)\mathbb{P}(\text{Outcome } j \,|\, \text{Prepared state } |\psi_k\rangle) \\ &= \sum_k p(k)\mathrm{Tr}\left[|\psi_k\rangle\langle\psi_k|M_j\right] . \end{aligned} \tag{1}$$

which is correct. However, it is a little cumbersome. In particular, the state of the system is now a collection $\{p(k), |\psi_k\rangle\}_k$ which is a bit verbose. A much cleaner way to handle such a thing is to define a *density matrix*

$$\rho = \sum_k p(k)|\psi_k\rangle\langle\psi_k| \tag{2}$$

which now represents the state of the system. By equation 1 and the linearity of the trace we then have

$$\mathbb{P}(\text{Outcome } j) = \sum_k p(k)\mathrm{Tr}\left[|\psi_k\rangle\langle\psi_k|M_j\right] = \mathrm{Tr}\left[\left(\sum_k p(k)|\psi_k\rangle\langle\psi_k|\right) M_j\right] = \mathrm{Tr}\left[\rho M_j\right] \tag{3}$$

and so we have a much cleaner and compact way of describing the state of a system and the outcome probabilities of measurements $\mathbb{P}(\text{Outcome } j) = \mathrm{Tr}[\rho M_j]$.

These density matrices ρ have two defining properties

1. Unit trace: $\mathrm{Tr}[\rho] = 1$

Proof.

$$\mathrm{Tr}\left[\rho\right] = \mathrm{Tr}\left[\sum_k p(k)|\psi_k\rangle\langle\psi_k|\right] = \sum_k p(k)\mathrm{Tr}\left[|\psi_k\rangle\langle\psi_k|\right] = \sum_k p(k) = 1\,. \tag{4}$$

□

2. Positive semidefinite (PSD): $\rho = \rho^\dagger$ and $\langle x|\rho|x\rangle \geq 0$ for all $|x\rangle \in \mathcal{H}$.

Proof. Checking Hermitian is straightforward. Then for any $|x\rangle \in \mathcal{H}$ we have

$$\langle x|\rho|x\rangle = \sum_k p(k)\langle x||\psi_k\rangle\langle\psi_k||x\rangle = \sum_k p(k)|\langle x|\psi_k\rangle|^2 \geq 0 \tag{5}$$

□

Overall we land at a new definition for quantum states which allows us to take into account probabilistic preparations of our systems.

Postulate 1 (Quantum state)**.** Let $\mathcal{H}$ be a Hilbert space associated with a quantum system. Then any state of that system is described by a density matrix acting on $\mathcal{H}$. I.e., $\rho \in \mathcal{L}(\mathcal{H})$ such that $\mathrm{Tr}\left[\rho\right] = 1$ and $\rho \geq 0$. The set of all density matrices for the Hilbert space $\mathcal{H}$ is denoted by $\mathcal{D}(\mathcal{H})$.

A quantum state of rank 1, i.e., $\rho = |\phi\rangle\langle\phi|$ is called a *pure state* and these correspond to exactly the states we covered previously. Note that density matrices also get rid of the problem of global phase equivalence, if $|\psi\rangle = e^{i\theta}|\phi\rangle$ then $\rho = |\psi\rangle\langle\psi| = |\phi\rangle\langle\phi|$. That is they have the same density matrix.

A state of larger rank is called a *mixed state*.

Remark 1 (All density matrices are quantum states). To get to our new definition we started with an ensemble $\{p(k), |\psi_k\rangle\}$ of states, defined a matrix $\rho = \sum_k p(k)|\psi_k\rangle\langle\psi_k|$ and noticed it had two properties (unit trace and PSD). For the definition to be a good one it should be the case that all density matrices really correspond to states. Fortunately, this is true! By the spectral decomposition we can write any matrix ρ in the form

$$\rho = \sum_j \lambda_j |\phi_j\rangle\langle\phi_j| \tag{6}$$

where λ_j are the eigenvalues of ρ and $|\phi_j\rangle$ are the corresponding normalized eigenvectors. As ρ is PSD we have that $\lambda_j \geq 0$ and as $\mathrm{Tr}\left[\rho\right] = 1$ we have $\sum_j \lambda_j = 1$. That is, $\{\lambda_j\}$ forms a probability distribution. Moreover the normalized eigenvectors are then quantum states $|\phi_j\rangle$ in the ket form. Together this is implies that the density matrix ρ corresponds to state of a system where we prepare the state $|\phi_j\rangle$ with probability λ_j. NB: this interpretation is not unique, different ensembles can lead to the same density operator!

Example 1. (Qubit states) For a qubit state a density matrix ρ takes the general form

$$\rho = \begin{pmatrix} a & \beta \\ \beta^* & 1-a \end{pmatrix} \tag{7}$$

for $a \in [0,1]$ and $|\beta| \leq \sqrt{a(1-a)}$. You can also write it as

$$\rho = \frac{I + r_x X + r_y Y + r_z Z}{2} \tag{8}$$

where X, Y and Z are the Pauli matrices and $r_x^2 + r_y^2 + r_z^2 \leq 1$. This latter constraint defines a ball in $\mathbb{R}^3$ which is normally referred to as the Bloch ball/sphere see Figure 1 for details as well as Exercise 4.

1.1 Composing and discarding systems

As we've seen previously if we have two systems A and B with corresponding Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ the Hilbert space describing the two systems simultaneously is given by the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$. It follows then from Postulate 1 that the set of quantum states describing the two systems is given by $\mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ – we will also use the notation $\mathcal{D}(AB)$. Note that there are many more states in this set than those of the form $\rho_A \otimes \rho_B$ (for instance entangled states).

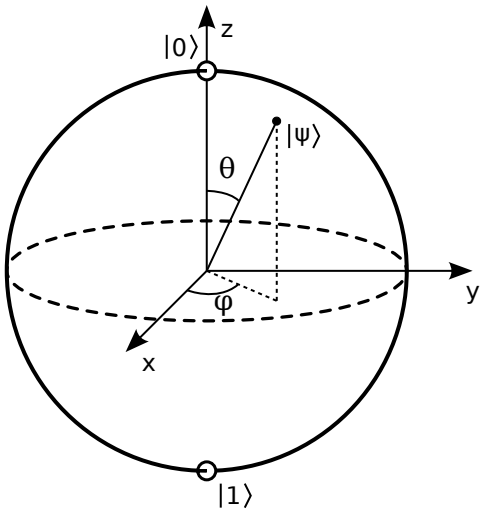

Figure 1: The Bloch-sphere/ball is a geometrical representation of single qubit states. A generic qubit pure state can be written as

$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle$$

with $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$ as depicted in the figure. The points inside the ball correspond to mixed states if we take the representation

$$\rho = \frac{I + r_x X + r_y Y + r_z Z}{2}$$

then the coordinates $(r_x, r_y, r_z) \in \mathbb{R}^3$ lie in the ball. And indeed if we take the spherical coordinates $(r_x, r_y, r_z) = (\cos(\phi)\sin(\theta), \sin(\phi)\sin(\theta), \cos(\theta))$ then we recover the pure states above $\rho = |\psi\rangle\langle\psi|$. With this representation the state at the center of the ball is $\rho = I/2$ the so-called *maximally mixed state*.

If we have a state $\rho_{AB} \in \mathcal{D}(AB)$ on the bipartite system AB, the density matrix formalism allows us to define a state on a subsystem using a map known as the partial trace.

Partial Trace
If we have a bipartite system AB the partial trace (over B) is a linear map

$$\mathrm{Tr}_B : L(\mathcal{H}_A \otimes \mathcal{H}_B) \to L(\mathcal{H}_A) \tag{9}$$

defined by linearly extending

$$\mathrm{Tr}_B : X \otimes Y \mapsto \mathrm{Tr}\,[Y]\, X\,. \tag{10}$$

If we have a multipartite system $A_1 A_2 \ldots A_n$ we can define the partial trace over system A_i in the same way by linearly extending

$$\mathrm{Tr}_{A_i}\,[X_1 \otimes X_2 \otimes \ldots X_n] = \mathrm{Tr}\,[X_i]\, X_1 \otimes \cdots \otimes X_{i-1} \otimes X_{i+1} \otimes \cdots \otimes X_n\,. \tag{11}$$

Example 2.

1. Consider a bipartite system AB. If $\{|i\rangle_A\}_i$ is the computational basis for system A and $\{|i\rangle_B\}_i$ is the computational basis for system B then any square matrix acting on the joint system AB can be written in the form

   $$X_{AB} = \sum_{ijkl} x_{ijkl} |i\rangle\langle j|_A \otimes |k\rangle\langle l|_B$$

   then

   $$X_A = \mathrm{Tr}_B\,[X_{AB}] = \sum_{ijkl} x_{ijkl}|i\rangle\langle j|_A \mathrm{Tr}\,[|k\rangle\langle l|_B] = \sum_{ijkl} x_{ijkl}|i\rangle\langle j|_A \delta_{kl} = \sum_{ijk} x_{ijkk}|i\rangle\langle j|_A$$

   where δ_{kl} is the Kronecker delta.

2. Let $|\psi\rangle_{AB} = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ be a maximally entangled state of two-qubits. The density operator representing this state is

$$\rho_{AB} = |\psi\rangle\langle\psi|_{AB} = \frac{|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|}{2} = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}$$

then the marginal state on system A (the state of knowledge of system A if we ignore system B) is given by $\rho_A = I/2$.

3. Another way to write the action of the partial trace over system B would be

$$\rho_A = \sum_i (I_A \otimes \langle i|_B)\rho_{AB}(I_A \otimes |i\rangle_B)\,.$$

Definition 1 (Purification)**.** Let ρ_A be a quantum state on some system A. A *purification* of ρ_A is a state $|\psi\rangle_{AB}$ on a joint system AB (system B is called the *purifying system*) such that

$$\rho_A = \mathrm{Tr}_B\left[|\psi\rangle\langle\psi|_{AB}\right]\,. \tag{12}$$

Pure states are in many respects simpler than mixed states so it can be convenient to expand the Hilbert space to a larger space and view your original system as a subsystem of a larger system. This way of thinking is sometimes referred to as the "church of the larger Hilbert space" and we will see that this way of thinking can be applied to many other concepts in quantum theory, extending the Hilbert space to view your object as a subobject of something simpler.

We note that purifications of quantum states always exist. For example consider an operator ρ_A, by the spectral theorem we can always write this operator in the form

$$\rho_A = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$$

where $\lambda_i \geq 0$ and $|\psi_i\rangle$ form an orthonormal basis of system A. Then let B be a system isomorphic to A. We can then define the following pure state on the joint system AB,

$$|\Phi\rangle_{AB} = \sum_i \sqrt{\lambda_i}|\psi_i\rangle_A \otimes |\psi_i\rangle_B\,.$$

A quick calculation shows that this is a purification of the state ρ_A, i.e., $\rho_A = \mathrm{Tr}_B\left[|\Phi\rangle\langle\Phi|\right]$.

Purifications bring with them an interesting interpretation of a mixed state. It says that we can see the mixture arising as a result of having incomplete information about a larger system with which our system is entangled.

Before moving onto transformations of quantum systems we note a very useful representation of a pure bipartite state known as the *Schmidt decomposition*.

Theorem 1 (Schmidt decomposition)**.** *Let $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a pure state on a bipartite system AB. Then there exist orthonormal bases $\{|i\rangle_A\}_i$, $\{|i\rangle_B\}$ of system A, B respectively and $\lambda_i \geq 0$ satisfying $\sum_i \lambda_i = 1$ such that*

$$|\psi\rangle = \sum_i \sqrt{\lambda_i}|i\rangle_A|i\rangle_B \tag{13}$$

The λ_i are called the Schmidt coefficients *of the state and the number of nonzero Schmidt coefficients is called the* Schmidt rank *of the state.*

Proof. We include a sketch of the proof because it is constructive. Start with the state

$$|\psi\rangle_{AB} = \sum_{mn} a_{mn}|m\rangle_A|n\rangle_B$$

written in any ONBs for systems A and B. We can define a matrix X by applying the map $|i\rangle \otimes |j\rangle \mapsto |i\rangle\langle j|$ (note this is actually an isomorphism between $\mathcal{H}_A \otimes \mathcal{H}_B$ and linear maps $\mathcal{L}(\mathcal{H}_B, \mathcal{H}_A)$), so we have

$$X = \sum_{mn} a_{mn}|m\rangle\langle n|\,.$$

By the *singular value decomposition* there exist unitary matrix U acting on $\mathcal{H}_A$, a unitary matrix V acting on $\mathcal{H}_B$ and a (possibly non-square) diagonal matrix D such that $X = UDV^\dagger$. The diagonal matrix contains the singular values of X and the columns of U, V are the left, right singular vectors of X respectively. Overall this implies that there exist scalars $\sqrt{\lambda_i}$ (singular values) and ONBs $\{|u_i\rangle_A\}_i$ and $\{|v_j\rangle\}_j$ for systems A and B (left/right singular vectors) such that

$$X = \sum_i \sqrt{\lambda_i} |u_i\rangle\langle v_i| \,.$$

By applying the inverse of the isomorphism from before we then find that

$$|\psi\rangle = \sum_i \sqrt{\lambda_i} |u_i\rangle|v_i\rangle$$

which is the desired result. □

Note that in the Schmidt decomposition the Hilbert spaces do not need to be of the same dimension – the orthonormal basis can have more elements than there are $\lambda_i > 0$.

Example 3.

1. The two-qubit state
$$\frac{|00\rangle + |11\rangle}{\sqrt{2}}$$
is already in Schmidt form and has Schmidt rank 2.
2. The two-qubit state $\frac{1}{\sqrt{2}}(|00\rangle + |01\rangle)$ can be written in Schmidt form $|0\rangle|+\rangle$ and has Schmidt rank 1.

1.2 Entanglement

Previously we said a pure bipartite state $|\psi\rangle_{AB}$ is entangled if it cannot be written as

$$|\psi\rangle_{AB} = |v\rangle_A \otimes |w\rangle_B \,.$$

States that factorize in such a way are called *product states*. For density matrices the theory of entanglement is a lot richer. For instance it is no longer enough to say that a state ρ_{AB} is entangled if it cannot be written in the form

$$\rho_{AB} = \sigma_A \otimes \tau_B \,. \tag{14}$$

This is because we now have access to probabilistic mixtures of states and so we can actually create more states than above (without the local systems interacting) by mixing such product states. This leads us to the following definition of entanglement for density matrices.

Definition 2 (Separable/entangled states)**.** Consider a bipartite system AB and let $\rho_{AB} \in \mathcal{D}(AB)$. The state ρ_{AB} is called *separable* if there exist $\sigma_i \in \mathcal{D}(A)$, $\tau_i \in \mathcal{D}(B)$ and a probability distribution p_i such that

$$\rho_{AB} = \sum_i p_i \sigma_i \otimes \tau_i \,. \tag{15}$$

Any state that is *not* separable is called *entangled*.

Separable states have a clear operational interpretation. We can create them by probabilistically preparing isolated (non-interacting) systems. If such a preparation is not possible the systems must have undergone some interaction and this is the source of their entanglement.

Example 4. We give some examples of separable and entangled states.

1. The maximally mixed state is separable
$$\rho_{AB} = I_{AB}/4 = (I_A/2) \otimes (I_B/2) \,.$$
2. The state $\rho_{AB} = p|00\rangle\langle 00| + (1-p)|11\rangle\langle 11|$ is separable as
$$\rho_{AB} = p|00\rangle\langle 00| + (1-p)|11\rangle\langle 11| = p|0\rangle\langle 0| \otimes |0\rangle\langle 0| + (1-p)|1\rangle\langle 1| \otimes |1\rangle\langle 1|$$

3. For any pure entangled state $|\psi\rangle_{AB}$ its associated density matrix $\rho_{AB} = |\psi\rangle\langle\psi|$ is also entangled under this new definition.

4. It turns out that the set of separable states forms a closed, convex subset of $\mathcal{D}(AB)$. This means that if we make a small enough perturbation to an entangled state then it should remain entangled. E.g., if we have a pure entangled state $|\psi\rangle_{AB}$ then we can choose $p > 0$ small enough such that the state

$$\rho_{AB} = (1-p)|\psi\rangle\langle\psi| + pI/(d_A d_B)$$

is still entangled.

2 Quantum channels (transformations)

We have previously seen that the evolution of a closed quantum system is given by a unitary transformation $|\psi\rangle \mapsto U|\psi\rangle$ or in the density matrix formalism $\rho \mapsto U\rho U^\dagger$. However, this is far from the whole picture. For example, suppose we have access to a system A which is not well isolated, i.e., it may interact with the environment which we'll denote by the system B. Let's assume that this larger system is now closed and so undergoes a global unitary evolution. If ρ_{AB} is the initial state of the joint system, after interacting there is some unitary U acting on the joint system such that the final state is $U\rho_{AB}U^\dagger$. We can describe the initial and final state of our system A by

$$\rho_A^{\text{initial}} = \text{Tr}_B\left[\rho_{AB}\right] \qquad \text{and} \qquad \rho_A^{\text{final}} = \text{Tr}_B\left[U\rho_{AB}U^\dagger\right]. \tag{16}$$

The question is then, does there exist a unitary V acting only on system A such that

$$\rho_A^{\text{final}} \stackrel{?}{=} V\rho_A^{\text{initial}}V^\dagger. \tag{17}$$

I.e., can we describe the dynamics we will observe at system A using the unitary formalism? In general the answer is no, as we shall soon see.

A simple mathematical argument to demonstrate that the unitary update from before is not enough is the fact that unitaries preserve eigenvalues. Hence ρ and $U\rho U^\dagger$ have the same eigenvalues, yet we know that mixed states can have very different eigenvalues so there should be a way to describe a transformation between the two systems. It turns out the general way to describe transformations of quantum systems is through something called a *quantum channel*.

Definition 3 (Quantum channel)**.** Let $\mathcal{H}_A$, $\mathcal{H}_B$ be Hilbert spaces. Then a linear map $\Phi : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$, is a *quantum channel* from system A to system B if it satisfies

1. **Trace preserving**: $\text{Tr}\left[X\right] = \text{Tr}\left[\Phi(X)\right]$ for all $X \in \mathcal{L}(\mathcal{H}_A)$.
2. **Positive**: $\Phi(X) \geq 0$ whenever $X \geq 0$.
3. **Completely positive**: For any additional Hilbert space $\mathcal{H}_C$ and $X_{AC} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_C)$ with $X_{AC} \geq 0$ we have

$$(\Phi \otimes \mathcal{I}_C)(X_{AC}) \geq 0$$

where $\mathcal{I}_C$ is the identity channel for system C, i.e., $\mathcal{I}_C(Y_C) = Y_C$.

That is, a quantum channel is a completely-positive trace preserving linear map (CPTP map).

In the above definition the tensor product of two channels should be interpreted in the following way. For a bipartite linear operator X_{AC} we can always write it as some linear combination of tensor products of operators $X_{AC} = \sum_i a_i Y_i \otimes Z_i$ where the Y_i are linear operators on system A and Z_i are linear operators on system C. Then for channels $\Phi : \mathcal{L}(A) \to \mathcal{L}(B)$ and $\Lambda : \mathcal{L}(C) \to \mathcal{L}(D)$ we have $(\Phi \otimes \Lambda) : \mathcal{L}(AC) \to \mathcal{L}(BD)$ and

$$(\Phi \otimes \Lambda)(X_{AC}) = \sum_i a_i \Phi(Y_i) \otimes \Lambda(Z_i).$$

This is exactly like how we defined the partial trace as a linear extension of a map applied to a subsystem. Indeed the partial trace is a quantum channel! Effectively a quantum channel is any linear map that sends density matrices to density matrices.

Postulate 2 (Evolution)**.** The evolution of a quantum system can be described by a quantum channel.

Example 5 (Examples of quantum channels). The following are all quantum channels

1. **Identity channel**: $\mathcal{I} : \mathcal{L}(H) \to \mathcal{L}(H)$ defined by $\mathcal{I}(\rho) = \rho$.
2. **Preparing σ channel**: $\Phi_\sigma : \mathcal{L}(A) \to \mathcal{L}(B)$ defined by
$$\Phi_\sigma(X) = \mathrm{Tr}\left[X\right]\sigma\,. \tag{18}$$
Operationally this channel corresponds to destroying system A and preparing the state σ on system B.
3. **Unitary channel**: $\Phi : \mathcal{L}(A) \to \mathcal{L}(A)$ defined as
$$\Phi(X) = UXU^\dagger \tag{19}$$
for some unitary matrix U. Operationally this corresponds to just applying a unitary evolution to the system, exactly like $|\psi\rangle \mapsto U|\psi\rangle$ from before.
4. **Unitary mixture channel**: $\Phi : \mathcal{L}(A) \to \mathcal{L}(A)$ defined as
$$\Phi(X) = \sum_{i=1}^{n} p_i U_i X U_i^\dagger \tag{20}$$
where U_i are all unitary matrices and p_i is a probability distribution. The action of this channel can be interpreted operationally as sampling a random integer $1 \leq i \leq n$ according to the distribution p_i and then updating the system according to the unitary U_i. To a person who knows that this was the procedure done but doesn't know what the outcome of the random variable was they would describe the evolution by the above channel.
5. **Depolarizing channel**: Let $p \in [0,1]$ and $d \in \mathbb{N}$ then the depolarizing channel with weight p is defined as $\Phi_p : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ where
$$\Phi_p(X) = (1-p)X + p\mathrm{Tr}\left[X\right] I/d\,. \tag{21}$$
Operationally this channel can be thought of as mixing the system with the maximally mixed state (uniform/white noise). It is a common simple model of noise.

Fortunately there are very nice characterizations of the set of quantum channels and standard ways to represent them. These characterizations are captured in the following lemma.

Lemma 1. *The following are equivalent.*

1. *$\Phi : \mathcal{L}(A) \to \mathcal{L}(B)$ is a quantum channel.*
2. ***Kraus representation**: There exist matrices $K_i \in \mathcal{L}(A,B)$ with $\sum_i K_i^\dagger K_i = I_A$ such that*
$$\Phi(X) = \sum_i K_i X K_i^\dagger\,. \tag{22}$$
The K_i are called Kraus operators and you need at most $\dim(A)\dim(B)$ such operators to define the channel.
3. ***Choi representation**: Define the* Choi matrix *$C \in \mathcal{L}(AB)$ of the map Φ by*
$$C_{AB} = \sum_{ij} |i\rangle\langle j|_A \otimes \Phi(|i\rangle\langle j|_A) = \begin{pmatrix} \Phi(|0\rangle\langle 0|) & \Phi(|0\rangle\langle 1|) & \dots & \Phi(|0\rangle\langle d-1|) \\ \Phi(|1\rangle\langle 0|) & \Phi(|1\rangle\langle 1|) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \Phi(|d-1\rangle\langle 0|) & \dots & \dots & \Phi(|d-1\rangle\langle d-1|) \end{pmatrix}. \tag{23}$$
Then $C_{AB} \geq 0$ and $\mathrm{Tr}_B\left[C_{AB}\right] = I_A$.

Proof. We won't prove the result here (you can find a proof in many textbooks, e.g. Wilde's Quantum Shannon theory book or Watrous' Quantum information book). However we will note some interesting connections that arise in the proof. In particular how to derive Kraus operators from the Choi matrix.

The condition $C_{AB} \geq 0$ for the Choi matrix is equivalent to the channel being completely positive. The second condition $\mathrm{Tr}_B\left[C_{AB}\right] = I_A$ is equivalent to the channel being trace preserving. This actually establishes an isomorphism

between completely positive maps from $A \to B$ and positive semidefinite matrices acting on AB. This isomorphism is known as the Choi-Jamiolkowski isomorphism.

A set of Kraus operators for the channel can actually be derived directly from the Choi matrix as well. As $C_{AB} \geq 0$ we have by the spectral theorem $C_{AB} = \sum_i |u_i\rangle\langle u_i|$ where $|u_i\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ are the potentially unnormalized eigenvectors of C_{AB}. Then define the map $\text{vec}^{-1}(|i\rangle_A |j\rangle_B) = |j\rangle\langle i|$ which sends vectors on $\mathcal{H}_A \otimes \mathcal{H}_B$ to matrices in $\mathcal{L}(A, B)$. Then a set of Kraus operators for the channel is given by $K_i = \text{vec}^{-1}(|u_i\rangle)$. Note Kraus operators are not unique. □

Example 6 (Kraus and Choi characterizations). For simplicity let us consider only qubit channels.

1. Let $\mathcal{I}$ be the identity channel (for a qubit). Then we can write

$$\mathcal{I}(M) = M = IMI$$

so a set of Kraus operators for the channel are $\{I\}$. The Choi matrix for this channel is given by

$$C = \sum_{ij} |i\rangle\langle j| \otimes |i\rangle\langle j| = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

It can be checked that $C \geq 0$ and $\text{Tr}_B[C] = I$.

2. Let $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ be the Pauli Z operator and $p \in [0, 1]$. Then define the channel

$$\Phi(M) = (1-p)M + pZMZ\,.$$

this channel is sometimes known as a dephasing channel. A set of Kraus operators for this channel are $\{\sqrt{1-p}I, \sqrt{p}Z\}$. It has a Choi matrix

$$C = \begin{pmatrix} \Phi(|0\rangle\langle 0|) & \Phi(|0\rangle\langle 1|) \\ \Phi(|1\rangle\langle 0|) & \Phi(|1\rangle\langle 1|) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1-2p \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1-2p & 0 & 0 & 1 \end{pmatrix}. \tag{24}$$

By computing the eigenvalues and eigenvectors of C one can derive that $C = \sum_i |v_i\rangle\langle v_i|$ where

$$|v_0\rangle = \sqrt{1-p}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \sqrt{1-p}(|00\rangle + |11\rangle) \qquad \text{and} \qquad |v_1\rangle = \sqrt{p}\begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \sqrt{p}(-|00\rangle + |11\rangle) \tag{25}$$

By the earlier remark in the proof of Lemma 1 we should then have a set of Kraus operators $K_0 = \sqrt{1-p}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \sqrt{1-p}I$ and $K_1 = \sqrt{p}(-|0\rangle\langle 0| + |1\rangle\langle 1|) = -\sqrt{p}Z$. Note that this is different from the set of Kraus operators we obtained by inspection as now the Z has a -1 phase. However, the action is the same!

Returning to the initial example of the section it turns out that all quantum channels can be viewed as a unitary transformation on a larger system! In the same spirit as purifications from the previous section we have the following result.

Theorem 2 (Stinespring dilation)**.** *Let $\Phi : \mathcal{L}(A) \to \mathcal{L}(B)$ be a quantum channel. Then*

1. *There exists a system C of dimension no larger than $\dim(A)\dim(B)$ and an isometry $V : A \to BC$ such that*

$$\Phi(M) = \text{Tr}_C\left[VMV^\dagger\right]. \tag{26}$$

2. *There exists a system C of dimension no larger than $\dim(A)\dim(B)$ and a unitary $U \in \mathcal{L}(AC)$ such that*

$$\Phi(M) = \text{Tr}_C\left[U(M \otimes |0\rangle\langle 0|_C)U^\dagger\right]. \tag{27}$$

Proof. If Φ is a quantum channel then there exist Kraus operators $\{K_a\}_a$ such that

$$\Phi(M) = \sum_a K_a M K_a^\dagger .$$

Now consider the following operator $V \in \mathcal{L}(A, BC)$ defined as

$$V = \sum_a K_a \otimes |a\rangle_C$$

where $\{|a\rangle\}$ is some orthonormal basis of C. We have that V is an isometry as

$$V^\dagger V = \left(\sum_a K_a^\dagger \otimes \langle a|\right)\left(\sum_b K_b \otimes |b\rangle\right) = \sum_{ab} K_a^\dagger K_b \langle a||b\rangle = \sum_a K_a^\dagger K_a = I_A .$$

Moreover, we have

$$\mathrm{Tr}_C\left[VMV^\dagger\right] = \mathrm{Tr}_C\left[\sum_{ab} K_a M K_b^\dagger \otimes |a\rangle\langle b|\right] = \sum_{ab} K_a M K_b^\dagger \mathrm{Tr}\left[|a\rangle\langle b|\right] = \sum_a K_a M K_a^\dagger . \tag{28}$$

which proves the first result.

The second result follows from the construction of the first. The idea is to consider the operator $U_0 \in \mathcal{L}(CA)$ defined by

$$U_0 = \sum_a |a\rangle\langle 0| \otimes K_a = \begin{pmatrix} K_0 & 0 & \dots \\ K_1 & 0 & \dots \\ \vdots & \vdots & \dots \end{pmatrix} . \tag{29}$$

Note that $\sum_a K_a^\dagger K_a = I$ tells us that if we think of the first $\dim(B)$ columns of the matrix as vectors then they should form an orthonormal set. By extending this orthonormal set to an orthonormal basis (which we can always do) we can fill in the remaining entries of U_0 to find a unitary U that satisfies $U(|0\rangle_C \otimes \rho_A) = \sum_a |a\rangle \otimes K_a \rho_A$. Note we reordered C to make an argument in the block matrix form but the ordering of the systems doesn't matter hence the result readily follows from the same logic as above. □

The second result in the Stinespring dilation has a clear operational interpretation and connects to the initial example of the section. Any quantum channel can be viewed as a unitary evolution on a larger system followed by a partial trace. That is, our non-unitary dynamics can be viewed as ignorance of some global unitary evolution (interaction with the environment).

3 Measurements

Finally we arrive at measurements. Let's begin with a definition we've already seen.

Definition 4 (POVMs)**.** Let $\mathcal{H}$ be a Hilbert space. A POVM on $\mathcal{H}$ is a collection of operators $\{M_i\}_i$ with $M_i \in \mathcal{L}(\mathcal{H})$ such that $M_i \geq 0$ for all i and $\sum_i M_i = I$.

A special case of POVMs is when the operators are all projectors, i.e., $M_i^2 = M_i$, this is sometimes referred to as a PVM or projective measurement. For a quantum system in a state ρ the probability of receiving the outcome i when performing the measurement $\{M_i\}_i$ is given by the Born rule

$$\mathbb{P}(i) = \mathrm{Tr}\left[\rho M_i\right] . \tag{30}$$

Like the case for states and measurements we can recover our simpler formalism of projective measurements by dilating our Hilbert space. In particular, as the following theorem will show, the measurements modeled by POVMs can be equivalently thought of as a projective measurement on a larger space.

Theorem 3 (Naimark dilation)**.** *Let $\{M_i\}_i$ be a POVM on a system A and let $\rho_A \in \mathcal{D}(A)$. Then there exists a system B, an isometry $V : A \to AB$ and a PVM $\{P_i\}_i$ such that*

$$\mathrm{Tr}\left[\rho_A M_i\right] = \mathrm{Tr}\left[(V\rho_A V^\dagger)P_i\right] . \tag{31}$$

Proof. Like the proof of the Stinespring dilation we construct an explicit isometry. Let

$$V = \sum_a \sqrt{M_a} \otimes |a\rangle_B$$

where $\{|a\rangle\}_a$ is an orthonormal basis for B. Then V is an isometry as $V^\dagger V = I$. Now define the projective measurement $\{P_a\}$ on the joint space AB by

$$P_a = I \otimes |a\rangle\langle a|$$

one can check this is a PVM. Then

$$\begin{aligned}
\mathrm{Tr}\left[V \rho_A V^\dagger P_a\right] &= \mathrm{Tr}\left[\left(\sum_{bc} \sqrt{M_b}\rho_A\sqrt{M_c} \otimes |b\rangle\langle c|\right)(I \otimes |a\rangle\langle a|)\right] \\
&= \mathrm{Tr}\left[\sum_b \sqrt{M_b}\rho_A\sqrt{M_a} \otimes |b\rangle\langle a|\right] \\
&= \mathrm{Tr}\left[\sqrt{M_a}\rho_A\sqrt{M_a}\right] \\
&= \mathrm{Tr}\left[\rho_A M_a\right]
\end{aligned}$$

where on the third line we used the identity $\mathrm{Tr}\left[X_{AB}\right] = \mathrm{Tr}\left[\mathrm{Tr}_B\left[X_A B\right]\right]$ (see exercises) and on the fourth line we used the cyclicity of the trace. □

3.1 State update rules

In the case of projective measurements we also defined a state update rule, if a measurement $\{P_i\}_i$ is projective then after we receive outcome i the quantum state ρ becomes

$$\frac{P_i \rho P_i}{\mathrm{Tr}\left[\rho P_i\right]}. \tag{32}$$

However, for general POVMs the state update is not well-defined. A common generalization of the projective measurement update rule is the Lüder's update rule which says that after measuring a POVM $\{M_a\}_a$ and receiving the outcome a the state becomes

$$\frac{\sqrt{M_a}\rho\sqrt{M_a}}{\mathrm{Tr}\left[\rho M_a\right]}. \tag{33}$$

However, note this is not a unique generalization. In particular let $M_i = K_i^\dagger K_i$ then the update rule

$$\frac{K_a \rho K_a^\dagger}{\mathrm{Tr}\left[\rho M_a\right]} \tag{34}$$

would also make sense mathematically but there are infinitely many such choices of K_a. For instance $K_a = U_a\sqrt{M_a}$ for any unitaries U_a would fit. Operationally this corresponds to performing the Lüder's update rule and then applying a unitary evolution to the system depending on the measurement outcome. Long story short, there's no unique update rule (nor should there be in general). Nevertheless we can still consider all the possible update rules by viewing a measurement in the language of quantum channels.

Definition 5 (Quantum instruments)**.** Let A and B be quantum systems and let $\{\Phi_i\}_i$ be a collection of completely positive maps $\Phi_i : \mathcal{L}(A) \to \mathcal{L}(B)$ that are *trace non-increasing* (i.e., $\mathrm{Tr}\left[\Phi_i(\rho)\right] \leq \mathrm{Tr}\left[\rho\right]$ for all inputs ρ). The collection $\{\Phi_i\}_i$ is called a *quantum instrument* if $\sum_i \Phi_i$ is trace-preserving (i.e., $\mathrm{Tr}\left[(\sum_i \Phi_i)(\rho)\right] = \sum_i \mathrm{Tr}\left[\Phi_i(\rho)\right] = \mathrm{Tr}\left[\rho\right]$ for all inputs ρ).

A quantum instrument is another way to model a quantum measurement. If $\{\Phi_i\}_i$ is a quantum instrument then when measuring a state ρ we receive the outcome i with probability

$$\mathbb{P}(i) = \mathrm{Tr}\left[\Phi_i(\rho)\right].$$

and we define the state update rule to be after outcome i is received the state becomes

$$\frac{\Phi_i(\rho)}{\mathrm{Tr}\left[\Phi_i(\rho)\right]}.$$

As quantum channels capture the most general transformations of quantum systems, quantum instruments capture the most general way to think about a measurement. In particular the update rule captures scenarios wherein for example we apply arbitrary channels to our quantum system depending on the outcome of our measurement.

Remark 2. Note that quantum instruments capture a broader notion of post-measurement states. However, they do not introduce anything new in terms of what properties of a quantum system can actually be measured. More formally, for any instrument $\{\Phi_i\}$ there exists a POVM $\{M_i\}_i$ such that for all states ρ we have $\mathrm{Tr}\left[\Phi_i(\rho)\right] = \mathrm{Tr}\left[\rho M_i\right]$. To see this note that for a CPTNI map Φ_i we have a Kraus decomposition $\{K_{a,i}\}_a$ with the Kraus operators satisfying $\sum_a K_{a,i}^\dagger K_{a,i} \leq I$ as it is trace non-increasing. We can then define a POVM $M_i = \sum_a K_{a,i}^\dagger K_{a,i}$. This is indeed a POVM due to the fact that $\sum_i \Phi_i$ is a CP map so $I = \sum_{i,a} K_{a,i}^\dagger K_{a,i} = \sum_i M_i$. Moreover,

$$\mathrm{Tr}\left[\Phi_i(\rho)\right] = \mathrm{Tr}\left[\sum_a K_{a,i}\rho K_{a,i}^\dagger\right] = \mathrm{Tr}\left[\sum_a K_{a,i}^\dagger K_{a,i}\rho\right] = \mathrm{Tr}\left[M_i\rho\right].$$

[PB: Add converse remark (as exercise?). Any quantum channel implementing a POVM will have $\sum_a K_{a,i}^\dagger K_{a,i} = M_i$.]

3.2 Classical systems

Frequently we want to describe 'classical' random variables in the framework of quantum theory. Suppose we have a random variable X with d different outcomes which is distributed according to some probability distribution p_x. To embed this in the framework of quantum theory we first fix some orthonormal basis $\{|x\rangle\}_x$ (say the standard basis) and we define the state

$$\rho_X = \sum_x p_x |x\rangle\langle x| \tag{35}$$

which is diagonal in this classical basis. Note that all classical states encoded this way will necessarily commute. The probability of the random variable being x is then found by measuring the projective measurement $\{|x\rangle\langle x|\}_x$. This allows us to model, amongst other things, classical information theory within the wider context of quantum information theory and also to model hybrid systems where quantum systems are correlated with classical random variables.

Returning to the example at the beginning of these notes. Suppose we sample the random variable X and depending on it's outcome we prepare a system A in some state $|\psi_x\rangle$. Then this entire system can be modeled as a *classical-quantum* state (cq-state)

$$\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes |\psi_x\rangle\langle\psi_x|.$$

We can also view the measurement outcome as a classical random variable. In particular suppose we measure a system A with some POVM $\{M_x\}_x$ and we record the outcome in some classical system X. We could view this in the quantum instrument formalism using the CPTNI maps $\Phi_i : \mathcal{L}(A) \to \mathcal{L}(X)$ where

$$\Phi_x(\rho) = \mathrm{Tr}\left[\rho M_x\right] |x\rangle\langle x|. \tag{36}$$

If we don't condition on the outcome of the measurement, after measuring the state becomes the mixed state

$$\sum_x \Phi_x(\rho) = \sum_x \mathrm{Tr}\left[\rho M_x\right] |x\rangle\langle x| = \sum_x p_x |x\rangle\langle x|.$$

If we wanted to also retain the post-measurement state then we could have something more general like some CPTNI maps $\Lambda_i : \mathcal{L}(A) \to \mathcal{L}(B)$ that model the quantum instruments from before and then $\Phi_i : \mathcal{L}(A) \to \mathcal{L}(XB)$ where

$$\Phi_x(\rho_A) = |x\rangle\langle x| \otimes \Lambda_x(\rho_A)$$

such an instrument captures the previous notion of a post-measurement state however we also have that the outcome of the measurement is stored into some new classical system X.

4 Exercises

1. **Density matrices**: Write down the density matrices that represent the following probabilistic preparations.

   (a) We prepare a qubit system in the state $|0\rangle$ with probability $1/2$ and $|+\rangle$ with probability $1/2$.

   (b) We prepare a two-qubit system in the state $\frac{|00\rangle+|11\rangle}{\sqrt{2}}$ with probability p and with probability $1-p$ we prepare it in the state $\frac{|00\rangle-|11\rangle}{\sqrt{2}}$.

   This preparation corresponds to probabilistically preparing entangled states. Is the resulting state entangled for all values of p?

   (c) Write down the two-qubit density matrix $\rho_{AB} = |\psi\rangle\langle\psi|$ where

   $$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|11\rangle\,.$$

   Compute also the marginal states ρ_A and ρ_B.

   (d) Show that a density matrix ρ is pure iff $\mathrm{Tr}\left[\rho^2\right] = 1$.

   (e) We motivated density matrices as a probabilistic mixture of state preparations, i.e., $\rho = \sum_x p_x|\psi_x\rangle\langle\psi_x|$. Give an example of two different preparations $\{p_x, |\psi_x\rangle\}_x$, $\{q_y, |\phi_y\rangle\}_x$ that lead to the same density matrix. That is

   $$\rho = \sum_x p_x|\psi_x\rangle\langle\psi_x| = \sum_y q_y|\phi_y\rangle\langle\phi_y|\,.$$

   (f) Prove that a density matrix has a unique preparation interpretation iff the density matrix is pure (preparations whose states differ by a global phase are considered the same).

2. **Trace identities**: Prove the following useful identities concerning the trace and partial trace maps.

   (a) Show that the trace is cyclic. That is for any two matrices X and Y we have

   $$\mathrm{Tr}\left[XY\right] = \mathrm{Tr}\left[YX\right]\,.$$

   (b) Recall we defined the trace of a square matrix X as the sum of its diagonal entries, i.e., $\mathrm{Tr}\left[X\right] = \sum_i \langle i|X|i\rangle$ where $\{|i\rangle\}_i$ is the standard orthonormal basis in which we write the matrix. Show that this definition is independent of the orthonormal basis we choose. that is

   $$\mathrm{Tr}\left[X\right] = \sum_i \langle v_i|X|v_i\rangle$$

   for any orthonormal basis $\{|v_i\rangle\}_i$.

   (c) Suppose we have a bipartite system AB and $X_{AB} \in \mathcal{L}(AB)$. Show that

   $$\mathrm{Tr}\left[X_{AB}\right] = \mathrm{Tr}\left[\mathrm{Tr}_B\left[X_{AB}\right]\right]\,.$$

   That is, to compute the trace we can first take a partial trace and then the trace of the remaining system.

   (d) Show that for any $X_{AB} \in \mathcal{L}(AB)$, $M, N \in \mathcal{L}(A)$ we have

   $$\mathrm{Tr}_B\left[(M \otimes I_B)X_{AB}(N \otimes I_B)\right] = M\mathrm{Tr}_B\left[X_{AB}\right]N\,.$$

   (e) Finally show that the partial trace is partially cyclic on the system being traced out. If $X_{AB} \in \mathcal{L}(AB)$, $M, N \in \mathcal{L}(A)$ then

   $$\mathrm{Tr}_A\left[(M \otimes I)X_{AB}(N \otimes I)\right] = \mathrm{Tr}_A\left[(NM \otimes I)X_{AB}(I \otimes I)\right]$$

3. **Positive semidefinite matrices**

   (a) Let $P \in \mathcal{L}(A)$ be a PSD matrix and let $X \in \mathcal{L}(A, B)$ be an arbitrary matrix. Show that

   $$XPX^\dagger \geq 0 \tag{37}$$

(b) Let $P, Q \in \mathcal{L}(A)$ be PSD matrices. Show that

$$P + Q \geq 0\,. \tag{38}$$

(c) (Jordon-Hanh decomposition) Let $X \in \mathcal{L}(A)$ be a Hermitian matrix. Show that there exist positive semidefinite matrices $P, Q \in \mathcal{L}(A)$ such that

$$X = P - Q \tag{39}$$

and $PQ = QP = 0$. (Hint: consider the spectral decomposition of X.)

(d) Let $P, Q \in \mathcal{L}(A)$ be PSD. Show that

$$\mathrm{Tr}\left[PQ\right] \geq 0\,.$$

4. **Qubits and the Bloch ball**: Let's consider density matrices for a qubit system.

(a) Show that any qubit state can be written in the form

$$\rho = \frac{I + r_x X + r_y Y + r_z Z}{2} \tag{40}$$

where $r_x, r_y, r_z \in \mathbb{R}$ satisfy $r_x^2 + r_y^2 + r_z^2 \leq 1$ and X, Y, Z are the three Pauli matrices.

(b) Show that for pure states we have $r_x^2 + r_y^2 + r_z^2 = 1$. I.e., pure states correspond to the surface of the Bloch ball.

(c) Let $\{P_0, P_1\}$ be a rank one projective qubit measurement i.e., $P_0 = |\psi\rangle\langle\psi|$ for some qubit state $|\psi\rangle$. Show that for such a measurement, P_0 and P_1 correspond to opposite points on the surface of the Bloch ball.

(d) Let ρ be a qubit state with a Bloch vector $r = (r_x, r_y, r_z)$ and let P_0 be a projective measurement element with a Bloch vector $s = (s_x, s_y, s_z)$. Show that

$$\mathrm{Tr}\left[\rho P_0\right] = \frac{1}{2} + \frac{1}{2} r \cdot s \tag{41}$$

where $r \cdot s = r_x s_x + r_y s_y + r_z s_z$. This gives a geometric interpretation to qubit measurement probabilities.

(e) Let $\rho = \frac{1}{2}(I + r_x X + r_y Y + r_z Z)$ and let $\sigma = \frac{1}{2}(I + s_x X + s_y Y + s_z Z)$. Show that

$$\|\rho - \sigma\|_1 = \sqrt{(r_x - s_x)^2 + (r_y - s_y)^2 + (r_z - s_z)^2} \tag{42}$$

where $\|\cdot\|_1$ is the trace norm defined in (52).

5. **Entanglement and the Schmidt decomposition**: Consider a pure bipartite state $|\psi\rangle_{AB} = \sum_i \sqrt{\lambda_i}|ii\rangle$ written in its Schmidt decomposition.

(a) Compute $\rho_A = \mathrm{Tr}_B\left[|\psi\rangle\langle\psi|\right]$ and $\rho_B = \mathrm{Tr}_A\left[|\psi\rangle\langle\psi|\right]$. What can you say about their eigenvalues and eigenvectors?

(b) Show that the state $|\psi\rangle_{AB}$ is entangled iff its Schmidt rank is greater than 1.

6. **A particular purification** Suppose we have a d-dimensional state $\rho_A \in \mathcal{D}(A)$. Let B be another d-dimensional system and define $|\Gamma\rangle_{AB} = \sum_{i=0}^{d-1} |ii\rangle$ to be the non-normalized maximally entangled state on AB. Show that

$$|\psi\rangle_{AB} := (\sqrt{\rho_A} \otimes I_B)|\Gamma\rangle$$

is a purification of ρ_A.

Use the above to purify your favourite qubit state. Compare this with the purification you get by constructing a Schmidt decomposition of a larger system, are they the same?

7. **Classical systems**: Consider a joint probability distribution of two classical binary random variables X and Y.

$$P_{XY}(x, y) = \begin{cases} 1/3 & \text{for } (x, y) = (0, 0) \\ 1/3 & \text{for } (x, y) = (0, 1) \\ 0 & \text{for } (x, y) = (1, 0) \\ 1/3 & \text{for } (x, y) = (1, 1) \end{cases} \tag{43}$$

(a) Compute the marginal distribution $P_X(x)$.

(b) How would you represent the joint random variable XY as a bipartite quantum state?

(c) Trace out system Y from part (b). Does the resulting quantum state ρ_X agree with the marginal distribution in part (a)?

8. **Quantum channels**:

(a) Suppose you have a qubit system A and you decide to transform it in the following manner: with probability 1/4 you do nothing (identity), with probability 1/4 you apply a Pauli X evolution, with probability 1/4 you apply a Pauli Y evolution and with probability 1/4 you apply a Pauli Z. Write down the qubit channel $\mathcal{E}$ that corresponds to this overall evolution.

Additionally show that $\mathcal{E}(\rho) = I/2$ for any qubit state ρ.

(b) Consider the replacement channel

$$\mathcal{E}_\sigma(\rho) := \mathrm{Tr}\left[\rho\right]\sigma$$

That takes any quantum state and replaces it with the state σ. Prove that this is a quantum channel.

(c) Consider the qubit depolarizing channel

$$\mathcal{E}_p(\rho) := (1-p)\rho + \mathrm{Tr}\left[\rho\right] I/2$$

find a Kraus representation for the channel. Interpret the action of this channel geometrically using the Bloch ball.

(d) Let $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ and $\mathcal{F} : \mathcal{L}(B) \to \mathcal{L}(C)$ be two quantum channels. Show that the concatenation of the two channels $(\mathcal{F} \circ \mathcal{E})$ is again a quantum channel where

$$(\mathcal{F} \circ \mathcal{E})(\rho) = \mathcal{F}(\mathcal{E}(\rho))\,. \tag{44}$$

9. **The transpose map**: Consider the following map $\mathcal{T} : \mathcal{L}(A) \to \mathcal{L}(A)$ defined with respect to the computational basis $\{|i\rangle\}_i$ as

$$\mathcal{T}(|i\rangle\langle j|) = |j\rangle\langle i|\,.$$

In other words $\mathcal{T}(X) = X^T$ where X^T denotes the transpose of X when X is written in the computational basis.

(a) Show that $\mathcal{T}$ is **not** a quantum channel. Hint: Show it is not completely positive by applying it to one of the subsystems of the two-qubit state $|\psi\rangle_{AB} = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$. That is, consider the output of

$$(\mathcal{T} \otimes \mathcal{I}_B)(|\psi\rangle\langle\psi|)$$

where $\mathcal{I}_B$ is the identity map on system B.

(b) In the previous part we showed that the transpose map is not a quantum channel by showing that the *partial transpose* $(\mathcal{T} \otimes \mathcal{I})$ when acting on an entangled state is not a positive map. It turns out positive but not completely positive maps are intricately related to entanglement.

Show that if ρ_{AB} is separable then

$$(\mathcal{T} \otimes \mathcal{I})(\rho_{AB}) \geq 0\,. \tag{45}$$

that is the partial transpose is always positive for separable states.

(c) By applying the partial transpose map $(\mathcal{T} \otimes \mathcal{I})$ on the two-qubit states

$$\rho_{AB} = (1-p)|\Phi^+\rangle\langle\Phi^+| + pI/4\,. \tag{46}$$

show that they are entangled for all $p < 2/3$ where $|\Phi^+\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$.

It turns out the partial transpose exactly captures entanglement in two-qubit systems. In particular, the positive partial transpose (PPT) criterion states that a two-qubit state ρ_{AB} is *separable* iff $(\mathcal{T} \otimes \mathcal{I})(\rho_{AB}) \geq 0$. This gives a very simple test to check whether a two-qubit state is entangled. Thus the state is part (c) is indeed separable for $p > 2/3$.

10. **Dilating dilations**: Suppose you have a quantum state $\rho_A \in \mathcal{D}(A)$ and a POVM $\{M_a\}_a$. Show that there always exists some pure state $|\psi\rangle$ and projective measurement $\{P_a\}_a$ on a potentially larger system such that

$$\mathrm{Tr}\left[\rho M_a\right] = \mathrm{Tr}\left[|\psi\rangle\langle\psi| P_a\right] . \tag{47}$$

11. **State discrimination**: A fundamental task in quantum information theory is state discrimination. Suppose you are told a system A is prepared in the state $\rho_0 \in \mathcal{D}(A)$ with probability $\lambda \in [0,1]$ and in the state $\rho_1 \in \mathcal{D}(A)$ with probability $1-\lambda$. You don't know which state was prepared so from your perspective the state of the system is

$$\lambda\rho_0 + (1-\lambda)\rho_1 .$$

Your goal is to measure the quantum system to try to determine which state was actually prepared. To this end you define a POVM $\{M_0, M_1\}$ which you will measure and if you receive outcome 0 you will guess that the system was prepared in state ρ_0 whereas if you receive outcome 1 you will guess that the state was prepared in system ρ_1. Overall the probability you guess correctly is given by

$$\begin{aligned}\mathbb{P}[\text{Guess correctly}] &= \mathbb{P}[\text{State is } \rho_0]\mathbb{P}[\text{Measure 0}|\text{State is } \rho_0] + \mathbb{P}[\text{State is } \rho_1]\mathbb{P}[\text{Measure 1}|\text{State is } \rho_1] \\ &= \lambda \mathrm{Tr}\left[\rho_0 M_0\right] + (1-\lambda)\mathrm{Tr}\left[\rho_1 M_1\right]\end{aligned}$$

(a) Show that

$$\mathbb{P}[\text{Guess correctly}] = \frac{1}{2} + \frac{1}{2}\mathrm{Tr}\left[(M_0 - M_1)(\lambda\rho_0 - (1-\lambda)\rho_1)\right]$$

(b) Using Hölder's inequality (see Theorem 4) or otherwise show that

$$\mathbb{P}[\text{Guess correctly}] \leq \frac{1}{2} + \frac{1}{2}\|\lambda\rho_0 - (1-\lambda)\rho_1\|_1$$

(c) By considering the Jordon-Hanh decomposition (see question 1c) of the operator $\lambda\rho_0 - (1-\lambda)\rho_1$ show that there exists a projective measurement $\{P_0, P_1\}$ such that

$$\mathbb{P}[\text{Guess correctly}] = \lambda \mathrm{Tr}\left[\rho_0 P_0\right] + (1-\lambda)\mathrm{Tr}\left[\rho_1 P_1\right] = \frac{1}{2} + \frac{1}{2}\|\lambda\rho_0 - (1-\lambda)\rho_1\|_1$$

This result, known as the *Holevo-Helstrom* theorem, establishes an operational meaning to the trace distance as characterizing how well one can distinguish between two quantum states.

12. **State discrimination example**: Suppose I prepare a qubit system in the state $|0\rangle$ with probability $1/2$ and otherwise I prepare it in the state $|-\rangle$ with probability $1/2$. Your task is to guess which state I prepared by measuring the qubit.

(a) By choosing a qubit measurement from the family of measurements

$$M_0 = \frac{I + \cos(a)Z + \sin(a)X}{2} \tag{48}$$

where $a \in [-\pi, \pi]$ and $M_1 = I - M_0$. Find the optimal probability of guessing correctly which state was prepared.

(b) Interpret the states and the optimal measurement geometrically using the Bloch ball.

(c) Compute

$$\frac{1}{2} + \frac{1}{2}\|\frac{1}{2}|0\rangle\langle 0| - \frac{1}{2}|-\rangle\langle -|\|_1 \tag{49}$$

and show it agrees with your answer from part (a).

13. **Unambiguous state discrimination** Suppose we want to distinguish two states but we never want to make a mistake. That is, we never want to say '0' when the state was ρ_1 or say '1' when the state was ρ_0. By allowing ourselves an additional output 'unsure' we can try to still maximize the probability we are successful but now without ever making a misidentification.

Using the same setup as the previous question. Define a 3-outcome qubit POVM $\{M_0, M_1, M_{\text{abort}}\}$. Such that

$$\mathbb{P}[\text{Guess correct}] > 0$$

and

$$\text{Tr}\left[\rho_0 M_1\right] = 0 = \text{Tr}\left[\rho_1 M_0\right]$$

i.e., whenever we guess 0/1 we never make a mistake.

14. **Trace distance contractivity** Let $\rho, \sigma \in \mathcal{D}(A)$.

    (a) Show that for any positive semidefinite operator P

    $$\|P\|_1 = \text{Tr}\left[P\right] .$$

    (b) Prove that for states ρ and σ we have

    $$\|\rho - \sigma\|_1 = 2 \max_P \{\text{Tr}\left[P(\rho - \sigma)\right] : 0 \leq P \leq I\} . \tag{50}$$

    (c) Let $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ be a quantum channel. Show that

    $$\|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1 \leq \|\rho - \sigma\|_1 . \tag{51}$$

    (d) Comment on what this result says in relation to the problem of distinguishing between the two states ρ and σ.

A Distances between operators

It is often useful to say that one state is close to another or one measurement is far from another. In order to quantify 'closeness' we use norms. A norm is just a map that gives a notion of length to an object in some vector space. In general it is a map $\|\cdot\| : V \to \mathbb{R}_+$ where V is some vector space that satisfies

1. $\|x\| \geq 0$ and $\|x\| = 0 \iff x = 0$
2. $\|\alpha x\| = |\alpha| \|x\|$ for any scalars α.
3. $\|x + y\| \leq \|x\| + \|y\|$

Given a norm we can then define a notion of distance between two objects in the vector space via $d(x, y) := \|x - y\|$. By the properties above you see that this distance is 0 iff $x = y$ and it also satisfies the other properties of distance measures known as *metrics*.

We will occasionally need some matrix norms. Given a matrix $X \in \mathcal{L}(A)$ we define:

1. The *trace norm*

   $$\|X\|_1 := \text{Tr}\left[\sqrt{X^\dagger X}\right] \tag{52}$$

2. The *Frobenius/Euclidean* norm

   $$\|X\|_2 := \text{Tr}\left[X^\dagger X\right]^{1/2} \tag{53}$$

3. The *infinity/operator* norm

   $$\|X\|_\infty := \max\{\|X|v\rangle\| : |v\rangle \in A,\ \||v\rangle\| \leq 1\} \tag{54}$$

   where $\||w\rangle\| := \sqrt{\langle w||w\rangle}$ is the standard Euclidean norm for vectors. Note the operator norm is also equal to the largest singular value of X.

The trace norm and the operator norm satisfy a duality. In particular we have

$$\|X\|_1 := \max\{|\text{Tr}\left[XY\right]| : \|Y\|_\infty \leq 1\} \tag{55}$$

and

$$\|X\|_\infty := \max\{|\text{Tr}\left[XY\right]| : \|Y\|_1 \leq 1\}. \tag{56}$$

This then leads us to the following useful result which can be seen as different kind of Cauchy-Schwarz inequality.

Theorem 4 (Hölder inequality)**.** *Let $X, Y \in \mathcal{L}(A)$. Then*

$$|\text{Tr}\left[XY\right]| \leq \|X\|_1 \|Y\|_\infty . \tag{57}$$

B Spectral theorem and matrix functions

Let $M \in \mathcal{L}(A)$ be a normal matrix, i.e., $M^\dagger M = MM^\dagger$. Then the spectral theorem says that there exists a unitary matrix U and a diagonal matrix D, whose entries are the eigenvalues of M, such that

$$M = UDU^\dagger \,. \tag{58}$$

Equivalently, there exists an orthonormal basis $\{|u_i\rangle\}_i$ of A such that

$$M = \sum_i \lambda_i |u_i\rangle\langle u_i|$$

where $\lambda_i \in \mathbb{C}$ are the eigenvalues of M and $|u_i\rangle$ are the corresponding eigenvectors.

Note that for positive semidefinite matrices the eigenvalues are all non-negative ($\lambda_i \geq 0$), for Hermitian matrices the eigenvalues are always real ($\lambda_i \in \mathbb{R}$) and for unitary matrices the eigenvalues have unit modulus ($|\lambda_i| = 1$).

Matrix functions
We can use the spectral theorem to lift any continuous function $f : \mathbb{C} \to \mathbb{C}$ to normal matrices (as long as it is defined on the spectrum of the matrix). In particular we define this lifted function by

$$f(M) = \sum_i f(\lambda_i)|u_i\rangle\langle u_i|. \tag{59}$$

Example 7.

1. If M is positive semidefinite then there is a unique positive semidefinite square root. Define $f : \mathbb{R}_+ \to \mathbb{R}_+$ by $f(x) = \sqrt{x}$. Then
$$M^{1/2} = \sum_i \sqrt{\lambda_i}|u_i\rangle\langle u_i|. \tag{60}$$
We see that $M^{1/2} \geq 0$ as its eigenvalues are all nonnegative and $M^{1/2}M^{1/2} = M$.

2. If M is positive definite then we define
$$M \log(M) = \sum_i \lambda_i \log(\lambda_i)|u_i\rangle\langle u_i| \,. \tag{61}$$
If M is positive semidefinite (can have eigenvalues equal to 0) then we define
$$M \log(M) = \sum_{i:\lambda_i > 0} \lambda_i \log(\lambda_i)|u_i\rangle\langle u_i| \tag{62}$$
where we have implicitly defined $0 \log 0 = 0$ which is justified by $\lim_{x\to 0} x \log x = 0$.