CS 766/QIC 820 Theory of Quantum Information (Fall 2011)

Lecture 3: States, measurements, and channels

We begin our discussion of quantum information in this lecture, starting with an overview of three mathematical objects that provide a basic foundation for the theory: states, measurements, and channels. We will also begin to discuss important notions connected with these objects, and will continue with this discussion in subsequent lectures.

3.1 Overview of states, measurements, and channels

The theory of quantum information is concerned with properties of abstract, idealized physical systems that will be called *registers* throughout this course. In particular, one defines the notions of *states* of registers; of *measurements* of registers, which produce classical information concerning their states; and of *channels*, which transform states of one register into states of another. Taken together, these definitions provide the basic model with which quantum information theory is concerned.

3.1.1 Registers

The term *register* is intended to be suggestive of a component inside a computer in which some finite amount of data can be stored and manipulated. While this is a reasonable picture to keep in mind, it should be understood that any physical system in which a finite amount of data may be stored, and whose state may change over time, could be modeled as a register. Examples include the entire memory of a computer, or a collection of computers, or any medium used for the transmission of information from one source to another. At an intuitive level, what is most important is that registers are viewed as a physical objects, or parts of a physical objects, that store information.

It is not difficult to formulate a precise mathematical definition of registers, but we will not take the time to do this in this course. It will suffice for our needs to state two simple assumptions about registers:

1. Every register has a unique name that distinguishes it from other registers.
2. Every register has associated to it a finite and nonempty set of *classical states*.

Typical names for registers in these notes are capital letters written in a *sans serif* font, such as X, Y, and Z, as well as subscripted variants of these names like $\mathsf{X}_1, \ldots, \mathsf{X}_n$, Y_A, and Y_B. In every situation we will encounter in this course, there will be a finite (but not necessarily bounded) number of registers under consideration.

There may be legitimate reasons, both mathematical and physical, to object to the assumption that registers have specified classical state sets associated to them. In essence, this assumption amounts to the selection of a preferred basis from which to develop the theory, as opposed to opting for a basis-independent theory. From a computational or information processing point of view, however, it is quite natural to assume the existence of a preferred basis, and little (or

perhaps nothing) is lost by making this assumption in the finite-dimensional setting in which we will work.

Suppose that X is a register whose classical state set is Σ. We then associate the complex Euclidean space $\mathcal{X} = \mathbb{C}^{\Sigma}$ with the register X. States, measurements, and channels connected with X will then be described in linear-algebraic terms that refer to this space. As a general convention, we will always name the complex Euclidean space associated with a given register with the same letter as the register, but in a scripted font rather than a *sans serif* font. For instance, the complex Euclidean spaces associated with registers Y_j and Z_A are denoted $\mathcal{Y}_j$ and $\mathcal{Z}_A$, respectively. This is done throughout these notes without explicit mention.

For any finite sequence $\mathsf{X}_1, \ldots, \mathsf{X}_n$ of distinct registers, we may view that the n-tuple

$$\mathsf{Y} = (\mathsf{X}_1, \ldots, \mathsf{X}_n)$$

is itself a register. Assuming that the classical state sets of the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are given by $\Sigma_1, \ldots, \Sigma_n$, respectively, we naturally take the classical state set of Y to be $\Sigma_1 \times \cdots \times \Sigma_n$. The complex Euclidean space associated with Y is therefore

$$\mathcal{Y} = \mathbb{C}^{\Sigma_1 \times \cdots \times \Sigma_n} = \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n.$$

3.1.2 States

A *quantum state* (or simply a *state*) of a register X is an element of the set

$$\mathrm{D}(\mathcal{X}) = \{\rho \in \mathrm{Pos}(\mathcal{X}) : \mathrm{Tr}(\rho) = 1\}$$

of *density operators* on $\mathcal{X}$. Every element of this set is to be considered a valid state of X.

A state $\rho \in \mathrm{D}(\mathcal{X})$ is said to be *pure* if it takes the form

$$\rho = uu^*$$

for some vector $u \in \mathcal{X}$. (Given that $\mathrm{Tr}(uu^*) = \|u\|^2$, any such vector is necessarily a unit vector.) An equivalent condition is that $\mathrm{rank}(\rho) = 1$. The term *mixed state* is sometimes used to refer to a state that is either not pure or not necessarily pure, but we will generally not use this terminology: it will be our default assumption that states are not necessarily pure, provided it has not been explicitly stated otherwise.

Three simple observations (the first two of which were mentioned briefly in the previous lecture) about the set of states $\mathrm{D}(\mathcal{X})$ of a register X are as follows.

1. The set $\mathrm{D}(\mathcal{X})$ is *convex*: if $\rho, \sigma \in \mathrm{D}(\mathcal{X})$ and $\lambda \in [0,1]$, then $\lambda\rho + (1-\lambda)\sigma \in \mathrm{D}(\mathcal{X})$.
2. The *extreme points* of $\mathrm{D}(\mathcal{X})$ are precisely the pure states uu^* for $u \in \mathcal{X}$ ranging over all unit vectors.
3. The set $\mathrm{D}(\mathcal{X})$ is *compact*.

One way to argue that $\mathrm{D}(\mathcal{X})$ is compact, starting from the assumption that the unit sphere $\mathcal{S} = \{u \in \mathcal{X} : \|u\| = 1\}$ in $\mathcal{X}$ is compact, is as follows. We first note that the function $f : \mathcal{S} \to \mathrm{D}(\mathcal{X}) : u \mapsto uu^*$ is continuous, so the set of pure states $f(\mathcal{S}) = \{uu^* : u \in \mathcal{X}, \|u\| = 1\}$ is compact (as continuous functions always map compact sets to compact sets). By the spectral theorem it is clear that $\mathrm{D}(\mathcal{X})$ is the convex hull of this set: $\mathrm{D}(\mathcal{X}) = \mathrm{conv}\{uu^* : u \in \mathcal{X}, \|u\| = 1\}$. As the convex hull of every compact set is compact, it follows that $\mathrm{D}(\mathcal{X})$ is compact.

Let $\mathsf{X}_1, \ldots, \mathsf{X}_n$ be distinct registers, and let Y be the register formed by viewing these n registers as a single, compound register: $\mathsf{Y} = (\mathsf{X}_1, \ldots, \mathsf{X}_n)$. A state of Y taking the form

$$\rho_1 \otimes \cdots \otimes \rho_n \in \mathrm{D}\left(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n\right),$$

for density operators $\rho_1 \in \mathrm{D}\left(\mathcal{X}_1\right), \ldots, \rho_n \in \mathrm{D}\left(\mathcal{X}_n\right)$, is said to be a *product state*. It represents the situation that $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are *independent*, or that their states are independent, at a particular moment. If the state of Y cannot be expressed as product state, it is said that $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are *correlated*. This includes the possibility that $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are *entangled*, which is a phenomenon that we will discuss in detail later in the course. Registers can, however, be correlated without being entangled.

3.1.3 Measurements

A *measurement* of a register X (or a measurement on a complex Euclidean space $\mathcal{X}$) is a function of the form

$$\mu : \Gamma \to \mathrm{Pos}\left(\mathcal{X}\right),$$

where Γ is a finite, nonempty set of *measurement outcomes*. To be considered a valid measurement, such a function must satisfy the constraint

$$\sum_{a \in \Gamma} \mu(a) = \mathbb{1}_{\mathcal{X}}.$$

It is common that one identifies the measurement μ with the collection of operators $\{P_a : a \in \Gamma\}$, where $P_a = \mu(a)$ for each $a \in \Gamma$. Each operator P_a is called the *measurement operator* associated with the outcome $a \in \Gamma$.

When a measurement of the form $\mu : \Gamma \to \mathrm{Pos}\left(\mathcal{X}\right)$ is applied to a register X whose state is $\rho \in \mathrm{D}\left(\mathcal{X}\right)$, two things happen:

1. An element of Γ is randomly selected as the outcome of the measurement. The probability associated with each possible outcome $a \in \Gamma$ is given by

$$p(a) = \langle \mu(a), \rho \rangle .$$

2. The register X ceases to exist.

This definition of measurements guarantees that the vector $p \in \mathbb{R}^{\Gamma}$ of outcome probabilities will indeed be a probability vector, for every choice of $\rho \in \mathrm{D}\left(\mathcal{X}\right)$. In particular, each $p(a)$ is a nonnegative real number because the inner product of two positive semidefinite operators is necessarily a nonnegative real number, and the probabilities sum to 1 due to the constraint $\sum_{a \in \Gamma} \mu(a) = \mathbb{1}_{\mathcal{X}}$. In more detail,

$$\sum_{a \in \Gamma} p(a) = \sum_{a \in \Gamma} \langle \mu(a), \rho \rangle = \langle \mathbb{1}_{\mathcal{X}}, \rho \rangle = \mathrm{Tr}(\rho) = 1.$$

It can be sown that *every* linear function that maps $\mathrm{D}\left(\mathcal{X}\right)$ to the set of probability vectors in $\mathbb{R}^{\Gamma}$ is induced by some measurement μ as we have just discussed. It is therefore not an arbitrary choice to define measurements as they are defined, but rather a reflection of the idea that every linear function mapping density operators to probability vectors is to be considered a valid measurement.

Note that the assumption that the register that is measured ceases to exist is not necessarily standard: you will find definitions of measurements in books and papers that do not make this assumption, and provide a description of the state that is left in the register after the measurement. No generality is lost, however, in making the assumption that registers cease to exist upon being measured. This is because standard notions of *nondestructive measurements*, which specify the states of registers after they are measured, can be described by composing channels with measurements (as we have defined them).

A measurement of the form

$$\mu : \Gamma_1 \times \cdots \times \Gamma_n \to \mathrm{Pos}\left(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n\right),$$

defined on a register of the form $\mathsf{Y} = (\mathsf{X}_1, \ldots, \mathsf{X}_n)$, is called a *product measurement* if there exist measurements

$$\begin{gathered} \mu_1 : \Gamma_1 \to \mathrm{Pos}\left(\mathcal{X}_1\right), \\ \vdots \\ \mu_n : \Gamma_n \to \mathrm{Pos}\left(\mathcal{X}_n\right) \end{gathered}$$

such that

$$\mu(a_1, \ldots, a_n) = \mu_1(a_1) \otimes \cdots \otimes \mu_n(a_n)$$

for all $(a_1, \ldots, a_n) \in \Gamma_1 \times \cdots \times \Gamma_n$. Similar to the interpretation of a product state, a product measurement describes the situation in which the measurements $\mu_1, \ldots, \mu_n$ are independently applied to registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$, and the n-tuple of measurement outcomes is interpreted as a single measurement outcome of the compound measurement μ.

A *projective measurement* $\mu : \Gamma \to \mathrm{Pos}\left(\mathcal{X}\right)$ is one for which $\mu(a)$ is a projection operator for each $a \in \Gamma$. The only way this can happen in the presence of the constraint $\sum_{a \in \Gamma} \mu(a) = \mathbb{1}_{\mathcal{X}}$ is for the measurement operators $\{P_a : a \in \Gamma\}$ to be projections onto mutually orthogonal subspaces of $\mathcal{X}$. When $\{x_a : a \in \Sigma\}$ is an orthonormal basis of $\mathcal{X}$, the projective measurement

$$\mu : \Sigma \to \mathrm{Pos}\left(\mathcal{X}\right) : a \mapsto x_a x_a^*$$

is referred to as the measurement with respect to the basis $\{x_a : a \in \Sigma\}$.

3.1.4 Channels

Quantum channels represent idealized physical operations that transform states of one register into states of another. In mathematical terms, a *quantum channel* from a register X to a register Y is a linear mapping of the form

$$\Phi : \mathrm{L}\left(\mathcal{X}\right) \to \mathrm{L}\left(\mathcal{Y}\right)$$

that satisfies two restrictions:

1. Φ must be *trace-preserving*, and
2. Φ must be *completely positive*.

These restrictions will be explained shortly.

When a quantum channel from X to Y is applied to X, it is to be viewed that the register X ceases to exist, having been replaced by or transformed into the register Y. The state of Y is determined by applying the mapping Φ to the state $\rho \in \mathrm{D}\left(\mathcal{X}\right)$ of X, yielding $\Phi(\rho) \in \mathrm{D}\left(\mathcal{Y}\right)$.

There is nothing that precludes the choice that $\mathsf{X} = \mathsf{Y}$, and in this case one simply views that the state of the register X has been changed according to the mapping $\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{X})$. A simple example of a channel of this form is the *identity channel* $\mathbb{1}_{\mathrm{L}(\mathcal{X})}$, which leaves each $X \in \mathrm{L}(\mathcal{X})$ unchanged. Intuitively speaking, this channel represents an ideal communication channel or a perfect component in a quantum computer memory, which causes no modification of the register it acts upon.

Along the same lines as states and measurements, tensor products of channels represent independently applied channels, collectively viewed as a single channel. More specifically, if $\mathsf{X}_1, \ldots, \mathsf{X}_n$ and $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ are registers, and

$$\begin{gathered} \Phi_1 : \mathrm{L}(\mathcal{X}_1) \to \mathrm{L}(\mathcal{Y}_1) \\ \vdots \\ \Phi_n : \mathrm{L}(\mathcal{X}_n) \to \mathrm{L}(\mathcal{Y}_n) \end{gathered}$$

are channels, the channel

$$\Phi_1 \otimes \cdots \otimes \Phi_n : \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n) \to \mathrm{L}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n)$$

is said to be a *product channel*. It is the channel that represents the action of channels $\Phi_1, \ldots, \Phi_n$ being independently applied to $\mathsf{X}_1, \ldots, \mathsf{X}_n$.

Now let us return to the restrictions of trace preservation and complete positivity mentioned in the definition of channels. Obviously, if we wish to consider that the output $\Phi(\rho)$ of a given channel $\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y})$ is a valid state of Y for every possible state $\rho \in \mathrm{D}(\mathcal{X})$ of X, it must hold that Φ maps density operators to density operators. What is more, this must be so for tensor products of channels: it must hold that

$$(\Phi_1 \otimes \cdots \otimes \Phi_n)(\rho) \in \mathrm{D}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n)$$

for every choice of $\rho \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$, given any choice of channels $\Phi_1, \ldots, \Phi_n$ transforming registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ into registers $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$. In addition, we make the assumption that the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}$ is a valid channel for every register Z.

In particular, for every legitimate channel $\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y})$, it must hold that $\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}$ is also a legitimate channel, for every choice of a register Z. Thus,

$$(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho) \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z})$$

for every choice of $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$. This is equivalent to the two conditions stated before: it must hold that Φ is *completely positive*, which means that $(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(P) \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Z})$ for every $P \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Z})$, and Φ must preserve trace: $\mathrm{Tr}(\Phi(X)) = \mathrm{Tr}(X)$ for every $X \in \mathrm{L}(\mathcal{X})$.

Once we have imposed the condition of complete positivity on channels, it is not difficult to see that any tensor product $\Phi_1 \otimes \cdots \otimes \Phi_n$ of such channels will also map density operators to density operators. We may view tensor products like this as a composition of the channels $\Phi_1, \ldots, \Phi_n$ tensored with identity channels like this:

$$\Phi_1 \otimes \cdots \otimes \Phi_n = (\Phi_1 \otimes \mathbb{1}_{\mathcal{X}_2} \otimes \cdots \otimes \mathbb{1}_{\mathcal{X}_n}) \cdots (\mathbb{1}_{\mathcal{Y}_1} \otimes \cdots \otimes \mathbb{1}_{\mathcal{Y}_{n-1}} \otimes \Phi_n).$$

On the right hand side, we have a composition of tensor products of channels, defined in the usual way that one composes mappings. Each one of these tensor products of channels maps density operators to density operators, by the definitions of complete positivity and trace-preservation, and so the same thing is true of the product channel on the left hand side.

We will study the condition of complete positivity (as well as trace-preservation) in much greater detail in a couple of lectures.

3.2 Information complete measurements

In the remainder of this lecture we will discuss a couple of basic facts about states and measurements. The first fact is that states of registers are uniquely determined by the measurement statistics they generate. More precisely, if one knows the probability associated with every outcome of every measurement that could possibly be performed on a given register, then that registers state has been uniquely determined.

In fact, something stronger may be said: for any choice of a register X, there are choices of measurements on X that uniquely determine every possible state of X by the measurement statistics that they alone generate. Such measurements are called *information-complete* measurements. They are characterized by the property that their measurement operators span the space $\mathrm{L}(\mathcal{X})$.

Proposition 3.1. *Let $\mathcal{X}$ be a complex Euclidean space, and let*

$$\mu : \Gamma \to \mathrm{Pos}(\mathcal{X}) : a \mapsto P_a$$

be a measurement on $\mathcal{X}$ with the property that the collection $\{P_a : a \in \Gamma\}$ spans all of $\mathrm{L}(\mathcal{X})$. The mapping $\phi : \mathrm{L}(\mathcal{X}) \to \mathbb{C}^\Gamma$, defined by

$$(\phi(X))(a) = \langle P_a, X \rangle$$

for all $X \in \mathrm{L}(\mathcal{X})$ and $a \in \Gamma$, is one-to-one on $\mathrm{L}(\mathcal{X})$.

Remark 3.2. Of course the fact that ϕ is one-to-one on $\mathrm{L}(\mathcal{X})$ implies that it is one-to-one on $\mathrm{D}(\mathcal{X})$, which is all we really care about for the sake of this discussion. It is no harder to prove the proposition for all of $\mathrm{L}(\mathcal{X})$, however, so it is stated in the more general way.

Proof. It is clear that ϕ is linear, so we must only prove $\ker(\phi) = \{0\}$. Assume $\phi(X) = 0$, meaning that $(\phi(X))(a) = \langle P_a, X \rangle = 0$ for all $a \in \Gamma$, and write

$$X = \sum_{a \in \Gamma} \alpha_a P_a$$

for some choice of $\{\alpha_a : a \in \Gamma\} \subset \mathbb{C}$. This is possible because $\{P_a : a \in \Gamma\}$ spans $\mathrm{L}(\mathcal{X})$. It follows that

$$\|X\|_2^2 = \langle X, X \rangle = \sum_{a \in \Gamma} \overline{\alpha_a} \langle P_a, X \rangle = 0,$$

and therefore $X = 0$ by the positive definiteness of the Frobenius norm. This implies $\ker(\phi) = \{0\}$, as required. □

Let us now construct a simple example of an information-complete measurement, for any choice of a complex Euclidean space $\mathcal{X} = \mathbb{C}^\Sigma$. We will assume that the elements of Σ have been ordered in some fixed way. For each pair $(a, b) \in \Sigma \times \Sigma$, define an operator $Q_{a,b} \in \mathrm{L}(\mathcal{X})$ as follows:

$$Q_{a,b} = \begin{cases} E_{a,a} & \text{if } a = b \\ E_{a,a} + E_{a,b} + E_{b,a} + E_{b,b} & \text{if } a < b \\ E_{a,a} + iE_{a,b} - iE_{b,a} + E_{b,b} & \text{if } a > b. \end{cases}$$

Each operator $Q_{a,b}$ is positive semidefinite, and the set $\{Q_{a,b} : (a, b) \in \Sigma \times \Sigma\}$ spans the space $\mathrm{L}(\mathcal{X})$. With the exception of the trivial case $|\Sigma| = 1$, the operator

$$Q = \sum_{(a,b) \in \Sigma \times \Sigma} Q_{a,b}$$

differs from the identity operator, which means that $\{Q_{a,b} : (a,b) \in \Sigma \times \Sigma\}$ is not generally a measurement. The operator Q is, however, positive definite, and by defining

$$P_{a,b} = Q^{-1/2} Q_{a,b} Q^{-1/2}$$

we have that $\mu : \Sigma \times \Sigma \to \mathrm{Pos}(\mathcal{X}) : (a,b) \mapsto P_{a,b}$ is an information-complete measurement.

It also holds that every state of an n-tuple of registers $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ is uniquely determined by the measurement statistics of all product measurements on $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$. This follows from the simple observation that for any choice of information-complete measurements

$$\begin{gathered} \mu_1 : \Gamma_1 \to \mathrm{Pos}(\mathcal{X}_1) \\ \vdots \\ \mu_n : \Gamma_n \to \mathrm{Pos}(\mathcal{X}_n) \end{gathered}$$

defined on $\mathsf{X}_1, \ldots, \mathsf{X}_n$, the product measurement given by

$$\mu(a_1, \ldots, a_n) = \mu_1(a_1) \otimes \cdots \otimes \mu_n(a_n)$$

is also necessarily information-complete.

3.3 Partial measurements

A natural notion concerning measurements is that of a *partial measurement*. This is the situation in which we have a collection of registers $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ in some state $\rho \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$, and we perform measurements on just a subset of these registers. These measurements will yield results as normal, but the remaining registers will continue to exist and have some state (which generally will depend on the particular measurement outcomes that resulted from the measurements).

For simplicity let us consider this situation for just a pair of registers (X, Y). Assume the pair has the state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$, and a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{X})$ is performed on X. Conditioned on the outcome $a \in \Gamma$ resulting from this measurement, the state of Y will become

$$\frac{\mathrm{Tr}_{\mathcal{X}}\left[(\mu(a) \otimes \mathbb{1}_{\mathcal{Y}})\rho\right]}{\langle \mu(a) \otimes \mathbb{1}_{\mathcal{Y}}, \rho \rangle}.$$

One way to see that this must indeed be the state of Y conditioned on the measurement outcome a is that it is the only state that is consistent with every possible measurement $\nu : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ that could independently be performed on Y.

To explain this in greater detail, let us write $\mathsf{A} = a$ to denote the event that the original measurement μ on X results in the outcome $a \in \Gamma$, and let us write $\mathsf{B} = b$ to denote the event that the new, hypothetical measurement ν on Y results in the outcome $b \in \Sigma$. We have

$$\Pr[(\mathsf{A} = a) \wedge (\mathsf{B} = b)] = \langle \mu(a) \otimes \nu(b), \rho \rangle$$

and

$$\Pr[\mathsf{A} = a] = \langle \mu(a) \otimes \mathbb{1}_{\mathcal{Y}}, \rho \rangle,$$

so by the rule of conditional probabilities we have

$$\Pr[\mathsf{B} = b | \mathsf{A} = a] = \frac{\Pr[(\mathsf{A} = a) \wedge (\mathsf{B} = b)]}{\Pr[\mathsf{A} = a]} = \frac{\langle \mu(a) \otimes \nu(b), \rho \rangle}{\langle \mu(a) \otimes \mathbb{1}_{\mathcal{Y}}, \rho \rangle}.$$

Noting that

$$\langle X \otimes Y, \rho \rangle = \langle Y, \mathrm{Tr}_{\mathcal{X}} \left[(X^* \otimes \mathbb{1}_{\mathcal{Y}}) \rho \right] \rangle$$

for all X, Y, and ρ, we see that

$$\frac{\langle \mu(a) \otimes \nu(b), \rho \rangle}{\langle \mu(a) \otimes \mathbb{1}_{\mathcal{Y}}, \rho \rangle} = \langle \nu(b), \xi_a \rangle$$

for

$$\xi_a = \frac{\mathrm{Tr}_{\mathcal{X}} \left[(\mu(a) \otimes \mathbb{1}_{\mathcal{Y}}) \rho \right]}{\langle \mu(a) \otimes \mathbb{1}_{\mathcal{Y}}, \rho \rangle}.$$

As states are uniquely determined by their measurement statistics, as we have just discussed, we see that $\xi_a \in \mathrm{D}(\mathcal{Y})$ must indeed be the state of Y, conditioned on the measurement μ having resulted in outcome $a \in \Gamma$. (Of course ξ_a is not well-defined when $\Pr[\mathsf{A} = a] = 0$, but we do not need to worry about conditioning on an event that will never happen.)

3.4 Observable differences between states

A natural way to measure the distance between probability vectors $p, q \in \mathbb{R}^{\Gamma}$ is by the 1-norm:

$$\| p - q \|_1 = \sum_{a \in \Gamma} | p(a) - q(a) | .$$

It is easily verified that

$$\| p - q \|_1 = 2 \max_{\Delta \subseteq \Gamma} \left(\sum_{a \in \Delta} p(a) - \sum_{a \in \Delta} q(a) \right).$$

This is a natural measure of distance because it quantifies the optimal probability that two known probability vectors can be distinguished, given a single sample from the distributions they specify.

As an example, let us consider a thought experiment involving two hypothetical people: Alice and Bob. Two probability vectors $p_0, p_1 \in \mathbb{R}^{\Gamma}$ are fixed, and are considered to be known to both Alice and Bob. Alice privately chooses a random bit $a \in \{0, 1\}$, uniformly at random, and uses the value a to randomly choose an element $b \in \Gamma$: if $a = 0$, she samples b according to p_0, and if $a = 1$, she samples b according to p_1. The sampled element $b \in \Gamma$ is given to Bob, whose goal is to identify the value of Alice's random bit a. Bob may only use the value of b, along with his knowledge of p_0 and p_1, when making his guess.

It is clear from Bayes' theorem what Bob should do to maximize his probability to correctly guess the value of a: if $p_0(b) > p_1(b)$, he should guess that $a = 0$, while if $p_0(b) < p_1(b)$ he should guess that $a = 1$. In case $p_0(b) = p_1(b)$, Bob may as well guess that $a = 0$ or $a = 1$ arbitrarily, for he has learned nothing at all about the value of a from such an element $b \in \Gamma$. Bob's probability to correctly identify the value of a using this strategy is

$$\frac{1}{2} + \frac{1}{4} \| p_0 - p_1 \|_1 ,$$

which can be verified by a simple calculation. This is an optimal strategy.

A slightly more general situation is one in which $a \in \{0, 1\}$ is not chosen uniformly, but rather

$$\Pr[a = 0] = \lambda \qquad \text{and} \qquad \Pr[a = 1] = 1 - \lambda$$

for some value of $\lambda \in [0,1]$. In this case, an optimal strategy for Bob is to guess that $a = 0$ if $\lambda p_0(b) > (1-\lambda)p_1(b)$, to guess that $a = 1$ if $\lambda p_0(b) < (1-\lambda)p_1(b)$, and to guess arbitrarily if $\lambda p_0(b) = (1-\lambda)p_1(b)$. His probability of correctness will be

$$\frac{1}{2} + \frac{1}{2} \left\| \lambda p_0 - (1-\lambda) p_1 \right\|.$$

Naturally, this generalizes the expression for the case $\lambda = 1/2$.

Now consider a similar scenario, except with quantum states $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ in place of probability vectors $p_0, p_1 \in \mathbb{R}^\Gamma$. More specifically, Alice chooses a random bit $a = \{0,1\}$ according to the distribution

$$\Pr[a = 0] = \lambda \qquad \text{and} \qquad \Pr[a = 1] = 1 - \lambda,$$

for some choice of $\lambda \in [0,1]$ (which is known to both Alice and Bob). She then hands Bob a register X that has been prepared in the quantum state $\rho_a \in \mathrm{D}(\mathcal{X})$. This time, Bob has the freedom to choose whatever measurement he wants in trying to guess the value of a.

Note that there is no generality lost in assuming Bob makes a measurement having outcomes 0 and 1. If he were to make any other measurement, perhaps with many outcomes, and then process the outcome in some way to arrive at a guess for the value of a, we could simply combine his measurement with the post-processing phase to arrive at the description of a measurement with outcomes 0 and 1.

The following theorem states, in mathematical terms, that Bob's optimal strategy correctly identifies a with probability

$$\frac{1}{2} + \frac{1}{2} \left\| \lambda \rho_0 - (1-\lambda) \rho_1 \right\|_1,$$

which is a similar expression to the one we had in the classical case. The proof of the theorem also makes clear precisely what strategy Bob should employ for optimality.

Theorem 3.3 (Helstrom)**.** *Let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be states and let $\lambda \in [0,1]$. For every choice of positive semidefinite operators $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ for which $P_0 + P_1 = \mathbb{1}_\mathcal{X}$, it holds that*

$$\lambda \langle P_0, \rho_0 \rangle + (1-\lambda) \langle P_1, \rho_1 \rangle \leq \frac{1}{2} + \frac{1}{2} \left\| \lambda \rho_0 - (1-\lambda) \rho_1 \right\|_1.$$

Moreover, equality is achieved for some choice of projection operators $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ with $P_0 + P_1 = \mathbb{1}_\mathcal{X}$.

Proof. First, note that

$$\begin{aligned}
&(\lambda \langle P_0, \rho_0 \rangle + (1-\lambda) \langle P_1, \rho_1 \rangle) - (\lambda \langle P_1, \rho_0 \rangle + (1-\lambda) \langle P_0, \rho_1 \rangle) = \langle P_0 - P_1, \lambda \rho_0 - (1-\lambda) \rho_1 \rangle \\
&(\lambda \langle P_0, \rho_0 \rangle + (1-\lambda) \langle P_1, \rho_1 \rangle) + (\lambda \langle P_1, \rho_0 \rangle + (1-\lambda) \langle P_0, \rho_1 \rangle) = 1,
\end{aligned}$$

and therefore

$$\lambda \langle P_0, \rho_0 \rangle + (1-\lambda) \langle P_1, \rho_1 \rangle = \frac{1}{2} + \frac{1}{2} \langle P_0 - P_1, \lambda \rho_0 - (1-\lambda) \rho_1 \rangle, \tag{3.1}$$

for any choice of $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ with $P_0 + P_1 = \mathbb{1}_\mathcal{X}$.

Now, for every unit vector $u \in \mathcal{X}$ we have

$$|u^*(P_0 - P_1)u| = |u^* P_0 u - u^* P_1 u| \leq u^* P_0 u + u^* P_1 u = u^*(P_0 + P_1)u = 1,$$

and therefore (as $P_0 - P_1$ is Hermitian) it holds that $\|P_0 - P_1\| \leq 1$. By Hölder's inequality (for Schatten p-norms) we therefore have

$$\langle P_0 - P_1, \lambda\rho_0 - (1-\lambda)\rho_1 \rangle \leq \|P_0 - P_1\| \, \|\lambda\rho_0 - (1-\lambda)\rho_1\|_1 \leq \|\lambda\rho_0 - (1-\lambda)\rho_1\|_1 ,$$

and so the inequality in the theorem follows from (3.1).

To prove equality can be achieved for projection operators $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ with $P_0 + P_1 = \mathbb{1}_{\mathcal{X}}$, we consider a spectral decomposition

$$\lambda\rho_0 - (1-\lambda)\rho_1 = \sum_{j=1}^{n} \eta_j x_j x_j^* .$$

Defining

$$P_0 = \sum_{j:\, \eta_j \geq 0} x_j x_j^* \qquad \text{and} \qquad P_1 = \sum_{j:\, \eta_j < 0} x_j x_j^* ,$$

we have that P_0 and P_1 are projections with $P_0 + P_1 = \mathbb{1}_{\mathcal{X}}$, and moreover

$$(P_0 - P_1)(\lambda\rho_0 - (1-\lambda)\rho_1) = \sum_{j=1}^{n} \big|\eta_j\big| \, x_j x_j^* .$$

It follows that

$$\langle P_0 - P_1, \lambda\rho_0 - (1-\lambda)\rho_1 \rangle = \sum_{j=1}^{n} \big|\eta_j\big| = \|\lambda\rho_0 - (1-\lambda)\rho_1\|_1 ,$$

and by (3.1) we obtain the desired equality. □