Classical information

To describe quantum information and how it works, we will begin with an overview of classical information. It is natural to wonder why so much attention is paid to classical information in a course on quantum information, but there are good reasons.

For one, although quantum and classical information are different in some spectacular ways, their mathematical descriptions are actually quite similar. Classical information also serves as a familiar point of reference when studying quantum information, as well as a source of analogy that goes a surprisingly long way. It is common that people ask questions about quantum information that have natural classical analogs, and often those questions have simple answers that can provide both clarity and insight into the original questions about quantum information. Indeed, it is not at all unreasonable to claim that one cannot truly understand quantum information without understanding classical information.

Some readers may already be familiar with the material to be discussed in this section, while others may not — but the discussion is meant for both audiences. In addition to highlighting the aspects of classical information that are most relevant to an introduction to quantum information, this section introduces the *Dirac notation*, which is often used to describe vectors and matrices in quantum information and computation. As it turns out, the Dirac notation is not specific to quantum information; it can equally well be used in the context of classical information, as well as for many other settings in which vectors and matrices arise.

Classical states and probability vectors

Suppose that we have a system that stores information. More specifically, we shall assume that this system can be in one of a finite number of *classical states* at each instant. Here, the term *classical state* should be

understood in intuitive terms, as a configuration that can be recognized and described unambiguously.

The archetypal example, which we will come back to repeatedly, is that of a *bit*, which is a system whose classical states are 0 and 1. Other examples include a standard six-sided die, whose classical states are 1, 2, 3, 4, 5, and 6 (represented by the corresponding number of dots on whatever face is on top); a nucleobase in a strand of DNA, whose classical states are *A*, *C*, *G*, and *T;* and a switch on an electric fan, whose classical states are (commonly) *high*, *medium*, *low*, and *off*. In mathematical terms, the specification of the classical states of a system are, in fact, the starting point: we *define* a bit to be a system that has classical states 0 and 1, and likewise for systems having different classical state sets.

For the sake of this discussion, let us give the name X to the system being considered, and let us use the symbol Σ to refer to the set of classical states of X. In addition to the assumption that Σ is finite, which was already mentioned, we naturally assume that Σ is *nonempty* — for it is nonsensical for a physical system to have no states at all. And while it does make sense to consider physical systems having *infinitely* many classical states, we will disregard this possibility, which is certainly interesting but is not relevant to this course. For these reasons, and for the sake of convenience and brevity, we will hereafter use the term *classical state set* to mean any finite and nonempty set.

Here are a few examples:

1. If X is a bit, then $\Sigma = \{0, 1\}$. In words, we refer to this set as the *binary alphabet*.
2. If X is a six-sided die, then $\Sigma = \{1, 2, 3, 4, 5, 6\}$.
3. If X is an electric fan switch, then $\Sigma = \{\text{high}, \text{medium}, \text{low}, \text{off}\}$.

When thinking about X as a carrier of information, the different classical states of X could be assigned certain meanings, leading to different outcomes or consequences. In such cases, it may be sufficient to describe X as simply being in one of its possible classical states. For instance, if X is a fan switch, we might happen to know with certainty that it is set to *high,* which might then lead us to switch it to *medium.*

Often in information processing, however, our knowledge is uncertain. One way to represent our knowledge of the classical state of a system X is to associate *probabilities* with its different possible classical states, resulting in what we shall call a *probabilistic state*.

For example, suppose X is a bit. Based on what we know or expect about what has happened to X in the past, we might perhaps believe that X is in the classical state 0 with probability $3/4$ and in the state 1 with probability $1/4$. We may represent these beliefs by writing this:

$$\operatorname{Pr}(\mathsf{X} = 0) = \frac{3}{4} \quad \text{and} \quad \operatorname{Pr}(\mathsf{X} = 1) = \frac{1}{4}.$$

A more succinct way to represent this probabilistic state is by a column vector.

$$\begin{pmatrix} \frac{3}{4} \\ \\ \frac{1}{4} \end{pmatrix}$$

The probability of the bit being 0 is placed at the top of the vector and the probability of the bit being 1 is placed at the bottom, because this is the conventional way to order the set $\{0, 1\}$.

In general, we can represent a probabilistic state of a system having any classical state set in the same way, as a vector of probabilities. The probabilities can be ordered in any way we choose, but it is typical that there is a natural or default way to do this. To be precise, we can represent any probabilistic state through a column vector satisfying two properties:

1. All entries of the vector are *nonnegative real numbers*.
2. The sum of the entries is equal to 1.

Conversely, any column vector that satisfies these two properties can be taken as a representation of a probabilistic state. Hereafter, we will refer to vectors of this form as *probability vectors*.

Alongside the succinctness of this notation, identifying probabilistic states as column vectors has the advantage that operations on probabilistic states are represented through matrix–vector multiplication, as will be discussed shortly.

Measuring probabilistic states

Next let us consider what happens if we *measure* a system when it is in a probabilistic state. In this context, by measuring a system we simply mean that we look at the system and recognize whatever classical state it is in without unambiguity. Intuitively speaking, we can't "see" a

probabilistic state of a system; when we look at it, we just see one of the possible classical states.

By measuring a system, we may also change our knowledge of it, and therefore the probabilistic state we associate with it can change. That is, if we recognize that X is in the classical state $a \in \Sigma$, then the new probability vector representing our knowledge of the state of X becomes the vector having a 1 in the entry corresponding to a and 0 for all other entries. This vector indicates that X is in the classical state a with certainty — which we know having just recognized it — and we denote this vector by $|a\rangle$, which is read as "ket a" for a reason that will be explained shortly. Vectors of this sort are also called *standard basis* vectors.

For example, assuming that the system we have in mind is a bit, the standard basis vectors are given by

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Notice that any two-dimensional column vector can be expressed as a linear combination of these two vectors. For example,

$$\begin{pmatrix} \frac{3}{4} \\ \frac{1}{4} \end{pmatrix} = \frac{3}{4}|0\rangle + \frac{1}{4}|1\rangle.$$

This fact naturally generalizes to any classical state set: any column vector can be written as a linear combination of standard basis states. Quite often we express vectors in precisely this way.

Returning to the change of a probabilistic state upon being measured, we may note the following connection to our everyday experiences. Suppose we flip a fair coin, but cover up the coin before looking at it. We would then say that its probabilistic state is

$$\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \frac{1}{2}|\text{heads}\rangle + \frac{1}{2}|\text{tails}\rangle.$$

Here, the classical state set of our coin is $\{\text{heads}, \text{tails}\}$. We'll choose to order these states as heads first, tails second.

$$|\text{heads}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |\text{tails}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

If we were to uncover the coin and look at it, we would see one of the two classical states: heads or tails. Supposing that the result were tails, we

would naturally update our description of the probabilistic state of the coin so that it becomes $|\text{tails}\rangle$. Of course, if we were then to cover up the coin, and then uncover it and look at it again, the classical state would still be tails, which is consistent with the probabilistic state being described by the vector $|\text{tails}\rangle$.

This may seem trivial, and in some sense it is. However, while quantum systems behave in an entirely analogous way, their measurement properties are frequently considered strange or unusual. By establishing the analogous properties of classical systems, the way quantum information works might seem less unusual.

One final remark concerning measurements of probabilistic states is this: probabilistic states describe knowledge or belief, not necessarily something actual, and measuring merely changes our knowledge and not the system itself. For instance, the state of a coin after we flip it, but before we look, is either heads or tails — we just don't know which until we look. Upon seeing that the classical state is tails, say, we would naturally update the vector describing our knowledge to $|\text{tails}\rangle$, but to someone else who didn't see the coin when it was uncovered, the probabilistic state would remain unchanged. This is not a cause for concern; different individuals may have different knowledge or beliefs about a particular system, and hence describe that system by different probability vectors.

Classical operations

In the last part of this brief summary of classical information, we will consider the sorts of operations that can be performed on a classical system.

Deterministic operations

First, there are deterministic operations, where each classical state $a \in \Sigma$ is transformed into $f(a)$ for some function f of the form $f : \Sigma \to \Sigma$.

For example, if $\Sigma = \{0, 1\}$, there are four functions of this form, f_1, f_2, f_3, and f_4, which can be represented by tables of values as follows:

a	$f_1(a)$
0	0
1	0

a	$f_2(a)$
0	0
1	1

a	$f_3(a)$
0	1
1	0

a	$f_4(a)$
0	1
1	1

The first and last of these functions are *constant:* $f_1(a) = 0$ and $f_4(a) = 1$ for each $a \in \Sigma$. The middle two are not constant, they are *balanced:* each of the two output values occurs the same number of times (once, in this case) as we range over the possible inputs. The function f_2 is the identity function: $f_2(a) = a$ for each $a \in \Sigma$. And f_3 is the function $f_3(0) = 1$ and $f_3(1) = 0$, which is better-known as the NOT function.

The actions of deterministic operations on probabilistic states can be represented by matrix-vector multiplication. Specifically, the matrix M that represents a given function $f : \Sigma \to \Sigma$ is the one that satisfies

$$M|a\rangle = |f(a)\rangle$$

for every $a \in \Sigma$. Such a matrix always exists and is uniquely determined by this requirement. Matrices that represent deterministic operations always have exactly one 1 in each column, and 0 for all other entries.

For instance, the matrices $M_1, \ldots, M_4$ corresponding to the functions $f_1, \ldots, f_4$ above are as follows:

$$M_1 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad M_4 = \begin{pmatrix}$$

Here's a quick verification showing that the first matrix is correct. The other three can be checked similarly.

$$M_1|0\rangle = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |0\rangle = |f_1(0)\rangle$$

$$M_1|1\rangle = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |0\rangle = |f_1(1)\rangle$$

A convenient way to represent matrices of these and other forms makes use of an analogous notation for row vectors to the one for column vectors discussed previously: we denote by $\langle a|$ the *row* vector having a 1 in the entry corresponding to a and zero for all other entries, for each $a \in \Sigma$. This vector is read as "bra a."

For example, if $\Sigma = \{0, 1\}$, then

$$\langle 0| = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad \text{and} \quad \langle 1| = \begin{pmatrix} 0 & 1 \end{pmatrix}.$$

For any classical state set Σ, we can view row vectors and column vectors as matrices, and perform the matrix multiplication $|b\rangle\langle a|$. We obtain a square matrix having a 1 in the entry corresponding to the pair (b, a), meaning that the row of the entry corresponds to the classical state b and the column corresponds to the classical state a, with 0 for all other entries. For example,

$$|0\rangle\langle 1| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Using this notation, we may express the matrix M that corresponds to any given function $f : \Sigma \to \Sigma$ as

$$M = \sum_{a \in \Sigma} |f(a)\rangle\langle a|.$$

For example, consider the function f_4 above, for which $\Sigma = \{0, 1\}$. We obtain the matrix

$$M_4 = |f_4(0)\rangle\langle 0| + |f_4(1)\rangle\langle 1| = |1\rangle\langle 0| + |1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The reason why this works is as follows. If we again think about vectors as matrices, and this time consider the multiplication $\langle a||b\rangle$, we obtain a 1×1 matrix, which we can think about as a scalar (that is, a number). For the sake of tidiness, we write this product as $\langle a|b\rangle$ rather than $\langle a||b\rangle$. This product satisfies the following simple formula:

$$\langle a|b\rangle = \begin{cases} 1 & a = b \\ 0 & a \neq b. \end{cases}$$

Using this observation, together with the fact that matrix multiplication is associative and linear, we obtain

$$M|b\rangle = \left(\sum_{a \in \Sigma} |f(a)\rangle\langle a|\right)|b\rangle = \sum_{a \in \Sigma} |f(a)\rangle\langle a|b\rangle = |f(b)\rangle,$$

for each $b \in \Sigma$, which is precisely what we require of the matrix M.

As we will discuss in greater detail later in a later lesson, $\langle a|b\rangle$ may also be seen as an *inner product* between the vectors $|a\rangle$ and $|b\rangle$. Inner products are critically important in quantum information, but we'll delay a discussion of them until they are needed.

At this point the names "bra" and "ket" may be evident: putting a "bra" $\langle a|$ together with a "ket" $|b\rangle$ yields a "bracket" $\langle a|b\rangle$. This notation and terminology is due to Paul Dirac, and for this reason is known as the *Dirac notation*.

Probabilistic operations and stochastic matrices

In addition to deterministic operations, we have *probabilistic operations*.

For example, consider the following operation on a bit. 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, so that it becomes 0 with probability $1/2$ and 1 with probability $1/2$. This operation is represented by the matrix

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

One can check that this matrix does the correct thing by multiplying the two standard basis vectors by it.

For an arbitrary choice of a classical state set, we can describe the set of all probabilistic operations in mathematical terms as those that are represented by stochastic matrices, which are matrices satisfying these two properties:

1. All entries are nonnegative real numbers.
2. The entries in every column sum to 1.

Equivalently, stochastic matrices are matrices whose columns all form probability vectors.

We can think about probabilistic operations at an intuitive level as ones where randomness might somehow be used or introduced during the operation, just like in the example above. With respect to the stochastic matrix description of a probabilistic operation, each column can be viewed as a vector representation of the probabilistic state that is generated given the classical state input corresponds to that column.

We can also think about stochastic matrices as being exactly those matrices that always map probability vectors to probability vectors. That is, stochastic matrices always map probability vectors to probability vectors, and any matrix that always maps probability vectors to probability vectors must be a stochastic matrix.

Finally, a different way to think about probabilistic operations is that they are random choices *of* deterministic operations. For instance, we can think about the operation in the example above as applying either the identity function or the constant 0 function, each with probability $1/2$. This is consistent with the equation

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

Such an expression is always possible, for an arbitrary choice of a classical state set and any stochastic matrix with rows and columns identified with that classical state set.

Compositions of probabilistic operations

Suppose that X is a system having classical state set Σ, and $M_1, \ldots, M_n$ are stochastic matrices representing probabilistic operations on the system X.

If the first operation M_1 is applied to the probabilistic state represented by a probability vector u, the resulting probabilistic state is represented by the vector $M_1 u$. If we then apply the second probabilistic operation M_2 to this new probability vector, we obtain the probability vector

$$M_2(M_1 u) = (M_2 M_1)u.$$

The equality follows from the fact that matrix multiplication (which includes matrix-vector multiplication as a special case) is an associative operation. Thus, the probabilistic operation obtained by composing the first and second probabilistic operations, where we first apply M_1 and then apply M_2, is represented by the matrix $M_2 M_1$, which is necessarily stochastic.

More generally, composing the probabilistic operations represented by the matrices $M_1, \ldots, M_n$ in this order, meaning that M_1 is applied first, M_2 is applied second, and so on, with M_n applied last, is represented by the matrix product

$$M_n \cdots M_1.$$

Note that the ordering is important here: although matrix multiplication is associative, it is not a commutative operation. For example, if

$$M_1 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad M_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

then

$$M_2 M_1 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad M_1 M_2 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

That is, the order in which probabilistic operations are composed matters; changing the order in which operations are applied in a composition can change the resulting operation.

Previous page Next page