A hopefully intuitive explanation of the lin alg behind quantum gates

July 27, 2022

We mentioned that all quantum gates (and operations) can be represented unitary matrices. But what are unitary matrices and how do we get a better feel for what they do? Can we do this without needing any lin alg background? First, let's look at just the real number setting to visualize what is happening here.

Vectors $v, w \in \mathbb{R}^n$ have notions of norm, $||v|| = \sqrt{\sum_i v_i^2}$, and dot product, which we will write as $\langle v|w\rangle = \sum_i v_i w_i$. These two notions are related:

$$||v||^2 = \sum_i v_i^2 = \langle v|v\rangle.$$

Norm also establishes a way of calculating the Euclidean distance between points, v, w: $||v - w||$. Meanwhile the dot product gives you a sense of the angle between two vectors, since for example $\langle v|w\rangle = 0$ implies v and w are orthogonal (perpendicular).

We will say an $n \times n$ matrix with real entries, A, is an **orthogonal matrix** if it preserves norm: for any vector $v \in \mathbb{R}^n$, $||Av|| = ||v||$. This implies that A also preserves distance between any two points, v, w: since $||Av - Aw|| = ||A(v - w)|| = ||v - w||$.

This kind of rigid transformation that doesn't stretch or compress the distance between two points is called an isometry. Remembering that the origin (zero vector) is always mapped to itself by a linear transformation, what kind of rigid geometric transformations can you think of like this? If you thought of a rotation (the axis of rotation going through the origin) or a reflection (the plane of reflection going through the origin) you basically understand what all orthogonal matrices do.

It also turns out that A must also preserve inner product as well. To see this:

$$||v + w||^2 = \langle v + w|v + w\rangle = \langle v|v\rangle + \langle v|w\rangle + \langle w|v\rangle + \langle w|w\rangle = ||v||^2 + 2\langle v|w\rangle + ||w||^2,$$

where we use the fact that dot products are symmetric: $\langle v|w\rangle = \langle w|v\rangle$, and distributive: $\langle v + w|u\rangle = \langle v|u\rangle + \langle w|u\rangle$. Similarly,

$$||A(v + w)||^2 = ||Av + Aw||^2 = ||Av||^2 + 2\langle Av|Aw\rangle + ||Aw||^2.$$

Since A preserves norm, $||v + w||^2 = ||A(v + w)||^2$ and so the expressions are equal:

$$||v||^2 + 2\langle v|w\rangle + ||w||^2 = ||Av||^2 + 2\langle Av|Aw\rangle + ||Aw||^2.$$

Since $||v|| = ||Av||$ and $||w|| = ||Aw||$, we can eliminate these from both sides leaving behind:

$$2\langle v|w\rangle = 2\langle Av|Aw\rangle.$$

This is the math behind the intuition that if a transformation doesn't stretch distances between points, it wont stretch angles between vectors either.

Sometimes you may see two other equivalent conditions for being an orthogonal matrix that are easier for a computer to check:

- The columns of A are orthonormal (fancy word meaning each column of A has norm 1 and each of the columns is orthogonal to all the others)
- $A^tA = I$ (where A^t is the transpose of A)

Let's see where they come from. It is standard in math to let $e_i \in \mathbb{R}^n$ denote the vector that has 0s everywhere except for a 1 in the i-th position. So for example

$$e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

You can check for yourself that Ae_i is the i-th column of the matrix A. Since A preserves norm $||Ae_i|| = ||e_i|| = 1$ and so the columns of A have norm 1. Since A preserves inner products, $\langle Ae_i|Ae_j\rangle = \langle e_i|e_j\rangle = 0$ whenever $i \neq j$ since the vectors e_i, e_j are orthogonal. Thus, the columns of A are orthonormal because A maps the orthonormal set of vectors $e_1, \ldots, e_n$ to an orthonormal set of vectors $Ae_1, \ldots, Ae_n$.

If you carefully look at what the entry of A^tA in the i-th row and j-th column (try a couple examples with 2 by 2 matrices), you'll see that the entry equals

$$A_{1i}A_{1j} + A_{2i}A_{2j} + \cdots + A_{ni}A_{nj},$$

the dot product of the i-th and j-th columns of A, $\langle Ae_i|Ae_j\rangle$. Thus, the entries along the diagonal of A^tA ($i = j$) will have value $\langle Ae_i|Ae_i\rangle = ||Ae_i||^2 = 1$ and all the other entries ($i \neq j$) will be $\langle Ae_i|Ae_j\rangle = 0$. Thus, $A^tA = I$.

The complex world

Now we try to recreate all of these facts for vectors and matrices with complex number entries. First, we have to slightly modify the definition of norm:

$$||v|| = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2},$$

we have to introduce the magnitudes $|v_i|$ since squaring a complex number doesn't necessarily make it positive and we don't want the norm (a naturally positive distance measuring quantity) of $\begin{bmatrix} i \\ 0 \end{bmatrix}$ to be $\sqrt{i^2 + 0^2} = \sqrt{-1}$. Now we can talk about the distance between two vectors $v, w \in \mathbb{C}^n$ by simply defining it is $||v - w||$. Do we still have the equation $||v||^2 = \langle v|v\rangle$. Well as its currently defined this wont work since $v_i^2 = |v_i|^2$ for complex numbers. However, it is true that $\overline{v_i}v_i = |v_i|^2$ for complex numbers (check this for yourself with $v_i = a + b_i$).

This motivates us to define something happens to be called the **inner product** (instead of dot product):

$$\langle v|w\rangle = \overline{v_1}w_1 + \overline{v_2}w_2 + \cdots + \overline{v_n}w_n,$$

note that if all the entries of v, w are real numbers this is just the plain-vanilla dot product. Now we indeed have

$$||v||^2 = \sum_i |v_i|^2 = \sum_i \overline{v_i}v_i = \langle v|v\rangle.$$

The inner product will still have the distributive property: $\langle v + w|u\rangle = \langle v|u\rangle + \langle w|u\rangle$, but it is no longer symmetric (now we have to write $\langle v|w\rangle = \overline{\langle w|v\rangle}$). Congratulations, you now know what a **Hilbert space** is (at least what a finite dimensional one is). In the world of quantum computing, Hilbert space is just a fancy word for the set of vectors $\mathbb{C}^n$ where we care about being able to take inner products of vectors ,$\langle v|w\rangle$.

So what is a unitary matrix, B? It is just like an orthogonal matrices but it preserves the complex norm instead of the real one. Just as before this means that B preserves distance (the way defined it) between points $v, w \in \mathbb{C}^n$ and it preserves the inner products: $\langle Bv|Bw\rangle = \langle v|w\rangle$ (the proof is very similar but slightly trickier since we cant use $\langle v|w\rangle = \langle w|v\rangle$). Just like before this will mean that the columns of B are orthonormal (we will say $v, w \in \mathbb{C}^n$ are orthogonal whenever $\langle v|w\rangle = 0$).

The entries of A^tA where dot products of between different columns of A. Since in the complex world we only want to deal with inner products, we somehow need to introduce complex conjugates into this product of matrices. It ends up looking like this:

$$\overline{B}^t B = I,$$

where $\overline{B}$ applies the complex conjugate to each entry of B. This means that if all the entries of B are real, then $\overline{B} = B$ and so $I = \overline{B}^t B = B^t B$, so B is a plain-vanilla orthogonal matrix in this special case. This is why unitary matrices are called generalizations of orthogonal matrices (they generalize rigid transformations like rotations and reflections to the complex vector spaces).

Because unitary matrices are so important the set of all n by n unitary matrices has a name $U(n)$.

Relation to qubits

We mentioned earlier how quantum gates can be represented as linear transformations that map state vectors to state vectors. We saw that the state vector of one or more qubits $v \in \mathbb{C}^n$ must have norm $||v||^2 = \sum_i |v_i|^2 = 1$, since $|v_i|^2$ is the probability of measuring $|i\rangle$. If there are b qubits then there will be $n = 2^b$ possible outcomes of measurements. This why single or multi-qubit gates must be unitary matrices.

Technically, we only need the matrix to map all norm 1 vectors v to norm 1 vectors Bv since we only care about state vectors but once this is satisfied B will to preserve the norm of all vectors.

Just like there is an interpretation of what norm means for state vectors (the sum of probabilities of a measuring various outcomes). There is an interpretation of what the inner product means: Given two state vectors, v, w, the magnitude of their inner product, $|\langle v|w\rangle|$, is a measure of how similar the quantum states are.

For example, if $v = |0\rangle$ and $w = |1\rangle$ then $\langle v|w\rangle = 0$ so these states are very different. Similar the state $|+\rangle$ is orthogonal to $|-\rangle$ and $|i\rangle$ is orthogonal to $|-i\rangle$. On the other hand, for two global phase equivalent states $v = |0\rangle$ and $w = i\,|0\rangle$, we have $|\langle v|w\rangle| = |i| = 1$ and these two states are indeed as similar as they can be (they are indistinguishable).

This gives us an intuition for why a unitary matrix preserves inner products between state vectors: since quantum gates are reversible we can't make two states easier or harder to tell apart. When we say global phase equivalent states are indistinguishable we mean that that is still the case even if we apply any quantum operations to the two states.

This is also related by the distance between two states in the Bloch sphere. The bigger the distance (or angle) between two states on the Bloch sphere the more dissimilar they are and the smaller the value of $|\langle v|w\rangle|$.

Since unitary matrices preserve inner product, they preserve the distance between points on the Bloch sphere! It's kind of we tried to make qubits in $\mathbb{C}^2$ easier to visualize by putting them on the Bloch sphere in $\mathbb{R}^3$, and the unitary matrix (a generalization of rigid transformations) still ends up being a rigid transformation on this Bloch sphere. Since only rigid transformation that fix the origin are rotations and reflections, any quantum logic **must** rotate or reflect the Bloch sphere.

However, you can check for yourself that reflections just end up being nonsensical when you try to interpret them as a logic gate. For example take the reflection through the xz-plane which would have fix states $|0\rangle, |1\rangle, |+\rangle, |-\rangle$ (or possibly apply a global phase to them) but also swap states $|i\rangle$ and $|-i\rangle$. Can you show

that you cant write down a 2 by 2 matrix that does this?

Eigenstates

Let's take the single qubit gate again represented by a 2 by 2 unitary matrix $B \in U(2)$. We saw that can be viewed as a rotation of the Bloch sphere about some axis. The axis intersects the Bloch sphere at two antipodal states $|a\rangle, |b\rangle$. These are called the **eigenstates** of B (their state vectors are called the **eigenvectors** of B).

Since the gate rotates about this axis it must fix the axis in place. Does this mean $B|a\rangle = |a\rangle$ and $B|b\rangle = |b\rangle$? Well not exactly since we are on the Bloch sphere which doesn't perceive global space, so it is possible that $B|a\rangle = \lambda_1|a\rangle$ and $B|b\rangle = \lambda_2|b\rangle$ for some global phases $\lambda_1, \lambda_2 \in \mathbb{C}$. Not it is important to keep track of this since if $\lambda_1 \neq \lambda_2$ there is a relative phase effect (think of the Z gate which would have $\lambda_1 = 1$ and $\lambda_2 = -1$. These two phases, λ_1, λ_2 are called the **eigenvalues** of B.

So for example if $B = X$, then the axis of rotation is the axis containing states $|+\rangle$ and $|-\rangle$. Here we have $X|+\rangle = |+\rangle$ and $X|-\rangle = -|-\rangle$ so $\lambda_1 = 1$ and $\lambda_2 = -1$. Thus, X behaves like a phase shift gate (in particular like the phase-flip gate Z) but with respect to an unconventional basis, the Hadamard basis, $|+\rangle, |-\rangle$.

This process of finding eigenvalues and eigenvectors is called **spectral decomposition**. Essentially, it allows you to see a potentially complicated matrix in a much simpler way. While X is not that bad imagine a big matrix (a multi-qubit operation) with nonzero entries all over the place. In comparison, a multi-controlled Z gate is much simpler (its matrix is just a diagonal matrix with 1s and a -1 on the diagonal). By finding the right basis of states (like the eigenstates, $|+\rangle, |-\rangle$ for X) we can imagine the complicated action of the multi-qubit as just a phase flip (multi-controlled Z-like gate) gate when acting on this basis.

Spectral Theorem

It turns out there is a theorem that assures us that this basis will always exist. Essentially it says: all gates are phase flip gates (but they might be flipping the phase of some non-standard states).

Spectral Theorem (Complex Version): If $A \in M_n(\mathbb{C})$ satisfies $\overline{A^t}A = A\overline{A^t}$, then there is a orthonormal basis $|v_1\rangle, |v_2\rangle, \ldots, |v_n\rangle$ such that $A|v_i\rangle = \lambda_i|v_i\rangle$. Thus, we can think of A as having the same action as the phase flip gate

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

Notice that if A is unitary then it satisfies $\overline{A^t}A = I_n = A\overline{A^t}$. Another important class of matrices with this property are the **Hermitian matrices**, matrices with $\overline{B^t} = B$. This is the complex generalization of a symmetric matrix. In this case $\overline{B^t}B = B^2 = B\overline{B^t}$. The real version of the spectral theorem only applies to symmetric matrices (matrices with $A^t = A$).