Multiple systems

Download the slides for this lesson.

Open the YouTube video for this lesson in a separate window.

Introduction

This lesson focuses on the basics of quantum information in the context of *multiple* systems. This context arises both commonly and naturally in information processing, classical and quantum; information-carrying systems are typically constructed from collections of smaller systems, such as bits or qubits.

A simple, yet critically important idea to keep in mind going into this lesson is that we can always choose to view multiple systems *together* as if they form a single, compound system — to which the discussion in the previous lesson applies. Indeed, this idea very directly leads to a description of how quantum states, measurements, and operations work for multiple systems.

There is, however, more to understanding multiple quantum systems than simply recognizing that they may be viewed collectively as single systems. For instance, we may have multiple quantum systems that are collectively in a particular quantum state, and then choose to measure some but not all of the individual systems. In general, this will affect the state of the systems that were not measured, and it is important to understand exactly how when analyzing quantum algorithms and protocols. An understanding of the sorts of *correlations* among multiple systems — and particularly a type of correlation known as *entanglement* — is also important in quantum information and computation.

Classical information

Like we did in the previous lesson, we'll begin this lesson with a discussion of classical information. Once again, the probabilistic and quantum descriptions are mathematically similar, and recognizing how the mathematics works in the familiar setting of classical information is helpful in understanding why quantum information is described in the way that it is.

Classical states via the Cartesian product

We'll start at a very basic level, with classical states of multiple systems. For simplicity, we'll begin by discussing just two systems, and then generalize to more than two systems.

To be precise, let X be a system whose classical state set is Σ, and let Y be a second system whose classical state set is Γ. Note that, because we have referred to these sets as *classical state sets*, our assumption is that Σ and Γ are both finite and nonempty. It could be that $\Sigma = \Gamma$, but this is not necessarily so — and regardless, it will be helpful to use different names to refer to these sets in the interest of clarity.

Now imagine that the two systems, X and Y, are placed side-by-side, with X on the left and Y on the right. If we so choose, we can view these two systems as if they form a single system, which we can denote by (X, Y) or XY depending on our preference. A natural question to ask about this compound system (X, Y) is, "What are its classical states?"

The answer is that the set of classical states of (X, Y) is the *Cartesian product* of Σ and Γ, which is the set defined as

$$\Sigma \times \Gamma = \big\{(a, b) \,:\, a \in \Sigma \text{ and } b \in \Gamma\big\}.$$

In simple terms, the Cartesian product is precisely the mathematical notion that captures the idea of viewing an element of one set and an element of a second set together, as if they form a single element of a single set. In the case at hand, to say that (X, Y) is in the classical state $(a, b) \in \Sigma \times \Gamma$ means that X is in the classical state $a \in \Sigma$ and Y is in the classical state $b \in \Gamma$; and if the classical state of X is $a \in \Sigma$ and the classical state of Y is $b \in \Gamma$, then the classical state of the joint system (X, Y) is (a, b).

For more than two systems, the situation generalizes in a natural way. If we suppose that $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are systems having classical state sets $\Sigma_1, \ldots, \Sigma_n$, respectively, for any positive integer n, the classical state set of the n-tuple $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$, viewed as a single joint system, is the Cartesian product

$$\Sigma_1 \times \cdots \times \Sigma_n = \big\{(a_1, \ldots, a_n) \,:\, a_1 \in \Sigma_1,\ \ldots,\ a_n \in \Sigma_n\big\}.$$

Of course, we are free to use whatever names we wish for systems, and to order them as we choose. In particular, if we have n systems like above, we could instead choose to name them $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$ and arrange them from right to left, so that the joint system becomes $(\mathsf{X}_{n-1}, \ldots, \mathsf{X}_0)$. Following the same pattern for naming the associated classical states and classical state sets, we might then refer to a classical state

$$(a_{n-1}, \ldots, a_0) \in \Sigma_{n-1} \times \cdots \times \Sigma_0$$

of this compound system. Indeed, this is the ordering convention used by Qiskit when naming multiple qubits. We'll come back to this convention and how it connects to quantum circuits in the next lesson, but we'll start using it now to help to get used to it.

Representing states as strings

It is often convenient to write a classical state of the form $(a_{n-1}, \ldots, a_0)$ as a string $a_{n-1} \cdots a_0$ for the sake of brevity, particularly in the very typical situation that the classical state sets $\Sigma_0, \ldots, \Sigma_{n-1}$ are associated with sets of *symbols* or *characters*. In this context, the term *alphabet* is commonly used to refer to sets of symbols used to form strings, but the mathematical definition of an alphabet is precisely the same as the definition of a classical state set: it is a finite and nonempty set.

For example, suppose that $\mathsf{X}_0, \ldots, \mathsf{X}_9$ are bits, so that the classical state sets of these systems are all the same.

$$\Sigma_0 = \Sigma_1 = \cdots = \Sigma_9 = \{0, 1\}$$

There are then $2^{10} = 1024$ classical states of the joint system $(\mathsf{X}_9, \ldots, \mathsf{X}_0)$, which are the elements of the set

$$\Sigma_9 \times \Sigma_8 \times \cdots \times \Sigma_0 = \{0, 1\}^{10}.$$

Written as strings, these classical states look like this:

$$\begin{array}{c} 0000000000 \\ 0000000001 \\ 0000000010 \\ 0000000011 \\ 0000000100 \\ \vdots \\ 1111111111 \end{array}$$

For the classical state 0000000110, for instance, we see that X_1 and X_2 are in the state 1, while all other systems are in the state 0.

Probabilistic states

Recall from the previous lesson that a *probabilistic state* associates a probability with each classical state of a system. Thus, a probabilistic state of multiple systems — viewed collectively as a single system — associates a probability with each element of the Cartesian product of the classical state sets of the individual systems.

For example, suppose that X and Y are both bits, so that their corresponding classical state sets are $\Sigma = \{0, 1\}$ and $\Gamma = \{0, 1\}$, respectively. Here is a probabilistic state of the pair (X, Y) :

$$\Pr\big((\mathsf{X},\mathsf{Y}) = (0,0)\big) = 1/2$$

$$\Pr\big((\mathsf{X},\mathsf{Y}) = (0,1)\big) = 0$$

$$\Pr\big((\mathsf{X},\mathsf{Y}) = (1,0)\big) = 0$$

$$\Pr\big((\mathsf{X},\mathsf{Y}) = (1,1)\big) = 1/2$$

This probabilistic state is one in which both X and Y are random bits — each is 0 with probability $1/2$ and 1 with probability $1/2$ — but the classical states of the two bits always agree. This is an example of a *correlation* between these systems.

Ordering Cartesian product state sets

Probabilistic states of systems can be represented by probability vectors, as was discussed in the previous lesson. In particular, the vector entries represent probabilities for the system to be in the possible classical states of that system, and the understanding is that a correspondence between the entries and the set of classical states has been selected. Choosing such a correspondence effectively means deciding on an ordering of the classical states, which is often natural or determined by a standard convention. For example, the binary alphabet $\{0,1\}$ is naturally ordered with 0 first and 1 second, so the first entry in a probability vector representing a probabilistic state of a bit is the probability for it to be in the state 0, and the second entry is the probability for it to be in the state 1.

None of this changes in the context of multiple systems, but there is a decision to be made. The classical state set of multiple systems together, viewed collectively as a single system, is a Cartesian product of the classical state sets of the individual systems — so we must decide how the elements of Cartesian products of classical state sets are ordered.

There is a simple convention that we follow for doing this, which is to start with whatever orderings are already in place for the individual classical state sets, and then to order the elements of the Cartesian product *alphabetically.* Another way to say this is that the entries in each n-tuple (or, equivalently, the symbols in each string) are treated as though they have significance that *decreases from left to right.* For example, according to this convention, the Cartesian product $\{1,2,3\} \times \{0,1\}$ is ordered like this:

$$(1,0),\ (1,1),\ (2,0),\ (2,1),\ (3,0),\ (3,1).$$

When n-tuples are written as strings and ordered in this way, we observe familiar patterns, such as $\{0,1\} \times \{0,1\}$ being ordered as $00, 01, 10, 11$, and the set $\{0,1\}^{10}$ being ordered as it was written earlier in the lesson. As another example, viewing the set $\{0,1,\ldots,9\} \times \{0,1,\ldots,9\}$ as a set of strings, we obtain the two-digit numbers 00 through 99, ordered numerically. This is obviously not a coincidence; our decimal number system uses precisely this sort of alphabetical ordering, where the word *alphabetical* should be understood as having a broad meaning that includes numerals in addition to letters.

Returning to the example of two bits from above, the probabilistic state described previously is therefore represented by the following probability vector, where the entries are labeled explicitly for the sake of clarity.

$$\begin{pmatrix} \frac{1}{2} \\ 0 \\ 0 \\ \frac{1}{2} \end{pmatrix} \begin{array}{l} \leftarrow \text{probability of being in the state } 00 \\ \leftarrow \text{probability of being in the state } 01 \\ \leftarrow \text{probability of being in the state } 10 \\ \leftarrow \text{probability of being in the state } 11 \end{array} \tag{1}$$

Independence of two systems

A special type of probabilistic state of two systems is one in which the systems are *independent*. Intuitively speaking, two systems are independent if learning the classical state of either system has no effect on the probabilities associated with the other. That is, learning what classical state one of the systems is in provides no information at all about the classical state of the other.

To define this notion precisely, let us suppose once again that X and Y are systems having classical state sets Σ and Γ, respectively. With respect to a given probabilistic state of these systems, they are said to be *independent* if it is the case that

$$\operatorname{Pr}((\mathsf{X},\mathsf{Y}) = (a,b)) = \operatorname{Pr}(\mathsf{X} = a)\operatorname{Pr}(\mathsf{Y} = b) \tag{2}$$

for every choice of $a \in \Sigma$ and $b \in \Gamma$.

To express this condition in terms of probability vectors, assume that the given probabilistic state of (X,Y) is described by a probability vector, written in the Dirac notation as

$$\sum_{(a,b)\in\Sigma\times\Gamma} p_{ab}|ab\rangle.$$

The condition (2) for independence is then equivalent to the existence of two probability vectors

$$|\phi\rangle = \sum_{a\in\Sigma} q_a|a\rangle \quad \text{and} \quad |\psi\rangle = \sum_{b\in\Gamma} r_b|b\rangle, \tag{3}$$

representing the probabilities associated with the classical states of X and Y, respectively, such that

$$p_{ab} = q_a r_b \tag{4}$$

for all $a \in \Sigma$ and $b \in \Gamma$.

For example, the probabilistic state of a pair of bits (X, Y) represented by the vector

$$\frac{1}{6}|00\rangle + \frac{1}{12}|01\rangle + \frac{1}{2}|10\rangle + \frac{1}{4}|11\rangle$$

is one in which X and Y are independent. Specifically, the condition required for independence is true for the probability vectors

$$|\phi\rangle = \frac{1}{4}|0\rangle + \frac{3}{4}|1\rangle \quad \text{and} \quad |\psi\rangle = \frac{2}{3}|0\rangle + \frac{1}{3}|1\rangle.$$

For instance, to make the probabilities for the 00 state match, we need $\frac{1}{6} = \frac{1}{4} \times \frac{2}{3}$, and indeed this is the case. Other entries can be verified in a similar manner.

On the other hand, the probabilistic state (1), which we may write as

$$\frac{1}{2}|00\rangle + \frac{1}{2}|11\rangle, \tag{5}$$

does not represent independence between the systems X and Y. A simple way to argue this follows.

Suppose that there did exist probability vectors $|\phi\rangle$ and $|\psi\rangle$, as in equation (3) above, for which the condition (4) is satisfied for every choice of a and b. It would then necessarily be that

$$q_0 r_1 = \Pr\big((\mathsf{X}, \mathsf{Y}) = (0, 1)\big) = 0.$$

This implies that either $q_0 = 0$ or $r_1 = 0$, because if both were nonzero, the product $q_0 r_1$ would also not be zero. This leads to the conclusion that either $q_0 r_0 = 0$ (in case $q_0 = 0$) or $q_1 r_1 = 0$ (in case $r_1 = 0$). We see, however, that neither of those equalities can be true because we must have $q_0 r_0 = 1/2$ and $q_1 r_1 = 1/2$. Hence, there do not exist vectors $|\phi\rangle$ and $|\psi\rangle$ satisfying the property required for independence.

Having defined independence between two systems, we can now define what is meant by *correlation:* it is a *lack of independence*. For example, because the two bits in the probabilistic state represented by the vector (5) are not independent, they are, by definition, correlated.

Tensor products of vectors

The condition of independence just described can be expressed succinctly through the notion of a *tensor product*. Although tensor products are a very general notion, and can be defined quite abstractly and applied to a variety of mathematical structures, we can adopt a simple and concrete definition in the case at hand.

Given two vectors

$$|\phi\rangle = \sum_{a\in\Sigma} \alpha_a |a\rangle \quad \text{and} \quad |\psi\rangle = \sum_{b\in\Gamma} \beta_b |b\rangle,$$

the tensor product $|\phi\rangle \otimes |\psi\rangle$ is the vector defined as

$$|\phi\rangle \otimes |\psi\rangle = \sum_{(a,b)\in\Sigma\times\Gamma} \alpha_a\beta_b |ab\rangle.$$

The entries of this new vector correspond to the elements of the Cartesian product $\Sigma \times \Gamma$, which are written as strings in the previous equation. Equivalently, the vector $|\pi\rangle = |\phi\rangle \otimes |\psi\rangle$ is defined by the equation

$$\langle ab|\pi\rangle = \langle a|\phi\rangle\langle b|\psi\rangle$$

being true for every $a \in \Sigma$ and $b \in \Gamma$.

We can now recast the condition for independence: for a joint system (X, Y) in a probabilistic state represented by a probability vector $|\pi\rangle$, the systems X and Y are independent if $|\pi\rangle$ is obtained by taking a tensor product

$$|\pi\rangle = |\phi\rangle \otimes |\psi\rangle$$

of probability vectors $|\phi\rangle$ and $|\psi\rangle$ on each of the subsystems X and Y. In this situation, $|\pi\rangle$ is said to be a *product state* or *product vector*.

We often omit the symbol $\otimes$ when taking the tensor product of kets, such as writing $|\phi\rangle|\psi\rangle$ rather than $|\phi\rangle \otimes |\psi\rangle$. This convention captures the idea that the tensor product is, in this context, the most natural or default way to take the product of two vectors. Although it is less common, the notation $|\phi \otimes \psi\rangle$ is also sometimes used.

When we use the alphabetical convention for ordering elements of Cartesian products, we obtain the following specification for the tensor product of two column vectors.

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} \otimes \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} \alpha_1\beta_1 \\ \vdots \\ \alpha_1\beta_k \\ \alpha_2\beta_1 \\ \vdots \\ \alpha_2\beta_k \\ \vdots \\ \alpha_m\beta_1 \\ \vdots \\ \alpha_m\beta_k \end{pmatrix}$$

As an important aside, notice the following expression for tensor products of standard basis vectors:

$$|a\rangle \otimes |b\rangle = |ab\rangle.$$

We could alternatively write (a, b) as an ordered pair, rather than a string, in which case we obtain $|a\rangle \otimes |b\rangle = |(a, b)\rangle$. It is, however, more common to omit the parentheses in this situation, instead writing $|a\rangle \otimes |b\rangle = |a, b\rangle$. This is typical in mathematics more generally; parentheses that don't add clarity or remove ambiguity are often simply omitted.

The tensor product of two vectors has the important property that it is *bilinear*, which means that it is linear in each of the two arguments separately, assuming that the other argument is fixed. This property can be expressed through these equations:

1. Linearity in the first argument:

$$\big(|\phi_1\rangle + |\phi_2\rangle\big) \otimes |\psi\rangle = |\phi_1\rangle \otimes |\psi\rangle + |\phi_2\rangle \otimes |\psi\rangle$$
$$\big(\alpha|\phi\rangle\big) \otimes |\psi\rangle = \alpha\big(|\phi\rangle \otimes |\psi\rangle\big)$$

2. Linearity in the second argument:

$$|\phi\rangle \otimes \big(|\psi_1\rangle + |\psi_2\rangle\big) = |\phi\rangle \otimes |\psi_1\rangle + |\phi\rangle \otimes |\psi_2\rangle$$
$$|\phi\rangle \otimes \big(\alpha|\psi\rangle\big) = \alpha\big(|\phi\rangle \otimes |\psi\rangle\big)$$

Considering the second equation in each of these pairs of equations, we see that scalars "float freely" within tensor products:

$$\big(\alpha|\phi\rangle\big) \otimes |\psi\rangle = |\phi\rangle \otimes \big(\alpha|\psi\rangle\big) = \alpha\big(|\phi\rangle \otimes |\psi\rangle\big).$$

There is therefore no ambiguity in simply writing $\alpha|\phi\rangle \otimes |\psi\rangle$, or alternatively $\alpha|\phi\rangle|\psi\rangle$ or $\alpha|\phi \otimes \psi\rangle$, to refer to this vector.

Independence and tensor products for three or more systems

The notions of independence and tensor products generalize straightforwardly to three or more systems. If $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$ are systems having classical state sets $\Sigma_0, \ldots, \Sigma_{n-1}$, respectively, then a probabilistic state of the combined system $(\mathsf{X}_{n-1}, \ldots, \mathsf{X}_0)$ is a *product state* if the associated probability vector takes the form

$$|\psi\rangle = |\phi_{n-1}\rangle \otimes \cdots \otimes |\phi_0\rangle$$

for probability vectors $|\phi_0\rangle, \ldots, |\phi_{n-1}\rangle$ describing probabilistic states of $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$. Here, the definition of the tensor product generalizes in a natural way: the vector

$$|\psi\rangle = |\phi_{n-1}\rangle \otimes \cdots \otimes |\phi_0\rangle$$

is defined by the equation

$$\langle a_{n-1} \cdots a_0|\psi\rangle = \langle a_{n-1}|\phi_{n-1}\rangle \cdots \langle a_0|\phi_0\rangle$$

being true for every $a_0 \in \Sigma_0, \ldots a_{n-1} \in \Sigma_{n-1}$.

A different, but equivalent, way to define the tensor product of three or more vectors is recursively in terms of tensor products of two vectors:

$$|\phi_{n-1}\rangle \otimes \cdots \otimes |\phi_0\rangle = |\phi_{n-1}\rangle \otimes \big(|\phi_{n-2}\rangle \otimes \cdots \otimes |\phi_0\rangle\big).$$

Similar to the tensor product of just two vectors, the tensor product of three or more vectors is linear in each of the arguments individually, assuming that all other arguments are fixed. In this case, it is said that the tensor product of three or more vectors is *multilinear*.

Like in the case of two systems, we could say that the systems $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$ are *independent* when they are in a product state, but the term *mutually independent* is more precise. There happen to be other notions of independence for three or more systems, such as *pairwise independence*, that are both interesting and important — but not in the context of this course.

Generalizing the observation earlier concerning tensor products of standard basis vectors, for any positive integer n and any classical states $a_0, \ldots, a_{n-1}$, we have

$$|a_{n-1}\rangle \otimes \cdots \otimes |a_0\rangle = |a_{n-1}\cdots a_0\rangle.$$

Measurements of probabilistic states

Now let us move on to measurements of probabilistic states of multiple systems. By choosing to view multiple systems together as single systems, we immediately obtain a specification of how measurements must work for multiple systems — provided that *all* of the systems are measured.

For example, if the probabilistic state of two bits (X, Y) is described by the probability vector

$$\frac{1}{2}|00\rangle + \frac{1}{2}|11\rangle,$$

then the outcome 00 — meaning 0 for the measurement of X and 0 for the measurement of Y — is obtained with probability $1/2$ and the outcome 11 is also obtained with probability $1/2$. In each case we update the probability vector description of our knowledge accordingly, so that the probabilistic state becomes $|00\rangle$ or $|11\rangle$, respectively.

Partial measurements

We could, however, choose to measure not *every* system, but instead just some of the systems. This will result in a measurement outcome for each system that gets measured, and will also (in general) affect our knowledge of the remaining systems that we didn't measure.

To explain how this works, we'll focus on the case of two systems, one of which is measured. The more general situation — in which some proper subset of three or more systems is measured — effectively reduces to the case of two systems when we view the systems that are measured collectively as if they form one system and the systems that are not measured as if they form a second system. To be precise, let's suppose that X and Y are systems whose classical state sets are Σ and Γ, respectively, and that the two systems together are in some probabilistic state. We'll consider what happens when we measure just X and do nothing to Y. The situation where just Y is measured and nothing happens to X is handled symmetrically.

First, we know that the probability to observe a particular classical state $a \in \Sigma$ when just X is measured must be consistent with the probabilities

we would obtain under the assumption that Y was also measured. That is, we must have

$$\operatorname{Pr}(\mathsf{X} = a) = \sum_{b \in \Gamma} \operatorname{Pr}\bigl((\mathsf{X}, \mathsf{Y}) = (a, b)\bigr).$$

This is the formula for the so-called *reduced* (or *marginal*) probabilistic state of X alone.

This formula makes perfect sense at an intuitive level, in the sense that something very strange would have to happen for it to be wrong. If it were wrong, that would mean that measuring Y could somehow influence the probabilities associated with different outcomes of the measurement of X, irrespective of the actual outcome of the measurement of Y. If Y happened to be in a distant location, such as somewhere in another galaxy for instance, this would allow for faster-than-light signaling — which we reject based on our understanding of physics. Another way to understand this comes from the interpretation of probability as reflecting a degree of belief. The mere fact that someone else might decide to look at Y cannot change the classical state of X, so without any information about what they did or didn't see, one's beliefs about the state of X should not change as a result.

Now, given the assumption that only X is measured and Y is not, there may still exist uncertainty over the classical state of Y. For this reason, rather than updating our description of the probabilistic state of (X, Y) to $|ab\rangle$ for some selection of $a \in \Sigma$ and $b \in \Gamma$, we must update our description so that this uncertainty about Y is properly reflected. The following *conditional probability* formula reflects this uncertainty.

$$\operatorname{Pr}(\mathsf{Y} = b \,|\, \mathsf{X} = a) = \frac{\operatorname{Pr}\bigl((\mathsf{X}, \mathsf{Y}) = (a, b)\bigr)}{\operatorname{Pr}(\mathsf{X} = a)}$$

Here, the expression $\operatorname{Pr}(\mathsf{Y} = b \,|\, \mathsf{X} = a)$ denotes the probability that $\mathsf{Y} = b$ *conditioned* on (or *given* that) $\mathsf{X} = a$. Technically speaking, this expression only makes sense if $\operatorname{Pr}(\mathsf{X} = a)$ is nonzero, for if $\operatorname{Pr}(\mathsf{X} = a) = 0$, then we're dividing by zero and we obtain indeterminate form $\frac{0}{0}$. This is not a problem, though, because if the probability associated with a is zero, then we'll never obtain a as an outcome of a measurement of X, so we don't need to be concerned with this possibility.

To express these formulas in terms of probability vectors, consider a probability vector $|\psi\rangle$ describing a joint probabilistic state of (X, Y).

$$|\psi\rangle = \sum_{(a,b)\in\Sigma\times\Gamma} p_{ab}|ab\rangle$$

Measuring X alone yields each possible outcome $a \in \Sigma$ with probability

$$\Pr(\mathsf{X} = a) = \sum_{c\in\Gamma} p_{ac}.$$

The vector representing the probabilistic state of X alone (i.e., the reduced probabilistic state of X) is therefore given by

$$\sum_{a\in\Sigma}\left(\sum_{c\in\Gamma} p_{ac}\right)|a\rangle.$$

Having obtained a particular outcome $a \in \Sigma$ of the measurement of X, the probabilistic state of Y is updated according to the formula for conditional probabilities, so that it is represented by this probability vector:

$$|\pi_a\rangle = \frac{\sum_{b\in\Gamma} p_{ab}|b\rangle}{\sum_{c\in\Gamma} p_{ac}}.$$

In the event that the measurement of X resulted in the classical state a, we therefore update our description of the probabilistic state of the joint system (X,Y) to $|a\rangle \otimes |\pi_a\rangle$.

One way to think about this definition of $|\pi_a\rangle$ is to see it as a *normalization* of the vector $\sum_{b\in\Gamma} p_{ab}|b\rangle$, where we divide by the sum of the entries in this vector to obtain a probability vector. This normalization effectively accounts for a conditioning on the event that the measurement of X has resulted in the outcome a.

For a specific example, suppose that classical state set of X is $\Sigma = \{0,1\}$, the classical state set of Y is $\Gamma = \{1,2,3\}$, and the probabilistic state of (X,Y) is

$$|\psi\rangle = \frac{1}{2}|0,1\rangle + \frac{1}{12}|0,3\rangle + \frac{1}{12}|1,1\rangle + \frac{1}{6}|1,2\rangle + \frac{1}{6}|1,3\rangle.$$

Our goal will be to determine the probabilities of the two possible outcomes (0 and 1), and to calculate what the resulting probabilistic state of Y is for the two outcomes, assuming the system X is measured.

Using the bilinearity of the tensor product, and specifically the fact that it is linear in the *second* argument, we may rewrite the vector $|\psi\rangle$ as follows:

$$|\psi\rangle = |0\rangle \otimes \left(\frac{1}{2}|1\rangle + \frac{1}{12}|3\rangle\right) + |1\rangle \otimes \left(\frac{1}{12}|1\rangle + \frac{1}{6}|2\rangle + \frac{1}{6}|3\rangle\right).$$

In words, what we've done is to isolate the distinct standard basis vectors for the first system (i.e., the one being measured), tensoring each with the linear combination of standard basis vectors for the second system we get by picking out the entries of the original vector that are consistent with the corresponding classical state of the first system. A moment's thought reveals that this is always possible, regardless of what vector we started with.

Having expressed our probability vector in this way, the effects of measuring the first system become easy to analyze. The probabilities of the two outcomes can be obtained by summing the probabilities in parentheses.

$$\begin{aligned}\Pr(\mathsf{X} = 0) &= \frac{1}{2} + \frac{1}{12} = \frac{7}{12}\\ \Pr(\mathsf{X} = 1) &= \frac{1}{12} + \frac{1}{6} + \frac{1}{6} = \frac{5}{12}\end{aligned}$$

These probabilities sum to one, as expected — but this is a useful check on our calculations.

And now, the probabilistic state of Y conditioned on each possible outcome can be quickly inferred by normalizing the vectors in parentheses. That is, we divide these vectors by the associated probabilities we just calculated, so that they become probability vectors. Thus, conditioned on X being 0, the probabilistic state of Y becomes

$$\frac{\frac{1}{2}|1\rangle + \frac{1}{12}|3\rangle}{\frac{7}{12}} = \frac{6}{7}|1\rangle + \frac{1}{7}|3\rangle,$$

and conditioned on the measurement of X being 1, the probabilistic state of Y becomes

$$\frac{\frac{1}{12}|1\rangle + \frac{1}{6}|2\rangle + \frac{1}{6}|3\rangle}{\frac{5}{12}} = \frac{1}{5}|1\rangle + \frac{2}{5}|2\rangle + \frac{2}{5}|3\rangle.$$

Operations on probabilistic states

To conclude this discussion of classical information for multiple systems, we'll consider *operations* on multiple systems in probabilistic states. Following the same idea as before, we can view multiple systems

collectively as single, compound systems, and then look to the previous lesson to see how this works.

Returning to the typical set-up where we have two systems X and Y, let us consider classical operations on the compound system (X, Y). Based on the previous lesson and the discussion above, we conclude that any such operation is represented by a stochastic matrix whose rows and columns are indexed by the Cartesian product $\Sigma \times \Gamma$.

For example, suppose that X and Y are bits, and consider an operation with the following description.

> If $\mathsf{X} = 1$, then perform a NOT operation on Y.
> Otherwise do nothing.

This is a deterministic operation known as a *controlled-NOT* operation, where X is the *control* bit that determines whether or not a NOT operation should be applied to the *target* bit Y. Here is the matrix representation of this operation:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Its action on standard basis states is as follows.

$$\begin{aligned} |00\rangle &\mapsto |00\rangle \\ |01\rangle &\mapsto |01\rangle \\ |10\rangle &\mapsto |11\rangle \\ |11\rangle &\mapsto |10\rangle \end{aligned}$$

If we were to exchange the roles of X and Y, taking Y to be the control bit and X to be the target bit, then the matrix representation of the operation would become

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

and its action on standard basis states would be like this:

$$
\begin{aligned}
|00\rangle &\mapsto |00\rangle \\
|01\rangle &\mapsto |11\rangle \\
|10\rangle &\mapsto |10\rangle \\
|11\rangle &\mapsto |01\rangle
\end{aligned}
$$

Another example is the operation having this description:

> Perform one of the following two operations, each with probability $1/2$:
> 1. Set Y to be equal to X.
> 2. Set X to be equal to Y.

The matrix representation of this operation is as follows:

$$
\begin{pmatrix}
1 & \frac{1}{2} & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & \frac{1}{2} & \frac{1}{2} & 1
\end{pmatrix}
= \frac{1}{2}
\begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1
\end{pmatrix}
+ \frac{1}{2}
\begin{pmatrix}
1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1
\end{pmatrix}.
$$

The action of this operation on standard basis vectors is as follows:

$$
\begin{aligned}
|00\rangle &\mapsto |00\rangle \\
|01\rangle &\mapsto \frac{1}{2}|00\rangle + \frac{1}{2}|11\rangle \\
|10\rangle &\mapsto \frac{1}{2}|00\rangle + \frac{1}{2}|11\rangle \\
|11\rangle &\mapsto |11\rangle
\end{aligned}
$$

In these examples, we are simply viewing two systems together as a single system and proceeding as in the previous lesson.

The same thing can be done for any number of systems. For example, imagine that we have three bits, and we increment the three bits modulo 8 — meaning that we think about the three bits as encoding a number between 0 and 7 using binary notation, add 1, and then take the remainder after dividing by 8. One way to express this operation is like this:

$$
\begin{aligned}
&|001\rangle\langle 000| + |010\rangle\langle 001| + |011\rangle\langle 010| + |100\rangle\langle 011| \\
&\quad + |101\rangle\langle 100| + |110\rangle\langle 101| + |111\rangle\langle 110| + |000\rangle\langle 111|.
\end{aligned}
$$

Another way to express it is as

$$\sum_{k=0}^{7} |(k+1) \bmod 8\rangle\langle k|,$$

assuming we've agreed that numbers from 0 to 7 inside of kets refer to the three-bit binary encodings of those numbers. A third option is to express this operation as a matrix.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Independent operations

Now suppose that we have multiple systems and we *independently* perform different operations on the systems separately.

For example, taking our usual set-up of two systems X and Y having classical state sets Σ and Γ, respectively, let us suppose that we perform one operation on X and, completely independently, another operation on Y. As we know from the previous lesson, these operations are represented by stochastic matrices — and to be precise, let us say that the operation on X is represented by the matrix M and the operation on Y is represented by the matrix N. Thus, the rows and columns of M have indices that are placed in correspondence with the elements of Σ and, likewise, the rows and columns of N correspond to the elements of Γ.

A natural question to ask is this: if we view X and Y together as a single, compound system (X, Y), what is the matrix that represents the combined action of the two operations on this compound system? To answer this question, we must first introduce tensor products of matrices — which are similar to the tensor product of vectors and are defined analogously.

Tensor products of matrices

The tensor product $M \otimes N$ of the matrices

$$M = \sum_{a,b \in \Sigma} \alpha_{ab} |a\rangle\langle b|$$

and

$$N = \sum_{c,d \in \Gamma} \beta_{cd} |c\rangle\langle d|$$

is the matrix

$$M \otimes N = \sum_{a,b \in \Sigma} \sum_{c,d \in \Gamma} \alpha_{ab} \beta_{cd} |ac\rangle\langle bd|$$

Equivalently, the tensor product of M and N is defined by the equation

$$\langle ac | M \otimes N | bd \rangle = \langle a | M | b \rangle \langle c | N | d \rangle$$

being true for every selection of $a, b \in \Sigma$ and $c, d \in \Gamma$.

An alternative, but equivalent, way to describe $M \otimes N$ is that it is the unique matrix that satisfies the equation

$$(M \otimes N)\big(|\phi\rangle \otimes |\psi\rangle\big) = \big(M|\phi\rangle\big) \otimes \big(N|\psi\rangle\big)$$

for every possible choice of vectors $|\phi\rangle$ and $|\psi\rangle$, assuming that the indices of $|\phi\rangle$ correspond to the elements of Σ and the indices of $|\psi\rangle$ correspond to Γ.

Following the convention described previously for ordering the elements of Cartesian products, we can also write the tensor product of two matrices explicitly as follows:

$$\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mm} \end{pmatrix} \otimes \begin{pmatrix} \beta_{11} & \cdots & \beta_{1k} \\ \vdots & \ddots & \vdots \\ \beta_{k1} & \cdots & \beta_{kk} \end{pmatrix}$$

$$= \begin{pmatrix} \alpha_{11}\beta_{11} & \cdots & \alpha_{11}\beta_{1k} & & \alpha_{1m}\beta_{11} & \cdots \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots \\ \alpha_{11}\beta_{k1} & \cdots & \alpha_{11}\beta_{kk} & & \alpha_{1m}\beta_{k1} & \cdots \\ & \vdots & & \ddots & & \vdots \\ \alpha_{m1}\beta_{11} & \cdots & \alpha_{m1}\beta_{1k} & & \alpha_{mm}\beta_{11} & \cdots \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots \\ \alpha_{m1}\beta_{k1} & \cdots & \alpha_{m1}\beta_{kk} & & \alpha_{mm}\beta_{k1} & \cdots \end{pmatrix}$$

Tensor products of three or more matrices are defined in an analogous way. If $M_0, \ldots, M_{n-1}$ are matrices whose indices correspond to classical state sets $\Sigma_0, \ldots, \Sigma_{n-1}$, then the tensor product $M_{n-1} \otimes \cdots \otimes M_0$ is defined by the condition that

$$\langle a_{n-1} \cdots a_0 | M_{n-1} \otimes \cdots \otimes M_0 | b_{n-1} \cdots b_0 \rangle = \langle a_{n-1} | M_{n-1} | b_{n-1} \rangle \cdots$$

for every choice of classical states $a_0, b_0 \in \Sigma_0, \ldots, a_{n-1}, b_{n-1} \in \Sigma_{n-1}$. Alternatively, tensor products of three or more matrices can be defined recursively, in terms of tensor products of two matrices, similar to what we observed for vectors.

The tensor product of matrices is sometimes said to be *multiplicative* because the equation

$$(M_{n-1} \otimes \cdots \otimes M_0)(N_{n-1} \otimes \cdots \otimes N_0) = (M_{n-1}N_{n-1}) \otimes \cdots \otimes (M$$

is always true, for any choice of matrices $M_0, \ldots, M_{n-1}$ and $N_0 \ldots, N_{n-1}$, provided that the products $M_0 N_0, \ldots, M_{n-1} N_{n-1}$ make sense.

Independent operations (continued)

We can now answer the question asked previously: if M is a probabilistic operation on X, N is a probabilistic operation on Y, and the two operations are performed independently, then the resulting operation on the compound system (X, Y) is the tensor product $M \otimes N$.

So, for both probabilistic states and probabilistic operations, *tensor products represent independence.* If we have two systems X and Y that are independently in the probabilistic states $|\phi\rangle$ and $|\pi\rangle$, then the compound system (X, Y) is in the probabilistic state $|\phi\rangle \otimes |\pi\rangle$; and if we apply probabilistic operations M and N to the two systems independently, then the resulting action on the compound system (X, Y) is described by the operation $M \otimes N$.

Let's take a look at an example, which recalls a probabilistic operation on a single bit from the previous lesson: if the classical state of the bit is 0, it is left alone; and if the classical state of the bit is 1, it is flipped to 0 with probability $1/2$. We observed that this operation is represented by the matrix

$$\begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}.$$

If this operation is performed on a bit X, and a NOT operation is (independently) performed on a second bit Y, then the joint operation on the compound system (X, Y) has the matrix representation

$$\begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

By inspection, we see that this is a stochastic matrix. This will always be the case: the tensor product of two or more stochastic matrices is always stochastic.

A common situation that we encounter is one in which one operation is performed on one system and *nothing* is done to another. In such a case, exactly the same prescription is followed, bearing in mind that *doing nothing* is represented by the identity matrix. For example, resetting the bit X to the 0 state and doing nothing to Y yields the probabilistic (and in fact deterministic) operation on (X, Y) represented by the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Quantum information

We're now prepared to move on to quantum information in the setting of multiple systems. Much like in the previous lesson on single systems, the mathematical description of quantum information for multiple systems is quite similar to the probabilistic case and makes use of similar concepts and techniques.

Quantum states

Multiple systems can be viewed collectively as single, compound systems. We've already observed this in the probabilistic setting, and the quantum setting is analogous. Quantum states of multiple systems are therefore represented by column vectors having complex number entries and Euclidean norm equal to 1, just like quantum states of single

systems. In the multiple system case, the entries of these vectors are placed in correspondence with the *Cartesian product* of the classical state sets associated with each of the individual systems, because that's the classical state set of the compound system.

For instance, if X and Y are qubits, then the classical state set of the pair of qubits (X, Y), viewed collectively as a single system, is the Cartesian product $\{0,1\} \times \{0,1\}$. By representing pairs of binary values as binary strings of length two, we associate this Cartesian product set with the set $\{00, 01, 10, 11\}$. The following vectors are therefore all examples of quantum state vectors of the pair (X, Y) :

$$\frac{1}{\sqrt{2}}|00\rangle - \frac{1}{\sqrt{6}}|01\rangle + \frac{i}{\sqrt{6}}|10\rangle + \frac{1}{\sqrt{6}}|11\rangle, \quad \frac{3}{5}|00\rangle - \frac{4}{5}|11\rangle, \quad \text{and}$$

There are variations on how quantum state vectors of multiple systems are expressed, and we can choose whichever variation suits our preferences. Here are some examples for the first quantum state vector above.

1. We may use the fact that $|ab\rangle = |a\rangle|b\rangle$ (for any classical states a and b) to instead write

$$\frac{1}{\sqrt{2}}|0\rangle|0\rangle - \frac{1}{\sqrt{6}}|0\rangle|1\rangle + \frac{i}{\sqrt{6}}|1\rangle|0\rangle + \frac{1}{\sqrt{6}}|1\rangle|1\rangle.$$

2. We may choose to write the tensor product symbol explicitly like this:

$$\frac{1}{\sqrt{2}}|0\rangle \otimes |0\rangle - \frac{1}{\sqrt{6}}|0\rangle \otimes |1\rangle + \frac{i}{\sqrt{6}}|1\rangle \otimes |0\rangle + \frac{1}{\sqrt{6}}|1\rangle \otimes |1\rangle.$$

3. We may subscript the kets to indicate how they correspond to the systems being considered, like this:

$$\frac{1}{\sqrt{2}}|0\rangle_{\mathsf{X}}|0\rangle_{\mathsf{Y}} - \frac{1}{\sqrt{6}}|0\rangle_{\mathsf{X}}|1\rangle_{\mathsf{Y}} + \frac{i}{\sqrt{6}}|1\rangle_{\mathsf{X}}|0\rangle_{\mathsf{Y}} + \frac{1}{\sqrt{6}}|1\rangle_{\mathsf{X}}|1\rangle_{\mathsf{Y}}.$$

Of course, we may also write quantum state vectors explicitly as column vectors:

$$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} \\ \frac{i}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}.$$

Depending upon the context in which it appears, one of these variations may be preferred — but they are all equivalent in the sense that they describe the same vector.

Tensor products of quantum state vectors

Similar to what we have for probability vectors, tensor products of quantum state vectors are also quantum state vectors — and again they represent *independence* among systems.

In greater detail, and beginning with the case of two systems, suppose that $|\phi\rangle$ is a quantum state vector of a system X and $|\psi\rangle$ is a quantum state vector of a system Y. The tensor product $|\phi\rangle \otimes |\psi\rangle$, which may alternatively be written as $|\phi\rangle|\psi\rangle$ or as $|\phi \otimes \psi\rangle$, is then a quantum state vector of the joint system (X, Y). Again we refer to a state of this form as a being a *product state*. Intuitively speaking, when a pair of systems (X, Y) is in a product state $|\phi\rangle \otimes |\psi\rangle$, we may interpret this as meaning that X is in the quantum state $|\phi\rangle$, Y is in the quantum state $|\psi\rangle$, and the states of the two systems have nothing to do with one another.

The fact that the tensor product vector $|\phi\rangle \otimes |\psi\rangle$ is indeed a quantum state vector is consistent with the Euclidean norm being *multiplicative* with respect to tensor products:

$$\begin{aligned}
\big\| |\phi\rangle \otimes |\psi\rangle \big\| &= \sqrt{\sum_{(a,b)\in\Sigma\times\Gamma} \big|\langle ab|\phi\otimes\psi\rangle\big|^2} \\
&= \sqrt{\sum_{a\in\Sigma}\sum_{b\in\Gamma} \big|\langle a|\phi\rangle\langle b|\psi\rangle\big|^2} \\
&= \sqrt{\left(\sum_{a\in\Sigma} \big|\langle a|\phi\rangle\big|^2\right)\left(\sum_{b\in\Gamma} \big|\langle b|\psi\rangle\big|^2\right)} \\
&= \big\||\phi\rangle\big\|\big\||\psi\rangle\big\|.
\end{aligned}$$

Because $|\phi\rangle$ and $|\psi\rangle$ are quantum state vectors, we have $\big\||\phi\rangle\big\| = 1$ and $\big\||\psi\rangle\big\| = 1$, and therefore $\big\||\phi\rangle \otimes |\psi\rangle\big\| = 1$, so $|\phi\rangle \otimes |\psi\rangle$ is also a quantum state vector.

This discussion may be generalized to more than two systems. If $|\psi_0\rangle, \ldots, |\psi_{n-1}\rangle$ are quantum state vectors of systems $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$, then $|\psi_{n-1}\rangle \otimes \cdots \otimes |\psi_0\rangle$ is a quantum state vector representing a *product state* of the joint system $(\mathsf{X}_{n-1}, \ldots, \mathsf{X}_0)$. Again, we know that this is a quantum state vector because

$$\big\| |\psi_{n-1}\rangle \otimes \cdots \otimes |\psi_0\rangle \big\| = \big\| |\psi_{n-1}\rangle \big\| \cdots \big\| |\psi_0\rangle \big\| = 1^n = 1.$$

Entangled states

Not all quantum state vectors of multiple systems are product states. For example, the quantum state vector

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle \tag{6}$$

of two qubits is not a product state. To reason this, we may follow exactly the same argument that we used to prove that the probabilistic state represented by the vector (5) is not a product state. That is, if (6) were a product state, there would exist quantum state vectors $|\phi\rangle$ and $|\psi\rangle$ for which

$$|\phi\rangle \otimes |\psi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle.$$

But then it would necessarily be the case that

$$\langle 0|\phi\rangle\langle 1|\psi\rangle = \langle 01|\phi \otimes \psi\rangle = 0$$

implying that $\langle 0|\phi\rangle = 0$ or $\langle 1|\psi\rangle = 0$ (or both). That contradicts the fact that

$$\langle 0|\phi\rangle\langle 0|\psi\rangle = \langle 00|\phi \otimes \psi\rangle = \frac{1}{\sqrt{2}}$$

and

$$\langle 1|\phi\rangle\langle 1|\psi\rangle = \langle 11|\phi \otimes \psi\rangle = \frac{1}{\sqrt{2}}$$

are both nonzero.

Notice that the specific value $1/\sqrt{2}$ is not important to this argument — all that is important is that this value is nonzero. Thus, for instance, the quantum state

$$\frac{3}{5}|00\rangle + \frac{4}{5}|11\rangle$$

is also not a product state, by the same argument.

It follows that the quantum state vector (6) represents a *correlation* between two systems, and specifically we say that the systems are *entangled*.

Entanglement is a quintessential feature of quantum information that will be discussed in much greater detail in later lessons. Entanglement can be complicated, particularly for the sorts of noisy quantum states that can be described by density matrices in the general formulation of quantum information, which was mentioned in the previous lesson. For quantum state vectors in the simplified formulation of quantum information that we're focusing on in this course, however, entanglement is equivalent to correlation: any quantum state vector that is not a product state represents an entangled state.

In contrast, the quantum state vector

$$\frac{1}{2}|00\rangle + \frac{i}{2}|01\rangle - \frac{1}{2}|10\rangle - \frac{i}{2}|11\rangle$$

is an example of a product state.

$$\frac{1}{2}|00\rangle + \frac{i}{2}|01\rangle - \frac{1}{2}|10\rangle - \frac{i}{2}|11\rangle = \left(\frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle\right) \otimes \left(\frac{1}{\sqrt{2}}|0\right)$$

Hence, this state is not entangled.

Bell states

We'll now take a look as some important examples of multiple-qubit quantum states, beginning with the *Bell states*. These are the following four two-qubit states:

$$\begin{aligned}
|\phi^+\rangle &= \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle \\
|\phi^-\rangle &= \frac{1}{\sqrt{2}}|00\rangle - \frac{1}{\sqrt{2}}|11\rangle \\
|\psi^+\rangle &= \frac{1}{\sqrt{2}}|01\rangle + \frac{1}{\sqrt{2}}|10\rangle \\
|\psi^-\rangle &= \frac{1}{\sqrt{2}}|01\rangle - \frac{1}{\sqrt{2}}|10\rangle
\end{aligned}$$

The Bell states are so-named in honor of John Bell. Notice that the same argument that establishes that $|\phi^+\rangle$ is not a product state reveals that none of the other Bell states is a product state either: all four of the Bell states represent entanglement between two qubits.

The collection of all four Bell states

$$\left\{|\phi^+\rangle, |\phi^-\rangle, |\psi^+\rangle, |\psi^-\rangle\right\}$$

is known as the *Bell basis.* True to its name, this is a basis; any quantum state vector of two qubits, or indeed any complex vector at all having entries corresponding to the four classical states of two bits, can be expressed as a linear combination of the four Bell states. For example,

$$|00\rangle = \frac{1}{\sqrt{2}}|\phi^+\rangle + \frac{1}{\sqrt{2}}|\phi^-\rangle.$$

GHZ and W states

Next we will consider two interesting examples of states of three qubits. The first example is the *GHZ state* (so named in honor of Daniel Greenberger, Michael Horne, and Anton Zeilinger, who first studied some of its properties):

$$\frac{1}{\sqrt{2}}|000\rangle + \frac{1}{\sqrt{2}}|111\rangle.$$

The second example is the so-called W state:

$$\frac{1}{\sqrt{3}}|001\rangle + \frac{1}{\sqrt{3}}|010\rangle + \frac{1}{\sqrt{3}}|100\rangle.$$

Neither of these states is a product state, meaning that they cannot be written as a tensor product of three qubit quantum state vectors. We'll examine both of these two states further when we discuss partial measurements of quantum states of multiple systems, and they'll arise from time to time throughout this series.

Additional examples

The examples of quantum states of multiple systems we've seen so far are states of two or three qubits, but we can also consider quantum states of multiple systems having different classical state sets.

For example, here's a quantum state of three systems, X, Y, and Z, where the classical state set of X is the binary alphabet (so X is a qubit) and the classical state set of Y and Z is $\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$:

$$\frac{1}{2}|0\rangle|\heartsuit\rangle|\heartsuit\rangle + \frac{1}{2}|1\rangle|\spadesuit\rangle|\heartsuit\rangle - \frac{1}{\sqrt{2}}|0\rangle|\heartsuit\rangle|\diamondsuit\rangle.$$

And here's an example of a quantum state of three systems, X, Y, and Z, that all share the same classical state set $\{0, 1, 2\}$:

$$\frac{|012\rangle - |021\rangle + |120\rangle - |102\rangle + |201\rangle - |210\rangle}{\sqrt{6}}.$$

Systems having the classical state set $\{0, 1, 2\}$ are often called *trits* or (assuming that they can be in a quantum state) *qutrits*. The term *qudit* refers to a system having classical state set $\{0, \ldots, d-1\}$ for an arbitrary choice of d.

Measurements of quantum states

Standard basis measurements of quantum states of single systems were discussed in the previous lesson: if a system having classical state set Σ is in a quantum state represented by the vector $|\psi\rangle$, and that system is measured (with respect to a standard basis measurement), then each classical state $a \in \Sigma$ appears with probability $|\langle a|\psi\rangle|^2$.

This tells us what happens when we have a quantum state of multiple systems and choose to measure the entire compound system (which is equivalent to measuring *all* of the systems). To state this precisely, let us suppose that $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$ are systems having classical state sets $\Sigma_0, \ldots, \Sigma_{n-1}$, respectively. We may then view $(\mathsf{X}_{n-1}, \ldots, \mathsf{X}_0)$ collectively as a single system whose classical state set is the Cartesian product $\Sigma_{n-1} \times \cdots \times \Sigma_0$. If a quantum state of this system is represented by the quantum state vector $|\psi\rangle$, and all of the systems are measured, then each possible outcome $(a_{n-1}, \ldots, a_0) \in \Sigma_{n-1} \times \cdots \times \Sigma_0$ appears with probability $|\langle a_{n-1} \cdots a_0|\psi\rangle|^2$.

For example, if systems X and Y are jointly in the quantum state

$$\frac{3}{5}|0\rangle|\heartsuit\rangle - \frac{4i}{5}|1\rangle|\spadesuit\rangle,$$

then measuring both systems with standard basis measurements yields the outcome $(0, \heartsuit)$ with probability $9/25$ and the outcome $(1, \spadesuit)$ with probability $16/25$.

Partial measurements

Now let us consider the situation in which we have multiple systems in some quantum state, and we measure a proper subset of the systems. As before, we will begin with two systems X and Y having classical state sets Σ and Γ, respectively.

In general, a quantum state vector of (X, Y) takes the form

$$|\psi\rangle = \sum_{(a,b)\in\Sigma\times\Gamma} \alpha_{ab}|ab\rangle,$$

where $\{\alpha_{ab} : (a,b) \in \Sigma \times \Gamma\}$ is a collection of complex numbers satisfying

$$\sum_{(a,b)\in\Sigma\times\Gamma} |\alpha_{ab}|^2 = 1,$$

which is equivalent to $|\psi\rangle$ being a unit vector.

We already know, from the discussion above, that if both X and Y are measured, then each possible outcome $(a,b) \in \Sigma \times \Gamma$ appears with probability

$$\bigl|\langle ab|\psi\rangle\bigr|^2 = |\alpha_{ab}|^2.$$

If we suppose instead that just the first system X is measured, the probability for each outcome $a \in \Sigma$ to appear must therefore be equal to

$$\sum_{b\in\Gamma}\bigl|\langle ab|\psi\rangle\bigr|^2 = \sum_{b\in\Gamma}|\alpha_{ab}|^2.$$

This is consistent with what we already saw in the probabilistic setting, as well as our current understanding of physics; the probability for each outcome to appear when X is measured can't possibly depend on whether or not Y was also measured, as that would allow for faster-than-light communication.

Having obtained a particular outcome $a \in \Sigma$ of a standard basis measurement of X, we naturally expect that the quantum state of X changes so that it is equal to $|a\rangle$, just like we had for single systems. But what happens to the quantum state of Y? To answer this question, we can first express the vector $|\psi\rangle$ as

$$|\psi\rangle = \sum_{a\in\Sigma} |a\rangle \otimes |\phi_a\rangle,$$

where

$$|\phi_a\rangle = \sum_{b\in\Gamma} \alpha_{ab}|b\rangle$$

for each $a \in \Sigma$. Here we're following the same methodology as in the probabilistic case, of isolating the standard basis states of the system being measured. The probability for the standard basis measurement of X to give each outcome a is as follows:

$$\sum_{b \in \Gamma} |\alpha_{ab}|^2 = \big\| |\phi_a\rangle \big\|^2.$$

And, as a result of the standard basis measurement of X giving the outcome a, the quantum state of the pair (X, Y) together becomes

$$|a\rangle \otimes \frac{|\phi_a\rangle}{\big\| |\phi_a\rangle \big\|}.$$

That is, the state "collapses" like in the single-system case, but only as far as is required for the state to be consistent with the measurement of X having produced the outcome a.

Informally speaking, $|a\rangle \otimes |\phi_a\rangle$ represents the component of $|\psi\rangle$ that is consistent with the a measurement of X resulting in the outcome a. We then *normalize* this vector — by dividing it by its Euclidean norm, which is equal to $\big\| |\phi_a\rangle \big\|$ — to obtain a valid quantum state vector having Euclidean norm equal to 1. This normalization step is analogous to what we did in the probabilistic setting when we divided vectors by the sum of their entries to obtain a probability vector.

As an example, consider the state of two qubits (X, Y) from the beginning of the section:

$$|\psi\rangle = \frac{1}{\sqrt{2}}|00\rangle - \frac{1}{\sqrt{6}}|01\rangle + \frac{i}{\sqrt{6}}|10\rangle + \frac{1}{\sqrt{6}}|11\rangle.$$

To understand what happens when the first system X is measured, we begin by writing

$$|\psi\rangle = |0\rangle \otimes \left(\frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{6}}|1\rangle\right) + |1\rangle \otimes \left(\frac{i}{\sqrt{6}}|0\rangle + \frac{1}{\sqrt{6}}|1\rangle\right).$$

We now see, based on the description above, that the probability for the measurement to result in the outcome 0 is

$$\left\| \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{6}}|1\rangle \right\|^2 = \frac{1}{2} + \frac{1}{6} = \frac{2}{3},$$

in which case the state of (X, Y) becomes

$$|0\rangle \otimes \frac{\frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{6}}|1\rangle}{\sqrt{\frac{2}{3}}} = |0\rangle \otimes \left(\frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle\right);$$

and the probability for the measurement to result in the outcome 1 is

$$\left\| \frac{i}{\sqrt{6}}|0\rangle + \frac{1}{\sqrt{6}}|1\rangle \right\|^2 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3},$$

in which case the state of (X, Y) becomes

$$|1\rangle \otimes \frac{\frac{i}{\sqrt{6}}|0\rangle + \frac{1}{\sqrt{6}}|1\rangle}{\sqrt{\frac{1}{3}}} = |1\rangle \otimes \left(\frac{i}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \right).$$

The same technique, used in a symmetric way, describes what happens if the second system Y is measured rather than the first. This time we rewrite the vector $|\psi\rangle$ as

$$|\psi\rangle = \left(\frac{1}{\sqrt{2}}|0\rangle + \frac{i}{\sqrt{6}}|1\rangle \right) \otimes |0\rangle + \left(-\frac{1}{\sqrt{6}}|0\rangle + \frac{1}{\sqrt{6}}|1\rangle \right) \otimes |1\rangle.$$

The probability that the measurement of Y gives the outcome 0 is

$$\left\| \frac{1}{\sqrt{2}}|0\rangle + \frac{i}{\sqrt{6}}|1\rangle \right\|^2 = \frac{1}{2} + \frac{1}{6} = \frac{2}{3},$$

in which case the state of (X, Y) becomes

$$\frac{\frac{1}{\sqrt{2}}|0\rangle + \frac{i}{\sqrt{6}}|1\rangle}{\sqrt{\frac{2}{3}}} \otimes |0\rangle = \left(\frac{\sqrt{3}}{2}|0\rangle + \frac{i}{2}|1\rangle \right) \otimes |0\rangle;$$

and the probability that the measurement outcome is 1 is

$$\left\| -\frac{1}{\sqrt{6}}|0\rangle + \frac{1}{\sqrt{6}}|1\rangle \right\|^2 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3},$$

in which case the state of (X, Y) becomes

$$\frac{-\frac{1}{\sqrt{6}}|0\rangle + \frac{1}{\sqrt{6}}|1\rangle}{\frac{1}{\sqrt{3}}} \otimes |1\rangle = \left(-\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \right) \otimes |1\rangle.$$

Remark on reduced quantum states

The previous example shows a limitation of the simplified description of quantum information, which is that it does not offer us a way to describe the reduced (or marginal) quantum state of just one of two systems (or of a proper subset of any number of systems) like in the probabilistic case.

Specifically, for a probabilistic state of two systems (X, Y) described by a probability vector

$$\sum_{(a,b)\in\Sigma\times\Gamma} p_{ab}|ab\rangle,$$

we can write the *reduced* (or *marginal*) probabilistic state of X alone as

$$\sum_{a\in\Sigma}\left(\sum_{b\in\Gamma} p_{ab}\right)|a\rangle = \sum_{(a,b)\in\Sigma\times\Gamma} p_{ab}|a\rangle.$$

For quantum state vectors, there is no analogous way to do this. In particular, for a quantum state vector

$$|\psi\rangle = \sum_{(a,b)\in\Sigma\times\Gamma} \alpha_{ab}|ab\rangle,$$

the vector

$$\sum_{(a,b)\in\Sigma\times\Gamma} \alpha_{ab}|a\rangle$$

is not a quantum state vector in general, and does not properly represent the concept of a reduced or marginal state.

What we may do instead is turn to the General formulation of quantum information, which is the subject of the third course in the *Understanding Quantum Information and Computation Series.* That formulation provides with a meaningful way to define reduced quantum states that is analogous to the probabilistic setting.

Partial measurements for three or more systems

Partial measurements for three or more systems, where some proper subset of the systems are measured, can be reduced to the case of two systems by dividing the systems into two collections, those that are measured and those that are not. Here is a specific example that illustrates how this can be done. It demonstrates how subscripting kets by the names of the systems they represent can be useful — in this case because it gives us a simple way to describe permutations of the systems.

For this example, we'll consider a quantum state of a 5-tuple of systems $(\mathsf{X}_4, \ldots, \mathsf{X}_0)$, where all five of these systems share the same classical state set $\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$:

$$\sqrt{\frac{1}{7}}|\heartsuit\rangle|\clubsuit\rangle|\diamondsuit\rangle|\spadesuit\rangle|\spadesuit\rangle + \sqrt{\frac{2}{7}}|\diamondsuit\rangle|\clubsuit\rangle|\diamondsuit\rangle|\spadesuit\rangle|\clubsuit\rangle + \sqrt{\frac{1}{7}}|\spadesuit\rangle|\spadesuit\rangle|\clubsuit\rangle|\diamondsuit\rangle$$
$$-i\sqrt{\frac{2}{7}}|\heartsuit\rangle|\clubsuit\rangle|\diamondsuit\rangle|\heartsuit\rangle|\heartsuit\rangle - \sqrt{\frac{1}{7}}|\spadesuit\rangle|\heartsuit\rangle|\clubsuit\rangle|\spadesuit\rangle|\clubsuit\rangle.$$

We'll consider the situation in which the first and third systems are measured, and the remaining systems are left alone.

Conceptually speaking, there's no fundamental difference between this situation and one in which one of two systems is measured. Unfortunately, because the measured systems are interspersed with the unmeasured systems, we face a hurdle in writing down the expressions needed to perform these calculations. One way to proceed, as suggested above, is to subscript the kets to indicate which systems they refer to. This gives us a way to keep track of the systems as we permute the ordering of the kets, which makes the mathematics simpler.

First, the quantum state vector above can alternatively be written as

$$\sqrt{\frac{1}{7}}|\heartsuit\rangle_4|\clubsuit\rangle_3|\diamondsuit\rangle_2|\spadesuit\rangle_1|\spadesuit\rangle_0 + \sqrt{\frac{2}{7}}|\diamondsuit\rangle_4|\clubsuit\rangle_3|\diamondsuit\rangle_2|\spadesuit\rangle_1|\clubsuit\rangle_0$$
$$+\sqrt{\frac{1}{7}}|\spadesuit\rangle_4|\spadesuit\rangle_3|\clubsuit\rangle_2|\diamondsuit\rangle_1|\clubsuit\rangle_0 - i\sqrt{\frac{2}{7}}|\heartsuit\rangle_4|\clubsuit\rangle_3|\diamondsuit\rangle_2|\heartsuit\rangle_1|\heartsuit\rangle_0$$
$$-\sqrt{\frac{1}{7}}|\spadesuit\rangle_4|\heartsuit\rangle_3|\clubsuit\rangle_2|\spadesuit\rangle_1|\clubsuit\rangle_0.$$

Nothing has changed, except that each ket now has a subscript indicating which system it corresponds to. Here we've used the subscripts $0, \ldots, 4$, but the names of the systems themselves could also be used (in a situation where we have system names such as X, Y, and Z, for instance).

We can now re-order the kets and collect terms as follows:

$$\begin{aligned}
&\sqrt{\frac{1}{7}}|\heartsuit\rangle_4|\diamondsuit\rangle_2|\clubsuit\rangle_3|\spadesuit\rangle_1|\spadesuit\rangle_0 + \sqrt{\frac{2}{7}}|\diamondsuit\rangle_4|\diamondsuit\rangle_2|\clubsuit\rangle_3|\spadesuit\rangle_1|\clubsuit\rangle_0 \\
&\quad + \sqrt{\frac{1}{7}}|\spadesuit\rangle_4|\clubsuit\rangle_2|\spadesuit\rangle_3|\diamondsuit\rangle_1|\clubsuit\rangle_0 - i\sqrt{\frac{2}{7}}|\heartsuit\rangle_4|\diamondsuit\rangle_2|\clubsuit\rangle_3|\heartsuit\rangle_1|\heartsuit\rangle_0 \\
&\quad - \sqrt{\frac{1}{7}}|\spadesuit\rangle_4|\clubsuit\rangle_2|\heartsuit\rangle_3|\spadesuit\rangle_1|\clubsuit\rangle_0 \\
&\qquad = |\heartsuit\rangle_4|\diamondsuit\rangle_2\left(\sqrt{\frac{1}{7}}|\clubsuit\rangle_3|\spadesuit\rangle_1|\spadesuit\rangle_0 - i\sqrt{\frac{2}{7}}|\clubsuit\rangle_3|\heartsuit\rangle_1|\heartsuit\rangle_0\right) \\
&\qquad\quad + |\diamondsuit\rangle_4|\diamondsuit\rangle_2\left(\sqrt{\frac{2}{7}}|\clubsuit\rangle_3|\spadesuit\rangle_1|\clubsuit\rangle_0\right) \\
&\qquad\quad + |\spadesuit\rangle_4|\clubsuit\rangle_2\left(\sqrt{\frac{1}{7}}|\spadesuit\rangle_3|\diamondsuit\rangle_1|\clubsuit\rangle_0 - \sqrt{\frac{1}{7}}|\heartsuit\rangle_3|\spadesuit\rangle_1|\clubsuit\rangle_0\right)
\end{aligned}$$

The tensor products are still implicit, even when parentheses are used, as in this example.

To be clear about permuting the kets, tensor products are not commutative: if $|\phi\rangle$ and $|\pi\rangle$ are vectors, then, in general, $|\phi\rangle \otimes |\pi\rangle$ is different from $|\pi\rangle \otimes |\phi\rangle$, and likewise for tensor products of three or more vectors. For instance, $|\heartsuit\rangle|\clubsuit\rangle|\diamondsuit\rangle|\spadesuit\rangle|\spadesuit\rangle$ is a different vector than $|\heartsuit\rangle|\diamondsuit\rangle|\clubsuit\rangle|\spadesuit\rangle|\spadesuit\rangle$. Re-ordering the kets as we have just done should not be interpreted as suggesting otherwise. Rather, for the sake of performing calculations, we're simply making a decision that it's more convenient to collect the systems together as $(\mathsf{X}_4, \mathsf{X}_2, \mathsf{X}_3, \mathsf{X}_1, \mathsf{X}_0)$ rather than $(\mathsf{X}_4, \mathsf{X}_3, \mathsf{X}_2, \mathsf{X}_1, \mathsf{X}_0)$. The subscripts on the kets serve to keep this all straight, and we're free to revert back to the original order later if we wish to do that.

We now see that, if the systems X_4 and X_2 are measured, the (nonzero) probabilities of the different outcomes are as follow:

– The measurement outcome $(\heartsuit, \diamondsuit)$ occurs with probability

$$\left\| \sqrt{\frac{1}{7}}|\clubsuit\rangle_3|\spadesuit\rangle_1|\spadesuit\rangle_0 - i\sqrt{\frac{2}{7}}|\clubsuit\rangle_3|\heartsuit\rangle_1|\heartsuit\rangle_0 \right\|^2 = \frac{1}{7} + \frac{2}{7} = \frac{3}{7}$$

– The measurement outcome $(\diamondsuit, \diamondsuit)$ occurs with probability

$$\left\| \sqrt{\frac{2}{7}}|\clubsuit\rangle_3|\spadesuit\rangle_1|\clubsuit\rangle_0 \right\|^2 = \frac{2}{7}$$

– The measurement outcome $(\spadesuit, \clubsuit)$ occurs with probability

$$\left\| \sqrt{\frac{1}{7}} |\spadesuit\rangle_3 |\diamondsuit\rangle_1 |\clubsuit\rangle_0 - \sqrt{\frac{1}{7}} |\heartsuit\rangle_3 |\spadesuit\rangle_1 |\clubsuit\rangle_0 \right\|^2 = \frac{1}{7} + \frac{1}{7} = \frac{2}{7}.$$

If the measurement outcome is $(\heartsuit, \diamondsuit)$, for instance, the resulting state of our five systems becomes

$$|\heartsuit\rangle_4 |\diamondsuit\rangle_2 \otimes \frac{\sqrt{\frac{1}{7}} |\clubsuit\rangle_3 |\spadesuit\rangle_1 |\spadesuit\rangle_0 - i\sqrt{\frac{2}{7}} |\clubsuit\rangle_3 |\heartsuit\rangle_1 |\heartsuit\rangle_0}{\sqrt{\frac{3}{7}}}$$

$$= \sqrt{\frac{1}{3}} |\heartsuit\rangle_4 |\clubsuit\rangle_3 |\diamondsuit\rangle_2 |\spadesuit\rangle_1 |\spadesuit\rangle_0 - i\sqrt{\frac{2}{3}} |\heartsuit\rangle_4 |\clubsuit\rangle_3 |\diamondsuit\rangle_2 |\heartsuit\rangle_1 |\heartsuit\rangle_0.$$

Here, for the final answer, we've reverted back to our original ordering of the systems, just to illustrate that we can do this. For the other possible measurement outcomes, the state can be determined in a similar way.

Finally, here are two examples promised earlier, beginning with the GHZ state

$$\frac{1}{\sqrt{2}} |000\rangle + \frac{1}{\sqrt{2}} |111\rangle.$$

If just the first system is measured, we obtain the outcome 0 with probability $1/2$, in which case the state of the three qubits becomes $|000\rangle$; and we also obtain the outcome 1 with probability $1/2$, in which case the state of the three qubits becomes $|111\rangle$.

For a W state, on the other hand, assuming again that just the first system is measured, we begin by writing this state like this:

$$\frac{1}{\sqrt{3}} |001\rangle + \frac{1}{\sqrt{3}} |010\rangle + \frac{1}{\sqrt{3}} |100\rangle$$

$$= |0\rangle \left(\frac{1}{\sqrt{3}} |01\rangle + \frac{1}{\sqrt{3}} |10\rangle \right) + |1\rangle \left(\frac{1}{\sqrt{3}} |00\rangle \right).$$

The probability that a measurement of the first qubit results in the outcome 0 is therefore equal to

$$\left\| \frac{1}{\sqrt{3}} |01\rangle + \frac{1}{\sqrt{3}} |10\rangle \right\|^2 = \frac{2}{3},$$

and conditioned upon the measurement producing this outcome, the quantum state of the three qubits becomes

$$|0\rangle \otimes \frac{\frac{1}{\sqrt{3}}|01\rangle + \frac{1}{\sqrt{3}}|10\rangle}{\sqrt{\frac{2}{3}}} = |0\rangle \left(\frac{1}{\sqrt{2}}|01\rangle + \frac{1}{\sqrt{2}}|10\rangle \right) = |0\rangle|\psi^+\rangle.$$

The probability that the measurement outcome is 1 is $1/3$, in which case the state of the three qubits becomes $|100\rangle$.

The W state is symmetric, in the sense that it doens't change if we permute the qubits. We therefore obtain a similar description for measuring the second or third qubit rather than the first.

Unitary operations

In principle, any unitary matrix whose rows and columns correspond to the classical states of a system represents a valid quantum operation on that system. This, of course, holds true for compound systems, whose classical state sets happen to be Cartesian products of the classical state sets of the individual systems.

Focusing in on two systems, if X is a system having classical state set Σ, and Y is a system having classical state set Γ, then the classical state set of the joint system (X, Y) is $\Sigma \times \Gamma$. Therefore, quantum operations on this joint system are represented by unitary matrices whose rows and columns are placed in correspondence with the set $\Sigma \times \Gamma$. The ordering of the rows and columns of these matrices is the same as the ordering used for quantum state vectors of the system (X, Y).

For example, let us suppose that $\Sigma = \{1, 2, 3\}$ and $\Gamma = \{0, 1\}$, and recall that the standard convention for ordering the elements of the Cartesian product $\{1, 2, 3\} \times \{0, 1\}$ is this:

$$(1,0),\ (1,1),\ (2,0),\ (2,1),\ (3,0),\ (3,1).$$

Here's an example of a unitary matrix representing an operation on (X, Y) :

$$U = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{i}{2} & -\frac{1}{2} & 0 & 0 & -\frac{i}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & -\frac{i}{2} & -\frac{1}{2} & 0 & 0 & \frac{i}{2} \\ 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}.$$

This unitary matrix isn't special, it's just an example. To check that U is unitary, it suffices to compute and check that $U^\dagger U = \mathbb{I}$, for instance. Alternatively, we can check that the rows (or the columns) are orthonormal, which is made simpler in this case given the particular form of the matrix U.

The action of U on the standard basis vector $|1,1\rangle$, for instance, is

$$U|1,1\rangle = \frac{1}{2}|1,0\rangle + \frac{i}{2}|1,1\rangle - \frac{1}{2}|2,0\rangle - \frac{i}{2}|3,0\rangle,$$

which we can see by examining the second column of U, considering our ordering of the set $\{1,2,3\} \times \{0,1\}$.

As with any matrix, it is possible to express U using Dirac notation, which would require 20 terms for the 20 nonzero entries of U. If we did write down all of these terms, however, rather than writing a 6×6 matrix, it would be messy and the patterns that are evident from the matrix expression would not likely be as clear. Simply put, Dirac notation is not always the best choice.

Unitary operations on three or more systems work in a similar way, with the unitary matrices having rows and columns corresponding to the Cartesian product of the classical state sets of the systems. We've already seen one example in this lesson: the three-qubit operation

$$\sum_{k=0}^{7} |(k+1) \bmod 8\rangle\langle k|,$$

where numbers in bras and kets mean their 3-bit binary encodings. In addition to being a deterministic operation, this is also a unitary operation. Operations that are both deterministic and unitary are called *reversible* operations. The conjugate transpose of this matrix can be written like this:

$$\sum_{k=0}^{7} |k\rangle\langle(k+1) \bmod 8| = \sum_{k=0}^{7} |(k-1) \bmod 8\rangle\langle k|.$$

This represents the *reverse*, or in mathematical terms the *inverse*, of the original operation — which is what we expect from the conjugate transpose of a unitary matrix. We'll see other examples of unitary operations on multiple systems as the lesson continues.

Unitary operations performed independently on individual systems

When unitary operations are performed independently on a collection of individual systems, the combined action of these independent operations is described by the tensor product of the unitary matrices that represent them. That is, if $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$ are quantum systems, $U_0, \ldots, U_{n-1}$ are unitary matrices representing operations on these systems, and the operations are performed independently on the systems, the combined action on $(\mathsf{X}_{n-1}, \ldots, \mathsf{X}_0)$ is represented by the matrix $U_{n-1} \otimes \cdots \otimes U_0$. Once again, we find that the probabilistic and quantum settings are analogous in this regard.

One would naturally expect, from reading the previous paragraph, that the tensor product of any collection of unitary matrices is unitary. Indeed this is true, and we can verify it as follows.

Notice first that the conjugate transpose operation satisfies

$$(M_{n-1} \otimes \cdots \otimes M_0)^\dagger = M_{n-1}^\dagger \otimes \cdots \otimes M_0^\dagger$$

for any chosen matrices $M_0, \ldots, M_{n-1}$. This can be checked by going back to the definition of the tensor product and of the conjugate transpose, and checking that each entry of the two sides of the equation are in agreement. This means that

$$(U_{n-1} \otimes \cdots \otimes U_0)^\dagger (U_{n-1} \otimes \cdots \otimes U_0) = (U_{n-1}^\dagger \otimes \cdots \otimes U_0^\dagger)(U_{n-1} \otimes$$

Because the tensor product of matrices is multiplicative, we find that

$$(U_{n-1}^\dagger \otimes \cdots \otimes U_0^\dagger)(U_{n-1} \otimes \cdots \otimes U_0) = (U_{n-1}^\dagger U_{n-1}) \otimes \cdots \otimes (U_0^\dagger U$$

Here we have written $\mathbb{I}_0, \ldots, \mathbb{I}_{n-1}$ to refer to the matrices representing the identity operation on the systems $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$, which is to say that these are identity matrices whose sizes agree with the number of classical states of $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$.

Finally, the tensor product $\mathbb{I}_{n-1} \otimes \cdots \otimes \mathbb{I}_0$ is equal to the identity matrix for which we have a number of rows and columns that agrees with the product of the number of rows and columns of the matrices $\mathbb{I}_{n-1}, \ldots, \mathbb{I}_0$. This larger identity matrix represents the identity operation on the joint system $(\mathsf{X}_{n-1}, \ldots, \mathsf{X}_0)$.

In summary, we have the following sequence of equalities:

$$\begin{aligned}
&(U_{n-1} \otimes \cdots \otimes U_0)^\dagger (U_{n-1} \otimes \cdots \otimes U_0) \\
&\quad = (U_{n-1}^\dagger \otimes \cdots \otimes U_0^\dagger)(U_{n-1} \otimes \cdots \otimes U_0) \\
&\quad = (U_{n-1}^\dagger U_{n-1}) \otimes \cdots \otimes (U_0^\dagger U_0) \\
&\quad = \mathbb{I}_{n-1} \otimes \cdots \otimes \mathbb{I}_0 \\
&\quad = \mathbb{I}.
\end{aligned}$$

We therefore conclude that $U_{n-1} \otimes \cdots \otimes U_0$ is unitary.

An important situation that often arises is one in which a unitary operation is applied to just one system — or a proper subset of systems — within a larger joint system. For instance, suppose that X and Y are systems that we can view together as forming a single, compound system (X, Y), and we perform an operation just on the system X. To be precise, let us suppose that U is a unitary matrix representing an operation on X, so that its rows and columns have been placed in correspondence with the classical states of X.

To say that we perform the operation represented by U just on the system X implies that we do nothing to Y, meaning that we independently perform U on X and the *identity operation* on Y. That is, "doing nothing" to Y is equivalent to performing the identity operation on Y, which is represented by the identity matrix $\mathbb{I}_{\mathsf{Y}}$. (Here, by the way, the subscript Y tells us that $\mathbb{I}_{\mathsf{Y}}$ refers to the identity matrix having a number of rows and columns in agreement with the classical state set of Y.) The operation on (X, Y) that is obtained when we perform U on X and do nothing to Y is therefore represented by the unitary matrix

$$U \otimes \mathbb{I}_{\mathsf{Y}}.$$

For example, if X and Y are qubits, performing a Hadamard operation on X (and doing nothing to Y) is equivalent to performing the operation

$$H \otimes \mathbb{I}_{\mathsf{Y}} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

on the joint system (X, Y).

Along similar lines, if an operation represented by a unitary matrix U is applied to Y and nothing is done to X, the resulting operation on (X, Y) is represented by the unitary matrix

$$\mathbb{I}_{\mathsf{X}} \otimes U.$$

For example, if we again consider the situation in which both X and Y are qubits and U is a Hadamard operation, the resulting operation on (X, Y) is represented by the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Not every unitary operation on a collection of systems can be written as a tensor product of unitary operations like this, just as not every quantum state vector of these systems is a product state. For example, neither the swap operation nor the controlled-NOT operation on two qubits, which are described below, can be expressed as a tensor product of unitary operations.

The swap operation

To conclude the lesson, let's take a look at two classes of examples of unitary operations on multiple systems, beginning with the *swap operation*.

Suppose that X and Y are systems that share the same classical state set Σ. The *swap* operation on the pair (X, Y) is the operation that exchanges the contents of the two systems, but otherwise leaves the systems alone — so that X remains on the left and Y remains on the right. We'll denote this operation as SWAP, and it operates like this for every choice of classical states $a, b \in \Sigma$:

$$\mathrm{SWAP}\, |a\rangle |b\rangle = |b\rangle |a\rangle.$$

One way to write the matrix associated with this operation using the Dirac notation is as follows:

$$\mathrm{SWAP} = \sum_{c,d \in \Sigma} |c\rangle\langle d| \otimes |d\rangle\langle c|.$$

It may not be immediately clear that this matrix represents SWAP, but we can check it satisfies the condition $\mathrm{SWAP}\, |a\rangle |b\rangle = |b\rangle |a\rangle$ for every choice of classical states $a, b \in \Sigma$. As a simple example, when X and Y are qubits, we find that

$$\text{SWAP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Controlled-unitary operations

Now let us suppose that Q is a qubit and R is an arbitrary system, having whatever classical state set we wish. For every unitary operation U acting on the system R, a *controlled* U operation is a unitary operation on the pair (Q, R) defined as follows:

$$CU = |0\rangle\langle 0| \otimes \mathbb{I}_{\mathsf{R}} + |1\rangle\langle 1| \otimes U.$$

For example, if R is also a qubit, and we consider the Pauli X operation on R, then a controlled-X operation is given by

$$CX = |0\rangle\langle 0| \otimes \mathbb{I}_{\mathsf{R}} + |1\rangle\langle 1| \otimes X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

We already encountered this operation in the context of classical information and probabilistic operations earlier in the lesson. Replacing the Pauli X operation on R with a Z operation gives operation:

$$CZ = |0\rangle\langle 0| \otimes \mathbb{I}_{\mathsf{R}} + |1\rangle\langle 1| \otimes Z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

If instead we take R to be two qubits, and we take U to be the *swap operation* between these two qubits, we obtain this operation:

$$\text{CSWAP} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

This operation is also known as a *Fredkin operation* (or, more commonly, a *Fredkin gate*), named for Edward Fredkin. Its action on standard basis states can be described as follows:

$$\text{CSWAP}\,|0bc\rangle = |0bc\rangle$$

$$\text{CSWAP}\,|1bc\rangle = |1cb\rangle$$

Finally, a *controlled-controlled-NOT operation*, which we may denote as CCX, is called a *Toffoli operation* (or *Toffoli gate*), named for Tommaso Toffoli. Its matrix representation looks like this:

$$CCX = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

We may alternatively express it using the Dirac notation as follows:

$$CCX = \big(|00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 10|\big) \otimes \mathbb{I} + |11\rangle\langle 11| \otimes X.$$

Qiskit implementations

In the previous lesson, we took a first look at Qiskit's `Statevector` and `Operator` classes, and used them to simulate operations and measurements on single qubits. In this section, we'll use these classes to explore the behavior of multiple qubits.

```
from qiskit import __version__
print(__version__)
```

Output:

```
1.3.1
```

We'll start by importing the `Statevector` and `Operator` classes, as well as the square root function from `NumPy`. Hereafter, generally speaking, we'll take care of all of our required imports first within each lesson.

```
from qiskit.quantum_info import Statevector, Oper
from numpy import sqrt
```

No output produced

Tensor products

The `Statevector` class has a `tensor` method, which returns the tensor product of that `Statevector` with another, given as an argument. The argument is interpreted as the tensor factor on the right.

For example, below we create two state vectors representing $|0\rangle$ and $|1\rangle$, and use the `tensor` method to create a new vector, $|\psi\rangle = |0\rangle \otimes |1\rangle$. Notice here that we're using the `from_label` method to define the states $|0\rangle$ and $|1\rangle$, rather than defining them ourselves.

```
zero = Statevector.from_label("0")
one = Statevector.from_label("1")
psi = zero.tensor(one)
display(psi.draw("latex"))
```

Output:

$$|01\rangle$$

Other allowed labels include "+" and "-" for the plus and minus states, as well as "r" and "l" (short for "right" and "left") for the states

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

Here's an example, of the tensor product of $|+\rangle$ and $|-i\rangle$.

```
plus = Statevector.from_label("+")
minus_i = Statevector.from_label("l")
phi = plus.tensor(minus_i)
display(phi.draw("latex"))
```

Output:

$$\frac{1}{2}|00\rangle - \frac{i}{2}|01\rangle + \frac{1}{2}|10\rangle - \frac{i}{2}|11\rangle$$

An alternative is to use the `^` operation for tensor products, which naturally gives the same results.

```
display((plus ^ minus_i).draw("latex"))
```

Output:

$$\frac{1}{2}|00\rangle - \frac{i}{2}|01\rangle + \frac{1}{2}|10\rangle - \frac{i}{2}|11\rangle$$

The `Operator` class also has a `tensor` method (as well as a `from_label` method), as we see in the following examples.

```
H = Operator.from_label("H")
I = Operator.from_label("I")
X = Operator.from_label("X")
display(H.tensor(I).draw("latex"))
display(H.tensor(I).tensor(X).draw("latex"))
```

Output:

$$\begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

$$\begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 \\ \frac{\sqrt{2}}{2} & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & -\frac{\sqrt{2}}{2} & 0 & 0 \\ \frac{\sqrt{2}}{2} & 0 & 0 & 0 & -\frac{\sqrt{2}}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & -\frac{\sqrt{2}}{2} & 0 \end{bmatrix}$$

Again, like in the vector case, the `^` operation is equivalent.

```
display((H ^ I ^ X).draw("latex"))
```

Output:

$$\begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 \\ \frac{\sqrt{2}}{2} & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & -\frac{\sqrt{2}}{2} & 0 & 0 \\ \frac{\sqrt{2}}{2} & 0 & 0 & 0 & -\frac{\sqrt{2}}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & -\frac{\sqrt{2}}{2} & 0 \end{bmatrix}$$

Compound states can be evolved using compound operations as we would expect — just like we saw for single systems in the previous lesson. For example, the following code computes the state $(H \otimes I)|\phi\rangle$ for $|\phi\rangle = |+\rangle \otimes |-i\rangle$ (which was already defined above).

```
display(phi.evolve(H ^ I).draw("latex"))
```

Output:

$$\frac{\sqrt{2}}{2}|00\rangle - \frac{\sqrt{2}i}{2}|01\rangle$$

Here is some code that defines a CX operation and calculates $CX|\psi\rangle$ for $|\psi\rangle = |+\rangle \otimes |0\rangle$. To be clear, this is a CX operation for which the left-hand qubit is the control and the right-hand qubit is the target. The result is the Bell state $|\phi^+\rangle$.

```
CX = Operator(
    [[1, 0, 0, 0],
     [0, 1, 0, 0],
     [0, 0, 0, 1],
     [0, 0, 1, 0]])
psi = plus.tensor(zero)
display(psi.evolve(CX).draw("latex"))
```

Output:

$$\frac{\sqrt{2}}{2}|00\rangle + \frac{\sqrt{2}}{2}|11\rangle$$

☐ Complete

Single systems Quantum circuits