Linear Algebra

```
from matplotlib import pyplot as plt
import numpy as np
from qiskit import *
from qiskit.visualization import plot_bloch_vector
```

Introduction

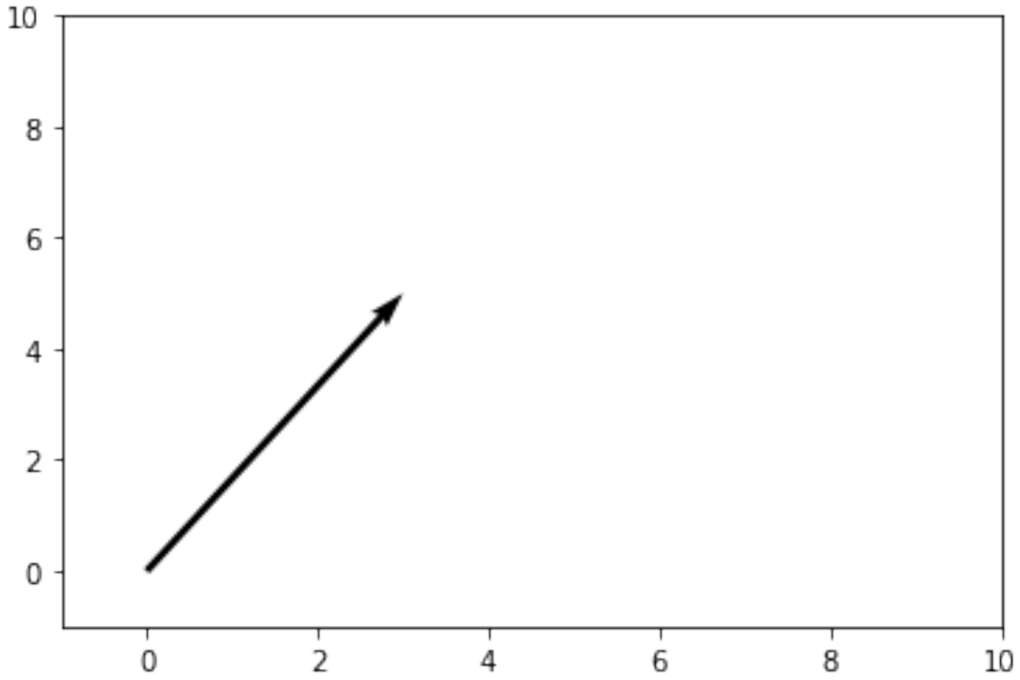
Linear algebra is the language of quantum computing. It is therefore crucial to develop a good understanding of the basic mathematical concepts that linear algebra is built upon, in order to arrive at many of the amazing and interesting constructions seen in quantum computation. The goal of this section is to create a foundation of introductory linear algebra knowledge, upon which the reader can build during their study of quantum computing.

Vectors and Vector Spaces

We will start our investigation into introductory linear algebra by first discussing one of the most important mathematical quantities in quantum computation: the vector.

Formally, a **vector** $|v\rangle$ $|v\rangle$ is defined as elements of a set known as a vector space. A more intuitive and geometric definition is that a vector "is a mathematical quantity with both direction and magnitude". For instance, consider a vector with x x and y y components of the form $\begin{pmatrix} 3 \\ 5 \end{pmatrix}$ $\begin{pmatrix} 3 \\ 5 \end{pmatrix}$. This vector can be visualized as an arrow pointing in the direction of 3 3 units down the x x axis and 5 5 units up the y y axis:

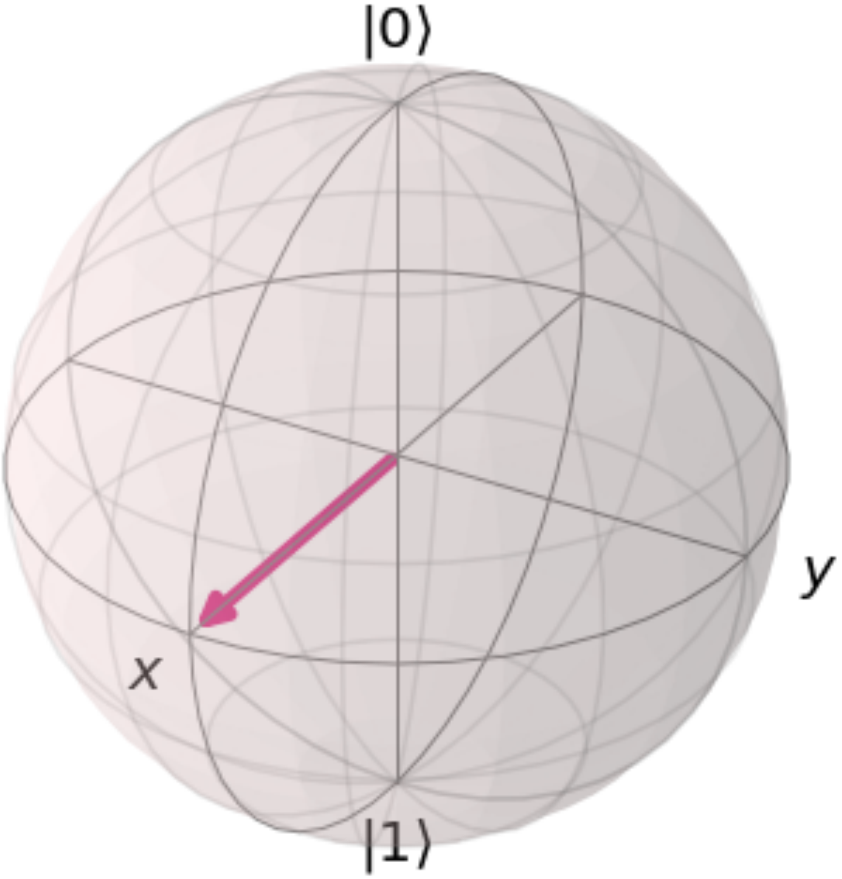
```
plt.figure()
ax = plt.gca()
ax.quiver([3], [5], angles='xy', scale_units='xy', scale=1)
ax.set_xlim([-1, 10])
ax.set_ylim([-1, 10])
plt.draw()
plt.show()
```



Note that "tail" of the vector doesn't have to be positioned at the origin; it only needs to point in the correct direction.

In quantum computing, we often deal with **state vectors**, which are simply vectors that point to a specific point in space that corresponds to a particular quantum state. This can be visualized using a Bloch sphere. For instance, a vector representing the state of a quantum system could look something like this arrow, enclosed inside the Bloch sphere, which is the so-called "state space" of all possible points to which our state vectors can "point":

```
plot_bloch_vector([1, 0, 0])
```



This particular state corresponds to an even superposition between $|0\rangle$ and $|1\rangle$ (the arrow is halfway between $|0\rangle$ at the top and $|1\rangle$ at the bottom of the sphere). Our vectors are allowed to rotate anywhere on the surface of the sphere, and each of these points represents a different quantum state.

Let's revisit our more formal definition of a vector, which is that a vector is an element of a vector space. We must now define a vector space. A **vector space** V over a **field** F is a set of objects (vectors), where two conditions hold. Firstly, **vector addition** of two vectors $|a\rangle, \ |b\rangle \ \in \ V$ will yield a third vector $|a\rangle \ + \ |b\rangle \ = \ |c\rangle$, also contained in V. The second condition is that **scalar multiplication** between some $|a\rangle \ \in \ V$ and some $n \ \in \ F$, denoted by $n|a\rangle$, is also contained within V.

We will now clarify this previous definition by working through a basic example. Let us demonstrate that the set $\mathbb{R}^2$ over the field $\mathbb{R}$ is a vector space. We assert that

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 & + & x_2 \\ y_1 & + & y_2 \end{pmatrix}$$

is contained within $\mathbb{R}^2$ $\mathbb{R}^2$. This is evidently the case, as the sum of two real numbers is a real number, making both components of the newly-formed vector real numbers; thus, the vector is contained in $\mathbb{R}^2$ $\mathbb{R}^2$ by definition. We also assert that:

$$n|v\rangle = \begin{pmatrix} nx \\ ny \end{pmatrix} \in V \quad \forall n \in \mathbb{R}$$

This is true as well, as the product of a real number and a real number is a real number, making the entire new vector real, and thus proving this statement.

Matrices and Matrix Operations

Let's turn our attention to another fundamental concept: a **matrix**. Matrices are mathematical objects that transform vectors to other vectors:

$$|v\rangle \rightarrow |v'\rangle = M|v\rangle$$

Generally, matrices are written as "arrays" of numbers, looking something like this:

$$M = \begin{pmatrix} 1 & -2 & 3 \\ 1 & 5i & 0 \\ 1 + i & 7 & -4 \end{pmatrix}$$

We can "apply" a matrix to a vector by performing matrix multiplication. In general, matrix multiplication between two matrices involves taking the first row of the first matrix, and multiplying each element by its "partner" in the first column of the second matrix (the first number of the row is multiplied by the first number of the column, second number of the row and second number of column, etc.). The sum of these new numbers becomes the first element of the first row of the new matrix. To fill in the rest of the first row, we repeat this process for the second, third, etc. columns of the second matrix. Then we take the second row of the first matrix, and repeat the process for each column of the second matrix, to produce the second row. We perform this process until we have used all rows of the first matrix. The resulting matrix is our new matrix. Here is an example:

$$\begin{pmatrix} 2 & 0 \\ 5 & -1 \end{pmatrix}\begin{pmatrix} -3 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} (2)(-3) + (0)(2) & (2)(1) + (0)(1) \\ (5)(-3) + (-1)(2) & (5)(1) + (-1)(1) \end{pmatrix}$$
$$= \begin{pmatrix} -6 & 2 \\ -17 & 4 \end{pmatrix}$$

To perform a quantum computation, we have some quantum state vector we manipulate by applying a matrix to that vector. A vector is simply a matrix with one column. To apply a matrix to a vector, therefore, we follow the same matrix multiplication procedure described above. We manipulate qubits in our quantum computer by applying sequences of **quantum gates**. Each quantum gate can be expressed as a matrix that can be applied to state vectors, thus changing the state. For instance, a commonly seen quantum gate is the Pauli-X gate, which is represented by the following matrix:

$$\sigma_x \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

This gate acts similarly to the classical NOT logic gate. It maps the computational basis state $|0\rangle$ $|0\rangle$ to $|1\rangle$ $|1\rangle$ and $|1\rangle$ $|1\rangle$ to $|0\rangle$ $|0\rangle$ (it "flips" the state). We write the two basis states as column vectors:

$$|0\rangle \;=\; \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad |1\rangle \;=\; \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

When we apply this matrix to each of the vectors:

$$\sigma_x|0\rangle \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \;=\; \begin{pmatrix} (0)(1) \;+\; (1)(0) \\ (1)(1) \;+\; (0)(0) \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ 1 \end{pmatrix} \;=\; |1\rangle$$

$$\sigma_x|1\rangle \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \;=\; \begin{pmatrix} (0)(0) \;+\; (1)(1) \\ (1)(0) \;+\; (0)(1) \end{pmatrix} \;=\; \begin{pmatrix} 1 \\ 0 \end{pmatrix} \;=\; |0\rangle$$

The matrix acts on the state vectors as expected.

Within quantum computation, we often encounter two important types of matrices: **Hermitian** and **Unitary** matrices. The former is more important in the study of quantum mechanics, but is still necessary to discuss in a study of quantum computation. The latter is of unparalleled importance in both quantum mechanics and quantum computation. If you take away only one concept from this section on linear algebra, it should be the concept of a unitary matrix.

A Hermitian matrix is simply a matrix that is equal to its **conjugate transpose** (denoted with a $\dagger$ $\dagger$ symbol). This means that flipping the sign of a Hermitian matrix's imaginary components, then reflecting its entries along its main diagonal (from the top left to bottom right corners), produces an equal matrix. For instance, the Pauli-Y matrix, commonly used in quantum computation, is Hermitian:

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \Rightarrow \sigma_y^{\dagger} = \begin{pmatrix} 0 & -(i) \\ -(-i) & 0 \end{pmatrix} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \sigma_y$$

Notice how we switched the places of the ii and the $-i-i$ (as we reflect across the main diagonal, the zeroes remain unchanged), and then flipped the sign.

A unitary matrix is very similar. Specifically, it is a matrix such that the **inverse matrix** is equal to the conjugate transpose of the original matrix.

The inverse of some matrix AA, denoted as $A^{-1}A^{-1}$, is a matrix such that:

$$A^{-1}A = AA^{-1} = \mathbb{I}$$

where $\mathbb{I}$ is the identity matrix. The identity matrix has 1s along the main diagonal (top left to bottom right), and 0s in all other places. It is called the identity matrix because it acts trivially on any other matrix (it has no effect). You can prove this on your own by multiplying an identity matrix by any other matrix.

When matrices get larger than 2×2, calculating the inverse becomes sufficiently complicated that it is usually left to computers to calculate. For a 2×2 matrix, the inverse is defined as:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

where $\det A$ is the **determinant** of the matrix. In the 2×2 case, $\det A = ad - bc$.

Calculating inverse matrices is rarely important in quantum computing. Since most of the matrices we encounter are unitary, we can assume that the inverse is simply given by taking the conjugate transpose.

Let's look at a basic example. The Pauli-Y matrix, in addition to being Hermitian, is also unitary (it is equal to its conjugate transpose, which is also equal to its inverse; therefore, the Pauli-Y matrix is its own inverse!). We can verify that this matrix is in fact unitary:

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_y^{\dagger} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \Rightarrow \sigma_y^{\dagger}\sigma_y$$
$$= \begin{pmatrix} (0)(0) + (-i)(i) & (0)(-i) + (-i)(0) \\ (i)(0) + (0)(i) & (i)(-i) + (0)(0) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbb{I}$$

The reason unitary matrices are important will become more apparent in the section on Hilbert spaces, and more so in the quantum mechanics subtopic of this textbook. The basic idea is that evolution of a quantum state by application of a unitary matrix "preserves" the quantum state.

Spanning Sets, Linear Dependence, and Bases

We are now in a position to discuss the construction of vector spaces. Consider some vector space $V V$. We say that some set of vectors $S S$ spans a subspace $V_S \subset V V_S \subset V$ (subset closed under vector space operations) of the vector space, if we can write any vector in the subspace as a **linear combination** of vectors contained within the spanning set.

A linear combination of some collection vectors $|v_1\rangle, \ ..., \ |v_n\rangle |v_1\rangle, \ ..., \ |v_n\rangle$ in some vector space over a field $F F$ is defined as an arbitrary sum of these vectors (which of course will be another vector that we will call $|v\rangle |v\rangle$):

$$|v\rangle \ = \ f_1|v_1\rangle \ + \ f_2|v_2\rangle \ + \ ... \ + \ f_n|v_n\rangle \ = \ \sum_i \ f_i|v_i\rangle$$

where each $f_i f_i$ is some element of $F F$. If we have a set of vectors that spans a space, we are saying that **any other vector** in the vector space can be written as a linear combination of these vectors.

A set of vectors $|v_1\rangle, \ ..., \ |v_n\rangle |v_1\rangle, \ ..., \ |v_n\rangle$ is said to be **linearly dependent** if there exist corresponding coefficients for each vector, $b_i \ \in \ F b_i \ \in \ F$, such that:

$$b_1|v_1\rangle \ + \ b_2|v_2\rangle \ + \ ... \ + \ b_n|v_n\rangle \ = \ \sum_i \ b_i|v_i\rangle \ = \ 0,$$

where at least one of the b_i b_i coefficients is non-zero. This is equivalent to the more intuitive statement that "the set of vectors can be expressed as linear combinations of each other". For example, let us have the set $\{|v_1\rangle, \ \dots, \ |v_n\rangle\}$ $\{|v_1\rangle, \ \dots, \ |v_n\rangle\}$ along with the corresponding coefficients $\{|b_1\rangle, \ \dots, \ |b_n\rangle\}$ $\{|b_1\rangle, \ \dots, \ |b_n\rangle\}$, such that the linear combination is equal to 0 0 . Since there is at least one vector with a non-zero coefficient, we choose a term in the linear combination $b_a|v_a\rangle$ $b_a|v_a\rangle$:

$$\begin{gathered} \sum_i \ b_i|v_i\rangle \ = \ b_a|v_a\rangle \ + \ \sum_{i, \, i \neq a} \ b_i|v_i\rangle \ = \ 0 \ \Rightarrow \ |v_a\rangle \ = \\ - \sum_{i, \, i \neq a} \ \frac{b_i}{b_a}|v_i\rangle \ = \ \sum_{i, \, i \neq a} \ c_i|v_i\rangle \end{gathered}$$

In the case that b_a b_a is the only non-zero coefficient, it is necessarily true that $|v_a\rangle$ $|v_a\rangle$ is the null vector, automatically making the set linearly dependent. If this is not the case, $|v_a\rangle$ $|v_a\rangle$ has been written as a linear combination of non-zero vectors, as was shown above. To prove the converse, we assume that there exists some vector $|v_a\rangle$ $|v_a\rangle$ in the subspace $|v_1\rangle, \dots, \ |v_n\rangle$ $|v_1\rangle, \dots, \ |v_n\rangle$ that can be written as a linear combination of other vectors in the subspace. This means that:

$$|v_a\rangle \ = \ \sum_s b_s|v_s\rangle$$

where s s is an index that runs over a subset of the subspace. It follows that:

$$|v_a\rangle \ - \ \sum_s b_s|v_s\rangle \ = \ |v_a\rangle \ - \ (|v_{s_1}\rangle \ + \ \ldots \ + \ |v_{s_r}\rangle) \ = \ 0$$

For all vectors in the subspace that are not included in the subset indexed by $s s$, we set their coefficients, indexed by $q q$, equal to $0 0$. Thus,

$$|v_a\rangle \ - \ (|v_{s_1}\rangle \ + \ \ldots \ + \ |v_{s_r}\rangle) \ + \ (0)(|v_{q_1}\rangle \ + \ \ldots \ + \ |v_{q_t}\rangle) \ = \ 0$$

which is a linear combination of all elements in the subspace $|v_1\rangle, \ \ldots, \ |v_n\rangle |v_1\rangle, \ \ldots, |v_n\rangle$. This is equal to $0 0$, thus completing the proof that the two definitions of linear dependence imply each other.

Let's now consider a basic example. Consider the set of two vectors in $\mathbb{R}^2 \mathbb{R}^2$, consisting of $|a\rangle \ = \ \begin{pmatrix} 1 \\ 0 \end{pmatrix} |a\rangle \ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|b\rangle \ = \ \begin{pmatrix} 2 \\ 0 \end{pmatrix} |b\rangle \ = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$. If we choose the field over our vector space to be $\mathbb{R} \mathbb{R}$, then we can create a linear combination of these vectors that equates to $0 0$. For example:

$$2|a\rangle \ - \ |b\rangle \ = \ 0$$

A set of vectors is said to be **linearly independent** if there is no vector in the set that can be expressed as a linear combination of all the others.

The notion of a **basis** is simply a **linearly independent spanning set**. In this sense, the basis of a vector space is the minimal possible set that spans the entire space. We call the size of the basis set the **dimension** of the vector space.

Bases and spanning sets are important because they allow us to "shrink down" vector spaces and express them in terms of only a few vectors. We can come to certain conclusions about our basis set that we can generalize to the entire vector space, simply because we know every vector in the space is just a linear combination of the basis vectors.

In quantum computation, one of the bases that we often encounter is $|0\rangle, \ |1\rangle |0\rangle, \ |1\rangle$. We can write any other qubit state as a linear combination of these basis vectors. For instance, the linear combination

$$\frac{|0\rangle \ + \ |1\rangle}{\sqrt{2}}$$

represents a superposition between the $|0\rangle |0\rangle$ and $|1\rangle |1\rangle$ basis state, with equal probability of measuring the state to be in either one of the basis vector states (this is intuitive, as the "weight" or the "amount of each basis vector" in the linear combination is equal, both being scaled by $1/\sqrt{2} 1/\sqrt{2}$).

Hilbert Spaces, Orthonormality, and the Inner Product

Hilbert Spaces are one of the most important mathematical constructs in quantum mechanics and quantum computation. A Hilbert space can be thought of as the state space in which all quantum state vectors "live". The main difference between a Hilbert space and any random vector space is that a Hilbert space is equipped with an **inner product**, which is an operation that can be performed between two vectors, returning a scalar.

In the context of quantum mechanics and quantum computation, the inner product between two state vectors returns a scalar quantity representing the amount to which the first vector lies along the second vector. From this, the probabilities of measurement in different quantum states (among other things) can be calculated (this will be discussed more in the quantum mechanics subtopic).

For two vectors $|a\rangle$ $|a\rangle$ and $|b\rangle$ $|b\rangle$ in a Hilbert space, we denote the inner product as $\langle a|b\rangle$ $\langle a|b\rangle$, where $\langle a|$ $\langle a|$ is equal to the conjugate transpose of $|a\rangle$ $|a\rangle$, denoted $|a\rangle^{\dagger}$ $|a\rangle^{\dagger}$. Thus, the inner product between two vectors of the Hilbert space looks something like:

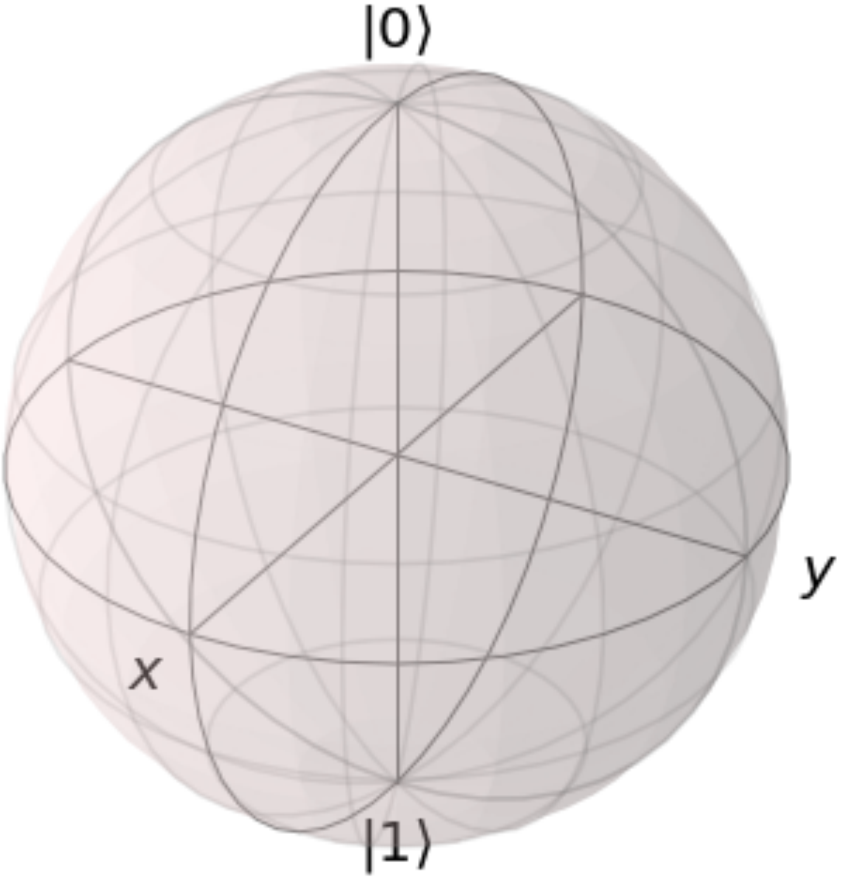
$$\langle a|b\rangle \ = \ \begin{pmatrix} a_1^* & a_2^* & ... & a_n^* \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{pmatrix} \ = \ a_1^* b_1 \ + \ a_2^* b_2 \ + \ ... \ + \ a_n^* b_n$$

where $*$ $*$ denotes the complex conjugate of the vector.

One of the most important conditions for a Hilbert space representing a quantum system is that the inner product of a vector with itself is equal to one: $\langle\psi|\psi\rangle \ = \ 1$ $\langle\psi|\psi\rangle \ = \ 1$. This is the so-called normalization condition, which states that the length of the vector squared (each component of the vector is squared and summed together, by defintion of the inner product) must be equal to one. The physical significance of this is that the length of a vector in a particular direction is representative of the "probability amplitude" of the quantum system with regards to measurement in that particular state. Obviously, the probability of the quantum system being measured in the state that it is in must be 1 1 (after all, the sum of the probabilities of finding the quantum system in any particular state must equal 1 1).

Let's consider the Bloch sphere:

```
plot_bloch_vector([0, 0, 0])
```



The surface of this sphere, along with the inner product between qubit state vectors, is a valid Hilbert space. In addition, the normalization condition holds true, as the radius of the Bloch sphere is 11, and thus the length squared of each vector must also equal 11.

A final note regarding Hilbert spaces and the inner product is their relationship to **unitary matrices**. Unitary matrices are important in quantum computation because they **preserve the inner product**, meaning that no matter how you transform a vector under a sequence of unitary matrices, the normalization condition still holds true. This can be demonstrated in the following short proof:

$$\langle\psi|\psi\rangle = 1 \Rightarrow |\psi\rangle \rightarrow U|\psi\rangle = |\psi'\rangle \Rightarrow \langle\psi'|\psi'\rangle = (U|\psi\rangle)^{\dagger}U|\psi\rangle$$
$$= \langle\psi|U^{\dagger}U|\psi\rangle = \langle\psi|\psi\rangle = 1$$

This means that unitary evolution sends quantum states to other valid quantum states. For a single-qubit Hilbert space, represented by the Bloch sphere, unitary transformations correspond to rotations of state vectors to different points on the sphere, not changing the length of the state vector in any way.

Eigenvectors and Eigenvalues

Consider the relationship of the form:

$$A|v\rangle = \lambda|v\rangle,$$

where AA is a matrix, and $\lambda\lambda$ is some number. If we are given some matrix AA, and need to find the vectors $|v\rangle|v\rangle$ and numbers $\lambda\lambda$ that satisfy this relationship, we call these vectors **eigenvectors**, and their corresponding number multipliers **eigenvalues**. Eigenvectors and eigenvalues have very important physical significance in the context of quantum mechanics, and therefore quantum computation. Given some AA, we exploit an interesting trick in order to find the set of eigenvectors and corresponding eigenvalues. Let us rearrange our equation as:

$$A|v\rangle - \lambda|v\rangle = 0 \Rightarrow (A - \lambda\mathbb{I})|v\rangle = 0$$

If we multiply both sides of this equation by the inverse matrix $(A - \lambda\mathbb{I})^{-1}(A - \lambda\mathbb{I})^{-1}$, we get $|v\rangle = 0|v\rangle = 0$. This is an extraneous solution (we don't allow eigenvectors to be the null vector, or else any eigenvalue/matrix combination would satisfy the eigenvector-eigenvalue relationship). Thus, in order to find the allowed eigenvectors and eigenvalues, we have to assume that the matrix $(A - \lambda\mathbb{I})(A - \lambda\mathbb{I})$ is **non-invertible**. Recall from earlier that the inverse of a matrix is of the form:

$$M^{-1} = \frac{1}{\det(M)} F(M),$$

where $F(M) F(M)$ is some new matrix (the particulars of which do not matter in this context) that depends on $M M$. The part of this equation in which we are interested is the inverse of the determinant. If the determinant of the matrix $M M$ is $0 0$, it follows that the inverse is undefined, and thus so is the inverse, making the matrix $M M$ non-invertible. We therefore require that:

$$\det(A - \lambda\mathbb{I}) = 0$$

From this, we can determine $\lambda \lambda$, then we plug each value of $\lambda \lambda$ back into the original equation to get the eigenvalues. Let's do an example, and find the eigenvectors/eigenvalues of the Pauli-Z matrix, $\sigma_z \sigma_z$. We start with:

$$\det(\sigma_z - \lambda\mathbb{I}) = \det\begin{pmatrix} 1 - \lambda & 0 \\ 0 & -1 - \lambda \end{pmatrix} = (-1 - \lambda)(1 - \lambda) = 1 \\ - \lambda^2 = 0 \Rightarrow \lambda = \pm 1$$

The equation, in terms of $\lambda \lambda$, that results when solving the determinant is called the **characteristic polynomial**. We can then plug each of these values back into the original equation. We'll start with $\lambda = 1 \lambda = 1$:

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} |v\rangle = |v\rangle \Rightarrow \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} \Rightarrow \begin{pmatrix} a \\ -b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

a can be any number, and b is 0; thus, the vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ forms a basis for all vectors that satisfy our relationship, and is therefore the eigenvector that corresponds to the eigenvalue of 1. We do the same thing for $\lambda = -1$:

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} |v\rangle = -|v\rangle \Rightarrow \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -a \\ -b \end{pmatrix} \Rightarrow \begin{pmatrix} a \\ -b \end{pmatrix} = \begin{pmatrix} -a \\ -b \end{pmatrix}$$

This time, b can be any number, and a is 0; thus, our basis vector (and our eigenvector corresponding to -1) is $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Notice how the eigenvectors of the Pauli-Z matrix are the quantum computational basis states $|0\rangle$ and $|1\rangle$. This is no coincidence. For instance, when we measure a qubit in the Z-basis, we are referring to a measurement that collapses the qubit's state into one of the eigenvectors of the Z matrix, either $|0\rangle$ or $|1\rangle$.

Matrix Exponentials

The notion of a matrix exponential is a very specific yet extremely important concept. We often see unitary transformations in the form:

$$U = e^{i\gamma H},$$

where H H is some Hermitian matrix and γ γ is some real number. It is fairly simple to prove that all matrices of this form are unitary. Taking the conjugate transpose of U U, we get:

$$U^\dagger = \left(e^{i\gamma H}\right)^\dagger = e^{-i\gamma H^\dagger}$$

But since H H is Hermitian, we know that $H^\dagger = H$ $H^\dagger = H$, thus:

$$e^{-i\gamma H^\dagger} = e^{-i\gamma H} \Rightarrow U^\dagger U = e^{-i\gamma H} e^{i\gamma H} = \mathbb{1}$$

You may wonder why a matrix inside of an exponential can still be considered a matrix. The answer becomes clearer when we expand our exponential function as a Taylor series. Recall from calculus that a Taylor series is essentially a way to write any function as an infinite-degree polynomial, and the main idea is to choose the terms of the polynomial and center it at some point x_0 x_0 lying on the function we are trying to transform into the polynomial, such that the zeroth, first, second, third, etc. derivative is the same for both the original function and the polynomial. Thus, we write our Taylor series in the form:

$$g(x) \;=\; \sum_{n=0}^{\infty} f^{(n)}(x_0)\,\frac{(x \;-\; x_0)^n}{n!},$$

where $g(x)g(x)$ is the polynomial, $f(x)f(x)$ is the original function, $f^{(n)}\,f^{(n)}$ is the nn-th derivative of $f\,f$, and $x_0\,x_0$ is the point at which we center the function. Since we are not approximating, $x_0\,x_0$ doesn't matter, so for simplicity, we choose $x_0 \;=\; 0x_0 \;=\; 0$, and the Taylor series becomes a Maclaurin series:

$$g(x) \;=\; \sum_{n=0}^{\infty} f^{(n)}(0)\,\frac{x^n}{n!}$$

If we choose $f(x) \;=\; e^x\,f(x) \;=\; e^x$, we can create an equivalent polynomial using the Maclaurin series. Since the derivative of $e^x\,e^x$ is simply $e^x\,e^x$, and evidently, $e^0 \;=\; 1e^0 \;=\; 1$, we get:

$$g(x) \;=\; \sum_{n=0}^{\infty} \frac{x^n}{n!} \;=\; e^x$$

Thus, for some matrix, $i\gamma H i\gamma H$, we get:

$$e^{i\gamma H} \; = \; \sum_{n=0}^{\infty} \frac{(i\gamma H)^n}{n!}$$

Therefore, the exponential of a matrix is a matrix. It is an infinite sum of powers of matrices, which admittedly looks overly complex...but the point here is that the matrix exponential is indeed a matrix.

We are now in a position to demonstrate a very important fact: if we have some matrix $B B$ such that $B^2 \; = \; \mathbb{I} B^2 \; = \; \mathbb{I}$ (this is called an **involutory matrix**), then:

$$e^{i\gamma B} \; = \; \cos(\gamma)\mathbb{I} \; + \; i \sin(\gamma) B$$

We start with the Maclaurin series:

$$e^{i\gamma B} \; = \; \sum_{n=0}^{\infty} \frac{(i\gamma B)^n}{n!}$$

Notice that we can split the summation into an imaginary part and a real part, based on whether $n n$ is even or odd in each term of the sum:

$$\sum_{n=0}^{\infty} \frac{(i\gamma B)^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n \gamma^{2n} B^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} \frac{(-1)^n \gamma^{2n+1} B^{2n+1}}{(2n+1)!}$$

Now, let us find the Maclaurin series for both $\sin x \sin x$ and $\cos x \cos x$. We'll start with $f(x) = \sin x f(x) = \sin x$:

$$\sin x = \sum_{n=0}^{\infty} f^n(0) \frac{x^n}{n!}$$

The derivative of $\sin x \sin x$ is **cyclical** in a sense (each arrow represents taking the derivative of the previous function):

$$\sin x \rightarrow \cos x \rightarrow -\sin x \rightarrow -\cos x \rightarrow \sin x$$

Since $\sin(0) = 0 \sin(0) = 0$ and $\cos(0) = 1 \cos(0) = 1$, all terms with even $n n$ become $0 0$, and we get:

$$\sum_{n=0}^{\infty} f^n(0) \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n + 1)!}$$

This looks similar to the odd term of our original equation. In fact, if we let $x = \gamma \boldsymbol{B}$$x = \gamma B$, they are exactly the same. We follow a process that is almost identical to show that the even terms are the same as the Maclaurin series for $f(x) = \cos x$$f(x) = \cos x$:

$$\cos x = \sum_{n=0}^{\infty} f^{n}(0)\frac{x^n}{n!}$$

$$\Rightarrow \cos x \rightarrow -\sin x \rightarrow -\cos x \rightarrow \sin x \rightarrow \cos x$$

$$\Rightarrow \sum_{n=0}^{\infty} f^{n}(0)\frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

Let us go back to the original equation. Recall that $\boldsymbol{B}^2 = \mathbb{I}$$B^2 = \mathbb{I}$. For any $n$$n$, we have:

$$\boldsymbol{B}^{2n} = \left(\boldsymbol{B}^2\right)^n = \mathbb{I}^n = \mathbb{I}$$

$$B^{2n+1} \;=\; B\left(B^2\right)^n \;=\; B\,\mathbb{I}^n \;=\; B\,\mathbb{I} \;=\; B$$

Substituting in this new information, we get:

$$\sum_{n=0}^{\infty} \frac{(-1)^n \gamma^{2n} B^{2n}}{(2n)!} + \; i \sum_{n=0}^{\infty} \frac{(-1)^n \gamma^{2n+1} B^{2n+1}}{(2n+1)!} \;=\; \mathbb{I} \sum_{n=0}^{\infty} \frac{(-1)^n \gamma^{2n}}{(2n)!} \;+\; iB \sum_{n=0}^{\infty} \frac{(-1)^n \gamma^{2n+1}}{(2n+1)!} \;=\; \cos(\gamma)\mathbb{I} \;+\; i\sin(\gamma)B$$

This fact is extremely useful in quantum computation. Consider the Pauli matrices:

$$\sigma_x \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\sigma_y \;=\; \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

$$\sigma_z \;=\; \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

These matrices are among the fundamental "quantum gates" used to manipulate qubits. These operations are not only unitary, they are also **Hermitian** and **Involutory**. This means that a matrix of the form $e^{i\gamma\sigma_k} \;\; k \;\in\; \{x,\; y,\; z\} e^{i\gamma\sigma_k} \;\; k \;\in\; \{x,\; y,\; z\}$ is not only a valid unitary matrix that can act upon a quantum state vector (a qubit), but it can be expressed using the sine-cosine relationship that we just proved. This is very powerful, and is seen throughout quantum computational theory, as gates of this type are used all the time.

One last important fact about matrix exponentials: if we have some matrix $M M$, with eigenvectors $|v\rangle |v\rangle$ and corresponding eigenvalues $v v$, then:

$$e^M |v\rangle \;=\; e^v |v\rangle$$

This one is much more straightforward to prove:

$$e^M |v\rangle \;=\; \sum_{n=0}^{\infty} \frac{B^n |v\rangle}{n!} \;=\; \sum_{n=0}^{\infty} \frac{v^n |v\rangle}{n!} \;=\; e^v |v\rangle$$

This fact is also very useful. When creating quantum circuits that simulate a certain Hamiltonian (especially for variational circuits), we frequently use gates of the form $e^{i\gamma\sigma_z} e^{i\gamma\sigma_z}$. Since $|0\rangle |0\rangle$ and $|1\rangle |1\rangle$ are eigenvalues of $\sigma_z \sigma_z$, we can easily determine mathematically that $e^{i\gamma\sigma_z} e^{i\gamma\sigma_z}$ will add a phase of $e^{i\gamma} e^{i\gamma}$ to $|0\rangle |0\rangle$, and will add a phase of $e^{-i\gamma} e^{-i\gamma}$ to $|1\rangle |1\rangle$. We can then construct this gate in terms of $CNOT CNOT$ and phase/rotation gates fairly easily, as we know the mathematical outcome of the gate on each of the computational basis states.

This fact doesn't only apply to exponentials of the $\sigma_z \sigma_z$ gate. For example, we can determine the outcome of a gate of the form $e^{i\gamma\sigma_x}$ $e^{i\gamma\sigma_x}$ on the eigenvectors of $\sigma_x \sigma_x$, $(|0\rangle + |1\rangle)/\sqrt{2}$ $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle - |1\rangle)/\sqrt{2}(|0\rangle - |1\rangle)/\sqrt{2}$. The same applies to exponentials of the σ_y σ_y matrix.

References

[1] Cayley, Arthur. “A Memoir on the Theory of Matrices.” Philosophical Transactions of the Royal Society of London, vol. 148, 1858, pp. 17–37. JSTOR.

[2] A New Branch of Mathematics: The Ausdehnungslehre of 1844 and Other Works: Hermann Grassmann, Lloyd C. Kannenberg: 9780812692761