QCI402 - Mathmatical Foundations for Quantum Computing - Northern University

Dr. Miao Yu - Mingjia Guan

Fall Semester 2025

The underpinnings of all scientific advancements is the ability to express natural phenomena with the art of Mathmatics; this is no different for the subject of Quantum Computing. While the boundaries of quantum computing have been pushed beyond limits in theoretical terms on university blackboards, it has become of great interest to realize the theoretical computational power with the advances of hardware and technology.

However, these notes mainly concerns itself with the mathmatical underpinnings of quantum computing that the course surrounds itself with. Mathmatical Foundations for Quantum Computingtakes a scaffolding approach designed to efficiently convey the required theoretical understanding of mathmatics in order to able to learn quantum computing. As of writing, we are basing the notes on verison one of the textbook published in March 2025. In this text, we will primarily be using dirac notation for the expression of vectors, operators, and their interactions.

Contents

1 Summation and Product Notations

This section primarily focuses on the common notations applied across mathmatics to denote and shorten addition and product notation.

1.1 Summation over a single Variable

The sigma notation is defined as follows

$$\sum_{i=1}^{n} f(i)$$

where we use sigma $\sum$ to represent the sum of a series. For example, the sum of all numbers in a series beginning with m and ending at index n is written as:

$$\sum_{i=m}^{n} a_i = a_m + a_{m+1} + a_{m+2} + \cdots + a_{n-1} + a_n$$

Sums can also be infinite, commonly seen when Sigma looks as follows: $\sum_{i=m}^{\infty}$. Infinite sums are either convergent or divergent. A few of the most common converging infinite sums are as follows:

$$\sum_{i=0}^{\infty} \frac{1}{2^i} = 1 + \frac{1}{2} + \frac{1}{4} + \cdots = 2$$

$$\sum_{i=0}^{\infty} \frac{1}{i^2} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots = \frac{\pi^2}{6}$$

The first example is an infinite geometric series, and the sum of the first n terms is given by:

$$S_n = \sum_{i=0}^{n} \frac{1}{2^i} = \frac{1 - \frac{1}{2^n}}{1 - \frac{1}{2}}$$

As $n \to \infty, \frac{1}{2^n} \to 0$. Consequently, $S_n \to \frac{1}{1-\frac{1}{2}} = 2$. A rigorous proof of the second example requires extensive calculus and is not immediately obvious. While any mathmatical symbol can be used for the index of a summation, it is more practical to use something other than i as in the context of complex numbers, i commonly denotes the complex number$\sqrt{-1}$. moreover, sume can also be specified using descriptions. For example,

$$\sum_{p \in P} f(p) \qquad P \in \mathbb{N}'$$

where $\mathbb{N}'$ is the set of all prime numbers. Summations can also contain parameters other than the index, which results in functions of those parameters. For example the discrete Fourier transform (DFT) is given by

$$\tilde{x}_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn}, \quad k = 0, 1, \cdots N-1$$

where x_n represents the N values index by n and $\tilde{x}_k$ are the Fourier coefficients. Here, i is the imaginary numebr and N is a positive integer representing the dimension fo the DFT, of which we will cover in greater depth in Chapter 3. The following are some useful summation forumae commonly encountered in quantum computing:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{i=1}^{n} i^3 = \left(\frac{n(n+1)}{2}\right)^2$$

$$\sum_{i=0}^{n} (a_0 + id) = (n+1)\left(a_0 + \frac{nd}{2}\right) \quad \text{(arithmetic series)}$$

$$\sum_{i=0}^{n} a^i = \frac{1-a^{n+1}}{1-a} \quad \text{(geometric series)}$$

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i \quad \text{(binomial theorem)}$$

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots \quad (|x| < 1)$$

$$\frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} n x^{n-1} = 1 + 2x + 3x^2 + 4x^3 + \cdots \quad (|x| < 1)$$

$$\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \quad (|x| < 1)$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$$

Below are also a list of the common summations rules and manupulations:

$$\sum_{i=m}^{n} a_i = \sum_{j=m}^{n} a_j \quad \text{(change of index variable)}$$

$$\sum_{i=s}^{t} f(i) = \sum_{n=s}^{t} f(n) \quad \text{(change of index variable)}$$

$$\sum_{n=s}^{t} f(n) = \sum_{n=s}^{j} f(n) + \sum_{n=j+1}^{t} f(n) \quad \text{(splitting a sum)}$$

$$\sum_{n=s}^{t} f(n) = \sum_{n=0}^{t-s} f(t-n) \quad \text{(reverse order)}$$

$$\sum_{n=s}^{t} f(n) = \sum_{n=s+p}^{t+p} f(n-p) \quad \text{(index shift)}$$

$$\sum_{n=s}^{t} a \cdot f(n) = a \cdot \sum_{n=s}^{t} f(n) \quad \text{(distributivity)}$$

$$\sum_{n=s}^{t} f(n) \pm \sum_{n=s}^{t} g(n) = \sum_{n=s}^{t} (f(n) \pm g(n)) \quad \text{(commutativity)}$$

1.2 Products and other Notations

Similar to the $\sum$ notation for addition, the $\prod$ (Pi) symbol is also more commonly used to dentoe the product of a series of terms. In this

$$\prod_{i=m}^{n} a_i = a_m \cdot a_{m+1} \cdot a_{m+2} \cdot \cdots \cdot a_{n-1} \cdot a_n$$

for example, the factorial of n is expressed as

$$\prod_{i=0}^{n} i = n!$$

and the relationship between $\sum$ and $\prod$, which are

$$b^{\sum_{n=s}^{t} f(n)} = \prod_{n=s}^{t} b^{f(n)}$$

$$\sum_{n=s}^{t} \log_b f(n) = log_b \prod_{n=s}^{t} f(n)$$

It is worth noting that in quantum computing and linear algebra, there are a few special notations such as the modulo-2 sum (bitwise XOR), or in other contexts the direct sum of linear spaces, represented by $\oplus$, and the tensor product represented by $\otimes$.

1.3 Summation over Multiple Variables

The double summation over a rectangular array is given by

$$\begin{aligned}
\sum_{i=1,j=1}^{n_1,n_2} a_{i,j} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} a_{i,j} = \sum_{j=1}^{n_2}\sum_{i=1}^{n_1} a_{i,j} \\
&= a_{1,1} + a_{1,2} + a_{1,3} + a_{1,4} + \cdots + a_{1,n_2} \\
&+ a_{2,1} + a_{2,2} + a_{2,3} + a_{2,4} + \cdots + a_{2,n_2} \\
&+ a_{3,1} + a_{3,2} + a_{3,3} + a_{3,4} + \cdots + a_{3,n_2} \\
&+ a_{4,1} + a_{4,2} + a_{4,3} + a_{4,4} + \cdots + a_{4,n_2} \\
&+ \cdots \\
&+ a_{n_1,1} + a_{n_1,2} + a_{n_1,3} + a_{n_1,4} + \cdots + a_{n_1,n_2}
\end{aligned}$$

Here, $\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}$ represents summing over each row first and then summing the results, while $\sum_{j=1}^{n_2}\sum_{i=1}^{n_1}$ will represent summing over the columns and then summing those results. The term $\sum_{i=1,j=1}^{n_1,n_2} a_{i,j}$ represents the summation over the rectangular array, irrespecive of the order. The product of two sums can be expanded into a double sum as follows:

$$\begin{aligned}
\left(\sum_{i=1}^{m} a_i\right)\left(\sum_{j=1}^{n} b_j\right) &= (a_1 + a_2 + \cdots + a_m)(b_1 + b_2 + \cdots + b_n) \\
&= a_1b_1 + a_1b_2 + a_1b_3 + a_1b_4 + \cdots + a_1b_n \\
&+ a_2b_1 + a_2b_2 + a_2b_3 + a_2b_4 + \cdots + a_2b_n \\
&+ a_3b_1 + a_3b_2 + a_3b_3 + a_3b_4 + \cdots + a_3b_n \\
&+ \cdots \\
&+ a_mb_1 + a_mb_2 + a_mb_3 + a_mb_4 + \cdots + a_mb_n \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} a_ib_j = \sum_{i=1}^{m} a_i \sum_{j=1}^{n} b_j
\end{aligned}$$

which is actually rather intuitive given how the expansion of the standard expansion of the term $(a+b)^2$ plays out, a more elementary application of the distributive property which the above equation generalizes over. For a triangular matrix, in this case the lower triangular matrix, the sum is given by

$$\begin{aligned}\sum_{1\le j\le n} a_{i,j} &= \sum_{i=1}^{n}\sum_{j=1}^{i} a_{i,j} = \sum_{j=1}^{n}\sum_{i=j}^{n} a_{i,j} = \sum_{j=0}^{n-1}\sum_{j=1}^{n-j} a_{i+j,i} \\ &= a_{1,1} \\ &+ a_{2,1} + a_{2,2} \\ &+ a_{3,1} + a_{3,2} + a_{3,3} \\ &+ a_{4,1} + a_{4,2} + a_{4,3} + a_{4,4} \\ &+ \cdots \\ &+ a_{n,1} + a_{n,2} + a_{n,3} + a_{n,4} + \cdots + a_{n,n}\end{aligned}$$

where the term $\sum_{1\le j\le n} a_{i,j}$ denotes the summation over all elements in a lower triangular array including the diagonal. The first notation variation will sum up each row to the ith element then aggregate while the second notation sums each column starting from the jth element downwards then aggregate the sums. The final expression will sum along the diagonal where $j = 0$ represents the main diagonal and $j = n - 1$ is the first off-diagonal, which is a single term.

Example. Say we would like to expand the product of $(1 + x_i)$ from 1 to n. We have

$$\prod_{i=1}^{n}(1 + x_i) = 1 + \sum_{k=1}^{n}\left(\sum_{1\le i_1<\cdots<i_k\le n}\prod_{j=1}^{k} x_{i_j}\right)$$

This formula represents the *multinomial expansion* of a product. When you expand the equation by hand, you get the product

$$\prod_{i=1}^{n}(1 + x_i) = (1 + x_1)(1 + x_2)\cdots(1 + x_n)$$

If we break this down, we see that the outer summation $\sum_{k=1}^{n}$ will go through each possible summation size in terms of the variables in question, and that the inner summation $\sum_{1\le i_1<\cdots<i_k\le n}$ will iterate through each possible unique product of the variables. while ensuring that they are unique. Not sure how this works, but if all x_i are the same, then wesee that the equation actually simplifies to a subset of the binomial theorem

$$(1 + x)^n = \sum_{k=0}^{n}\binom{n}{k}x^k$$

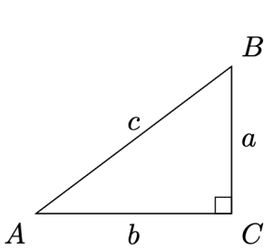
where $\binom{n}{k}$ is the binomial coefficient representing the numebr of ways to choose k elements from a set of n distinct elements.

2 Trigonometry

You can't escape this.

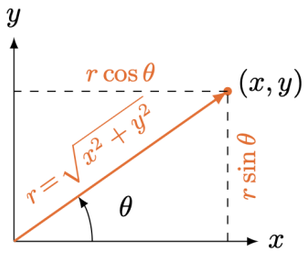
2.1 Definitions

I mean, where do I start? The basic trigonometric functions are defined as the ratios between the angles of a right triangle. I will not show how these ratios remain the same given the same angle, nor will I go into great mathmatical detail of how to prove these items. However, we still have to go over this. Don't ask me why.



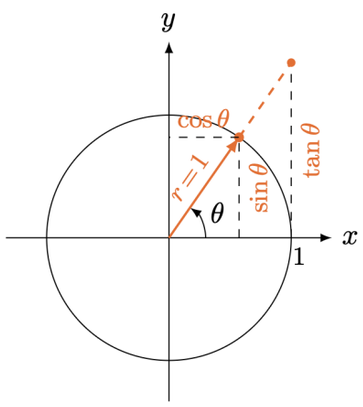
$$\sin A = \frac{a}{c},\ \cos A = \frac{b}{c}$$

$$\tan A = \frac{a}{b},\ \cot A = \frac{b}{a}$$

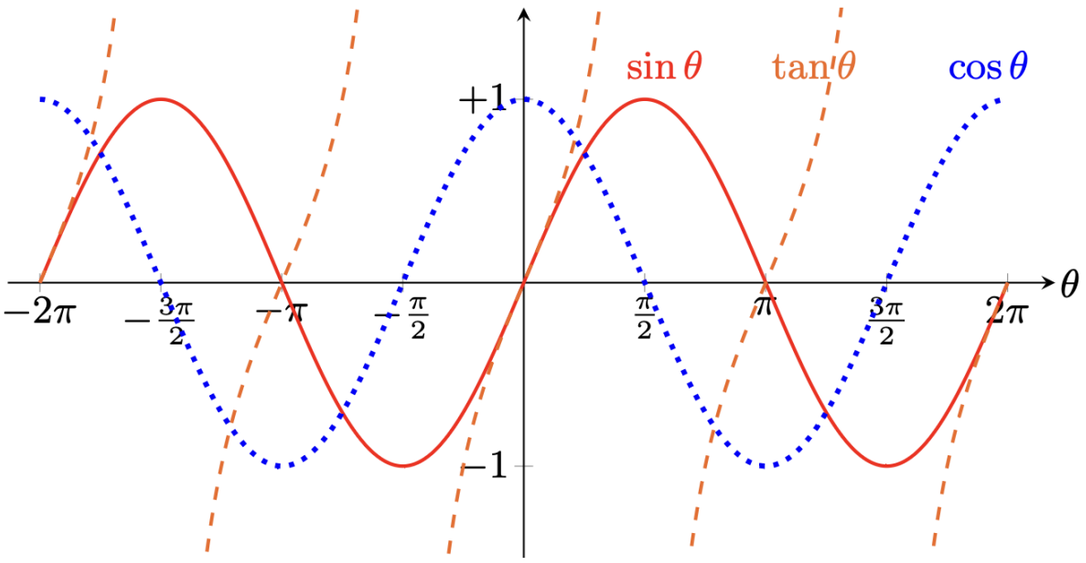


$$\sin\theta = \frac{y}{r},\ \cos\theta = \frac{x}{r}$$

$$\tan\theta = \frac{y}{x},\ \cot\theta = \frac{x}{y}$$



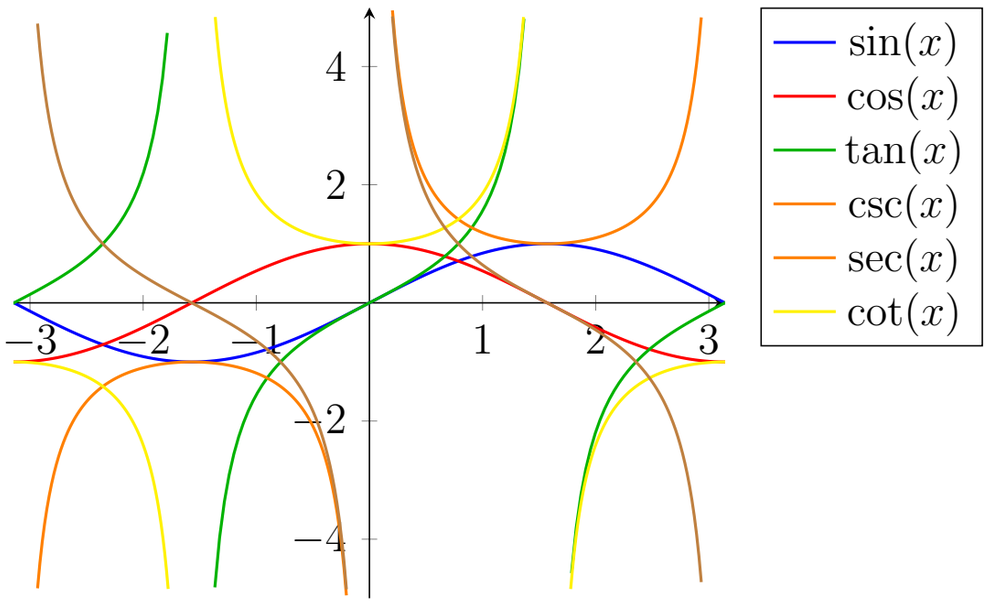
The functions of trigonometric functions can also be plotted out as follows:



2.2 Basic Properties and Inverse Functions

	$\sin\theta$	$\cos\theta$	$\tan\theta$	$\csc\theta$	$\sec\theta$	$\cot\theta$
Definition	y/r	x/r	y/x	r/y	r/x	x/y
Period	2π	2π	π	2π	2π	π
Range	$[-1,1]$	$[-1,1]$	$(-\infty,\infty)$	$(-\infty,-1]\cup[1,\infty)$		$(-\infty,\infty)$
Zeros	$n\pi$	$(n+\frac{1}{2})\pi$	$n\pi$			$(n+\frac{1}{2})\pi$
Poles			$(n+\frac{1}{2})\pi$	$n\pi$	$(n+\frac{1}{2})\pi$	$n\pi$

Note: n is an integer.



We can also see that there are certain useful symmetric properties of the trigonometric functions

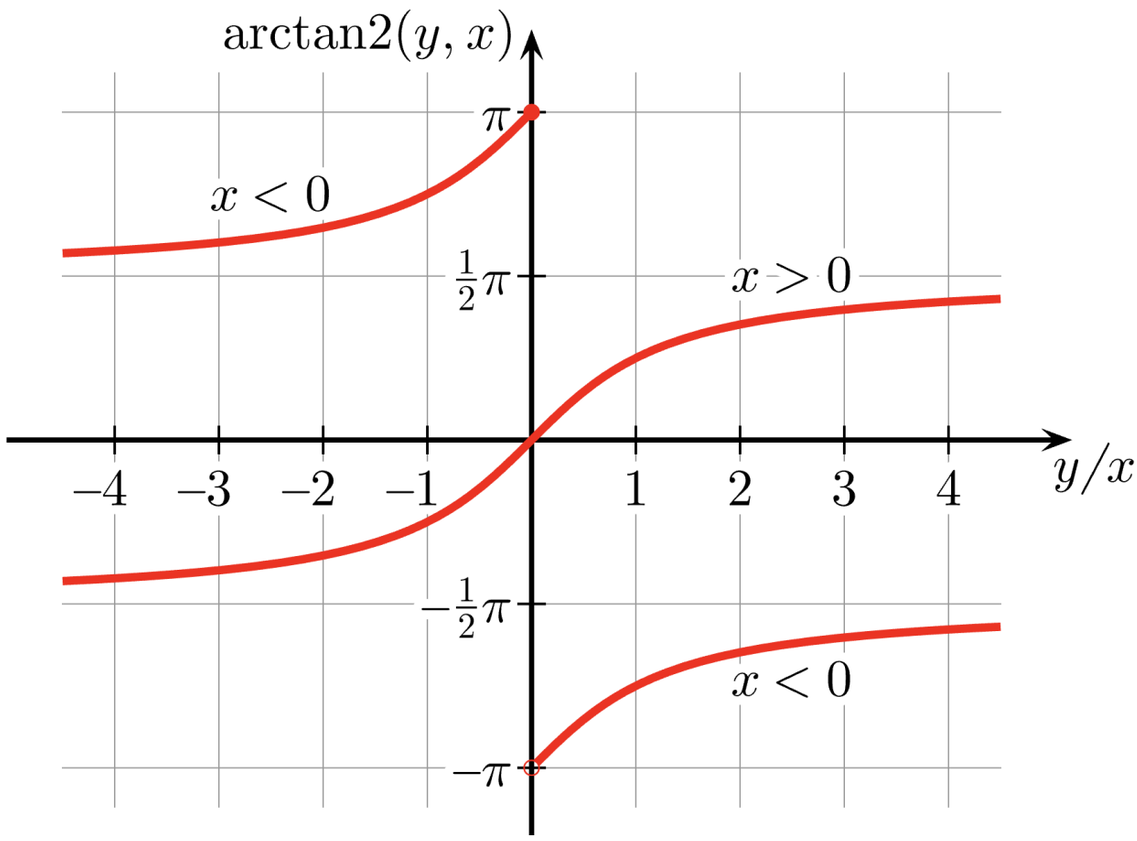
$$\begin{aligned}
&\sin(-\theta) = -\sin(\theta) && \sin(\pi-\theta) = \sin(\theta) && \sin(\pi+\theta) = -\sin(\theta)\\
&\cos(-\theta) = \cos(\theta) && \cos(\pi-\theta) = -\cos(\theta) && \cos(\pi+\theta) = -\cos(\theta)\\
&\tan(-\theta) = -\tan(\theta) && \tan(\pi-\theta) = -\tan(\theta) && \tan(\pi+\theta) = \tan(\theta)
\end{aligned}$$

There are also some common inverse functions associated with the functions.

Function	**sin**	**cos**	**tan**	**csc**	**sec**	**cot**
Inverse	$\textbf{sin}^{-1}$ **arcsin**	$\textbf{cos}^{-1}$ **arccos**	$\textbf{tan}^{-1}$ **arctan**	$\textbf{csc}^{-1}$	$\textbf{sec}^{-1}$	$\textbf{cot}^{-1}$
Domain	$[-1,1]$	$[-1,1]$	$(-\infty,\infty)$	$(-\infty,-1]\cup[1,\infty)$		$(-\infty,\infty)$
Range	$[-\frac{\pi}{2},\frac{\pi}{2}]$	$[0,\pi]$	$(-\frac{\pi}{2},\frac{\pi}{2})$	$[-\frac{\pi}{2},\frac{\pi}{2}]\backslash\{0\}$	$[0,\pi]\backslash\{\frac{\pi}{2}\}$	$(-\frac{\pi}{2},\frac{\pi}{2})$

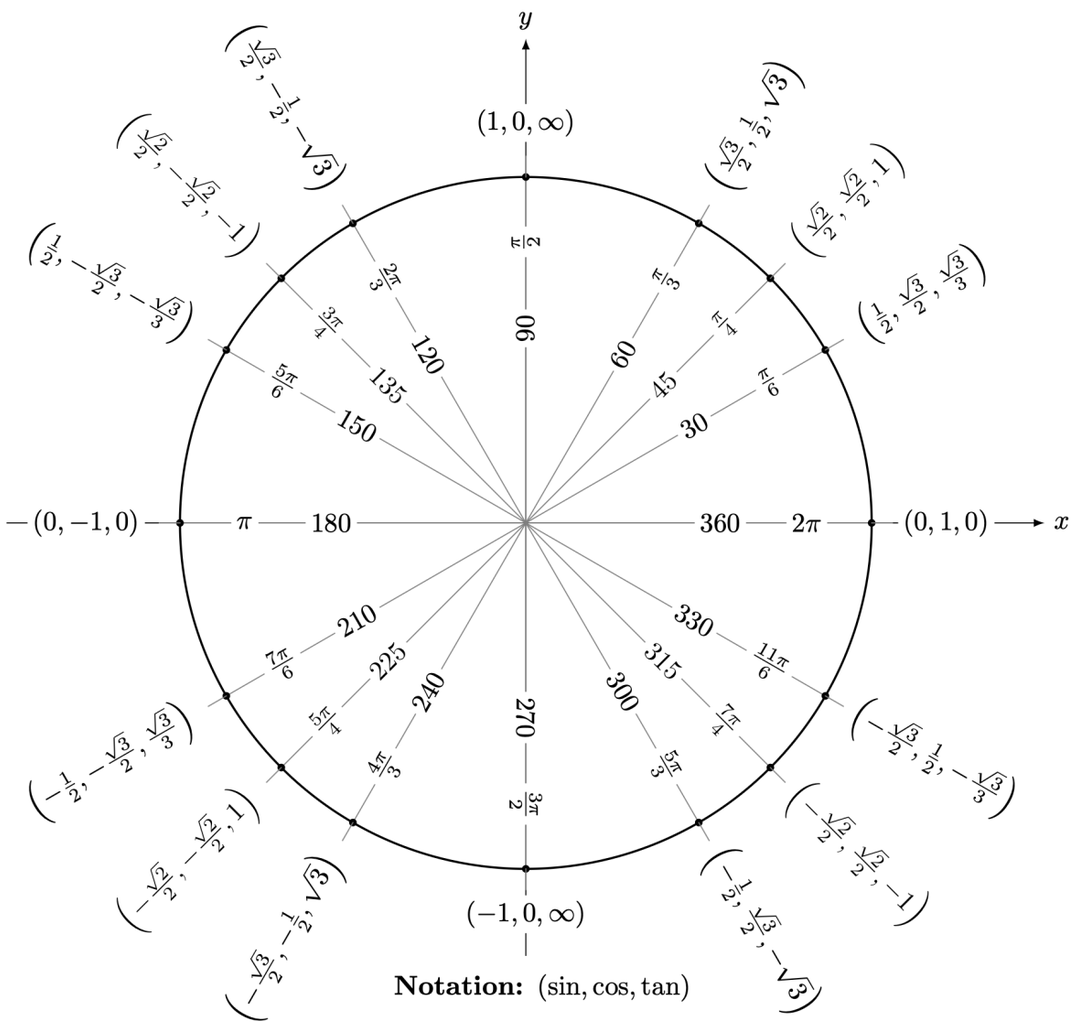
But also some interesting extensions of the commonly known inverse functions, with the example being arctan2, a function that effectively doubles the domain of the function while preserving its properties for the purpose of, say, converting values from cartesian to spherical coordinates for the azimuthal angle ϕ.

$$\operatorname{arctan2}(y,x) = \begin{cases}
\arctan\left(\frac{y}{x}\right) & \text{if } x > 0,\\
\arctan\left(\frac{y}{x}\right) + \pi & \text{if } x < 0 \text{ and } y \geq 0,\\
\arctan\left(\frac{y}{x}\right) - \pi & \text{if } x < 0 \text{ and } y < 0,\\
+\frac{\pi}{2} & \text{if } x = 0 \text{ and } y > 0,\\
-\frac{\pi}{2} & \text{if } x = 0 \text{ and } y < 0,\\
0 & \text{if } x = 0 \text{ and } y = 0.
\end{cases}$$



2.3 Special Angles and Function Values

There are a few special angles that are worth remembering for the trigonometric functions mentioned above, given by the wheel below.



2.4 Trigonometric Identities

Reciprocal and Quotient Identities

$$\csc\theta = \frac{1}{\sin\theta} \qquad \sec\theta = \frac{1}{\cos\theta} \qquad \cot\theta = \frac{1}{\tan\theta}$$

$$\tan\theta = \frac{\sin\theta}{\cos\theta}, \qquad \cot\theta = \frac{\cos\theta}{\sin\theta}$$

Cofunction Identities

$$\begin{aligned} \sin\left(\frac{\pi}{2} - \theta\right) &= \cos\theta, & \cos\left(\frac{\pi}{2} - \theta\right) &= \sin\theta \\ \tan\left(\frac{\pi}{2} - \theta\right) &= \cot\theta, & \cot\left(\frac{\pi}{2} - \theta\right) &= \tan\theta \\ \sec\left(\frac{\pi}{2} - \theta\right) &= \csc\theta, & \csc\left(\frac{\pi}{2} - \theta\right) &= \sec\theta \end{aligned}$$

Pythagorean Identities

$$\begin{aligned} \sin^2\theta + \cos^2\theta &= 1 \\ 1 + \tan^2\theta &= \sec^2\theta \\ 1 + \cot^2\theta &= \csc^2\theta \end{aligned}$$

Even–Odd Symmetry

$$\begin{aligned} \sin(-\theta) &= -\sin\theta, & \cos(-\theta) &= \cos\theta, & \tan(-\theta) &= -\tan\theta \\ \csc(-\theta) &= -\csc\theta, & \sec(-\theta) &= \sec\theta, & \cot(-\theta) &= -\cot\theta \end{aligned}$$

Sum and Difference Formulas

$$\begin{aligned} \sin(\alpha \pm \beta) &= \sin\alpha\cos\beta \pm \cos\alpha\sin\beta \\ \cos(\alpha \pm \beta) &= \cos\alpha\cos\beta \mp \sin\alpha\sin\beta \\ \tan(\alpha \pm \beta) &= \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta} \end{aligned}$$

Double Angle Formulas

$$\begin{aligned} \sin(2\theta) &= 2\sin\theta\cos\theta \\ \cos(2\theta) &= \cos^2\theta - \sin^2\theta \\ &= 2\cos^2\theta - 1 \\ &= 1 - 2\sin^2\theta \\ \tan(2\theta) &= \frac{2\tan\theta}{1 - \tan^2\theta} \end{aligned}$$

Half Angle Formulas

$$\begin{aligned} \sin^2\left(\frac{\theta}{2}\right) &= \frac{1 - \cos\theta}{2} \\ \cos^2\left(\frac{\theta}{2}\right) &= \frac{1 + \cos\theta}{2} \\ \tan\left(\frac{\theta}{2}\right) &= \frac{\sin\theta}{1 + \cos\theta} = \frac{1 - \cos\theta}{\sin\theta} \end{aligned}$$

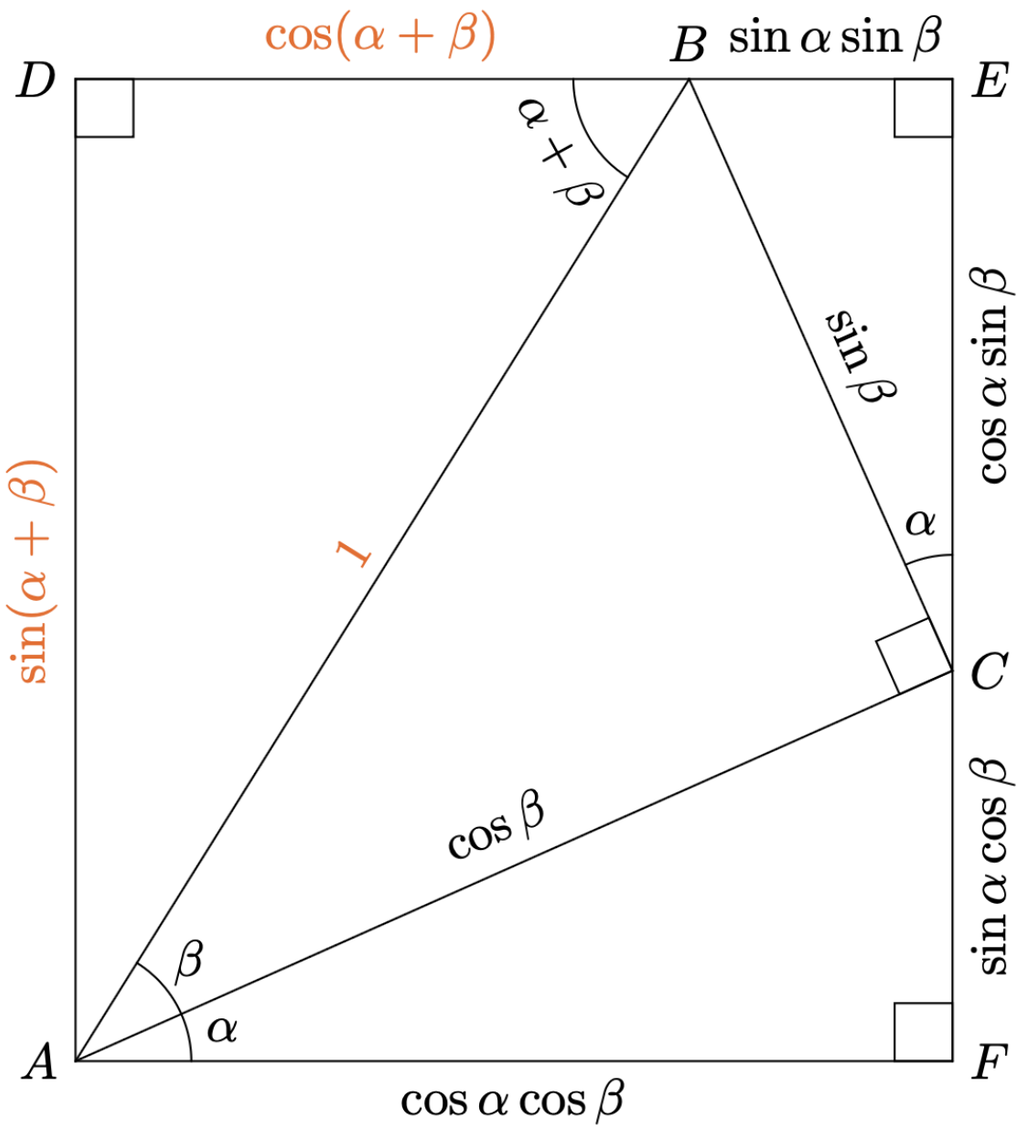
Product-to-Sum Identities

$$\sin\alpha\sin\beta = \frac{1}{2}\big[\cos(\alpha-\beta)-\cos(\alpha+\beta)\big]$$
$$\cos\alpha\cos\beta = \frac{1}{2}\big[\cos(\alpha-\beta)+\cos(\alpha+\beta)\big]$$
$$\sin\alpha\cos\beta = \frac{1}{2}\big[\sin(\alpha+\beta)+\sin(\alpha-\beta)\big]$$

Sum-to-Product Identities

$$\sin\alpha+\sin\beta = 2\sin\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right)$$
$$\sin\alpha-\sin\beta = 2\cos\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right)$$
$$\cos\alpha+\cos\beta = 2\cos\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right)$$
$$\cos\alpha-\cos\beta = -2\sin\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right)$$

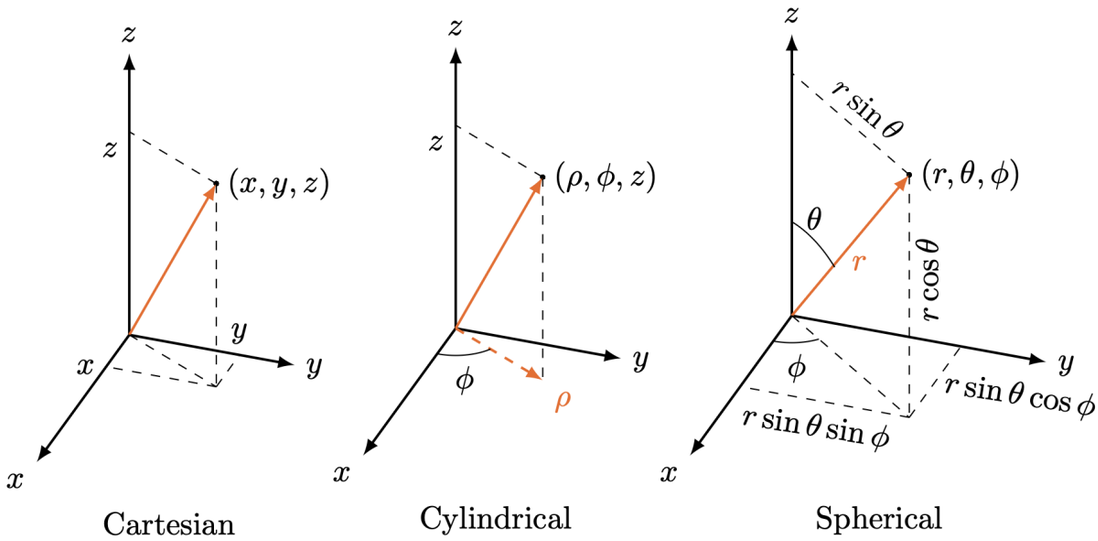
A rather nice photo to sum the secion up is by relating the angles to each other using the following image



2.5 The Spherical Coordinate System

While the expansion of the cartesian coordinate system into three dimensions is the logical linear expansion to take, some interesting basis begin to form if we consider the angles as a

unit of measurement in three dimensions



The cylindrical coordinate system elevates the polar coordinate system into three dimensions with the addition of the z axis, while the spherical system is a more organic translation of the polar coordinate system into three dimensions. While the former is best suited for describing not only cylinder-like structures yet also helices, while the latter is best suited for rotations in three-dimensional space, which is particularly potent in the field of quantum computing.

For example, the spherical coordinate system is commonly used to represent qubit states on the Bloch sphere, employing a radius (r), a polar angle (θ), and an azimuthal angle (ϕ) to represent a point in three-dimensional space. The azimuthal angle ϕ is measured in the xy-plane from the positive x-axis with common values ranging from $(-\pi, \pi]$ or $(0, 2\pi]$. The polar angle is commonly measured from the positive z axis towards the xy-plane, with values ranging from $[0, \pi]$. Note that $r \in \mathbb{R}$, meaning that we can cover the other half of the range simply by flipping the sign around.

Conversion is relatively simple, with conversion to and from spherical to cartesian being as follows

$$\begin{aligned} x &= r\sin\theta\cos\phi & r &= \sqrt{x^2+y^2+z^2} \\ y &= r\sin\theta\sin\phi & \phi &= \operatorname{arctan2}(y, x) \\ z &= r\cos\theta & \theta &= \arccos\frac{z}{r} \end{aligned}$$

where arctan2 was previously defined as a optimal inverse mapping onto the range of $[-\pi, \pi]$. We also have a few definitions

Definition 2.1. The Law of Sines is defined as

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$$

Definition 2.2. The Law of Cosines is defined as

$$a^2 = b^2 + c^2 - 2bc\cos A$$

which can be rewritten as

$$\cos A = \frac{b^2 + c^2 - 2bc}{a^2}$$

Definition 2.3. The Law of Tangents is defined as

$$\frac{a-b}{a+b} = \frac{\tan\frac{1}{2}(A-B)}{\tan\frac{1}{2}(A+B)}$$

3 Complex Numbers

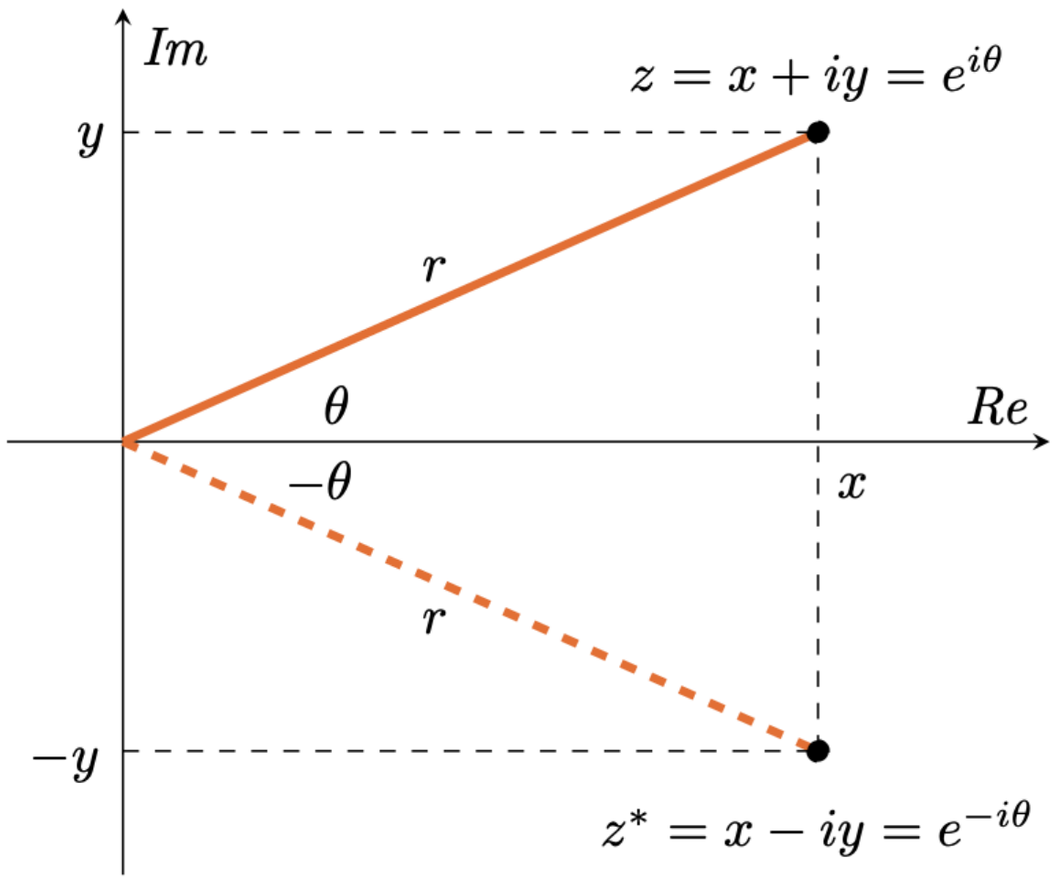
We consider numbers to be complex when they compose of a real and imaginary part, and they are not only fundamental to a complete understanding of algebra and mathmatics as a whole, but also form the backbone of quantum mechanics, and, by extension, quantum computing. Mastering complex numbers is like Rosie mastering the rivet gun, so we have to study it.

3.1 Cartesian Form

Definition 3.1. A complex number z is defined as

$$z = x + iy, \qquad x, y \in \mathbb{R}, \quad i^2 = -1$$

This is called the **cartesian form** of the complex number z and corresponds to a point in the two-dimensional complex plane. We commonly refer to i as the imaginary unit. It may seem ironic that we need imaginary numbers in quantum computing, or that we really need the imaginary number. Take it as you may.



Complex numbers not motivated by quantum computing. In the numbers system, we have the real numbers $\mathbb{N}$, the integer numbers $\mathbb{Z}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$. the set incursions go this way $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{R} \subset \mathbb{C}$, all supersets of the preceeding set. The set of all complex numbers $\mathbb{C}$ is closed over all algabraic operations, which include addition, substraction, multiplication, division, power, and root, and is considered the superset of all numbers.

Definition 3.2. The basic components of a complex number are defined as follows.

$$\Re(z) = x, \Im(z) = y$$

which are the real and imaginary components of a complex number z. Of course the complex number itself has a few interesting properties, such as $i^2 = -1$, $i^3 = -i$, and $i^4 = 1$. The complex conjugate of a complex number z is defined as

$$z^* = x - iy$$

defined as inverting the sign of the imaginary component. We can express the modulus (vector length) and argument (angle with respect to the real axis) (which are r and θ in polar coordinates), as follows:

$$r = |z| = \sqrt{zz^*} = \sqrt{x^2 + y^2}$$

A very convenient property derived from algebra is that $zz^* = x^2 + y^2$.

$$\theta = arg(z)$$

For the angle, we note that

$$\tan \frac{y}{x} \Longrightarrow \text{arctan2}(x, y)$$

where arctan2 has been defined in the previous section.

Example. Given $z = 1 + \sqrt{3}i$, we have

$$z^* = 1 - \sqrt{3}\, i.$$

$$|z| = \sqrt{1^2 + (\sqrt{3})^2} = 2.$$

$$zz^* = (1 + \sqrt{3}\, i)(1 - \sqrt{3}\, i) = 1 - (\sqrt{3}\, i)^2 = 1 + 3 = 4 = |z|^2.$$

$$\theta = \arctan\left(\frac{\sqrt{3}}{1}\right) = \frac{\pi}{3}.$$

3.2 Exponential Form

Now it is worth noting that while we commonly write $z = x + iy$ to represent a complex numnber, we like to use the following definitions of the complex number in polar form to represent a complex number itself, defined as $z = r(\cos\theta + i \sin\theta)$. However, multiplication and its inverse operation, division, becomes unnecessarily difficult givne the presence of another notation, namely **exponential form**.

Definition 3.3. The exponential/euler forms of the complex numbers can be thought of as a circular form of the function $z = x + iy$. In polar coordinates, we can rewrite this number as

$$z = r\cos\theta + i\sin\theta, \qquad r \in \mathbb{R}$$

Conversely, the conversion between cartesian and polar are

$$x = r\cos\theta \qquad y = r\sin\theta$$

The formula for z above can be rewritten as

$$z = re^{i\theta}$$

Theorem 3.4. Euler's formula states that for any complex number $z = r\cos\theta + i\sin\theta$, we have:

$$e^{i\theta} = \cos\theta + i\sin\theta$$

Proof. Euler's formula can be proven using the Taylor series expansion for the functions:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$$

if we replace x in the formula for e^x with e^{ix}, we then have

$$\begin{aligned} e^{ix} &= 1 + (ix) + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \frac{(ix)^6}{6!} + \frac{(ix)^7}{7!} + \cdots \\ &= 1 + ix - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - i\frac{x^7}{7!} + \cdots \\ &= \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots\right) + \left(ix - i\frac{x^3}{3!} + i\frac{x^5}{5!} - i\frac{x^7}{7!} + \cdots\right) \\ &= \cos\theta + i\sin\theta \end{aligned}$$

which sums up essentially what mathmaticians call the most beautiful proof man has known. If this proof were a female robot, mathematicians would compose harmonic waves and produce digital flowers in LaTeXto please Euler's genius. □

As noted before, we know that the set of all algebraic operations is well defined and closed on the set of all complex numbers $\mathbb{C}$. Addition will be easier in cartesian form, while multiplication will be considerably simpler in exponential form. Conversion between the two is also not difficult:

	Cartesian Form	**Exponential Form**
	$z = x + iy$	$z = re^{i\theta}$
Conjugate	$z^* = x - iy$	$z^* = re^{-i\theta}$
Modulus	$\lvert z\rvert = \sqrt{zz^*} = \sqrt{x^2 + y^2}$	$\lvert z\rvert = r$
Conversion	$x = r\cos\theta$	$r = \sqrt{x^2 + y^2}$
	$y = r\sin\theta$	$\theta = \arctan2(y, x)$

3.3 Basic Operations

As we touched upon earlier, the set of all complex numbers are closed on operations of addition and subtraction:

$$z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2).$$

$$z_1 - z_2 = (x_1 - x_2) + i(y_1 - y_2).$$

As well on multiplication and division. For

$$z_1 = r_1 e^{i\theta_1} \qquad z_2 = r_2 e^{i\theta_2}$$

We have:

$$z_1 \cdot z_2 = r_1 r_2 e^{i(\theta_1 + \theta_2)}.$$

$$\frac{z_1}{z_2} = \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}.$$

We also have the following properties for the conjugates of complex numbers

$$\begin{aligned} |z| &= |z^*| \\ (z_1 \pm z_2)^* &= z_1^* \pm z_2^* \\ (z_1 \cdot z_2)^* &= z_1^* \cdot z_2^* \\ (z_1/z_2)^* &= z_1^*/z_2^* \\ (z^x)^* &= (z^*)^x \quad x \in \mathbb{R} \\ (x^z)^* &= x^{z^*} \quad x \in \mathbb{R} \end{aligned}$$

Where the last two are not immediately obvious. To prove that $(z^x)^* = (z^*)^x$, it is useful to write out z using the complex notation $re^{i\theta}$, and the last property is best proven using the identity $a^b = e^{b \ln a}$. As for powers and roots of complex numbers, we have

Theorem 3.5. De Moivre's theorem states that

$$(\cos\theta + i\sin\theta)^s = \cos s\theta + i\sin s\theta$$

which is conveniently derived from the fact that $z^s = r^s e^{is\theta}$.

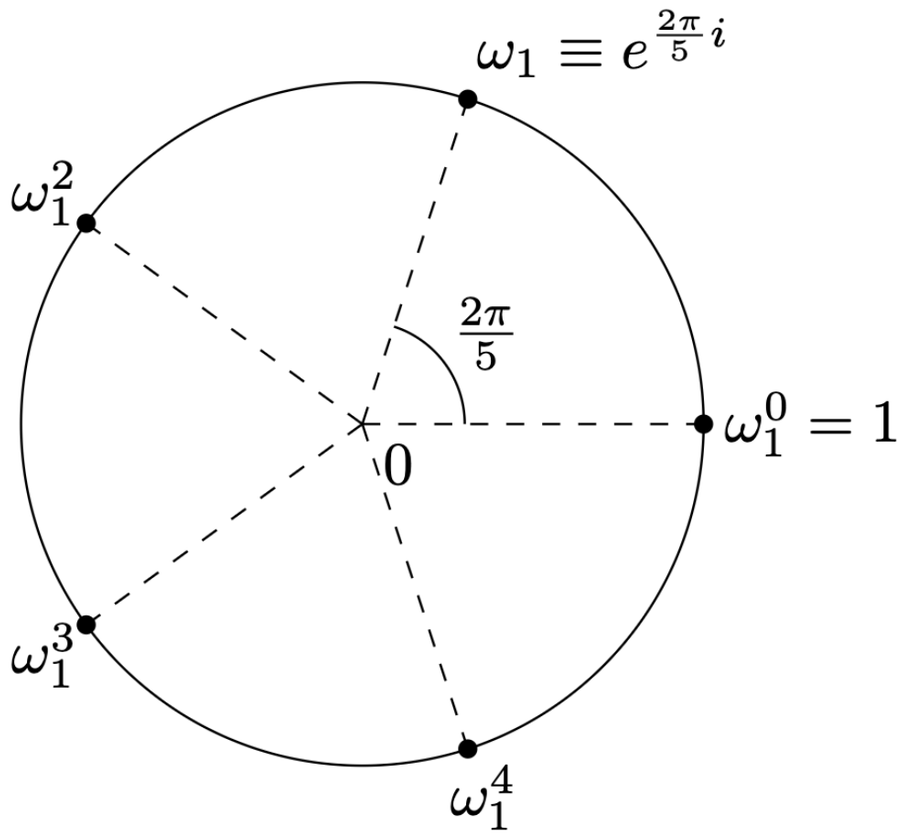
In particular, an application of this theorem is where we describe the roots of unity. Any root of unity can be descibed by a power of the first root of unity,

$$\omega_1 = e^{\frac{2\pi i}{n}}$$

the n-th roots of unity ($n \in \mathbb{N}$) are given by

$$\omega_k = e^{\frac{2\pi i}{n}k} = w_1^k \qquad k = 1, 2, \ldots, n-1$$

which essentially says that there are n roots to the complex polynomial. For $n = 5$, we have the 5 roots of unity given by



In general, we say that there are n values of k that satisfy the equation $\omega_1^n = e^{(\frac{2\pi i}{n})^n} = 1$. From this, we can generalize what we know into the summations over ω_k, which is any k-th root of unity except for $\omega = 1$.

$$\sum_{k=0}^{n-1} \omega_k = \sum_{k=0}^{n-1} \omega_1^k = 0$$

This formula can be conviniently proven by applying the formula for summing a geometric sequence to the summation. From this, we can conviniently derive a useful mathmatical condition, being

Example. The DFT Orthonormality condition depends on two parameters k and l, and is stated as follows

$$\frac{1}{N}\sum_{n=0}^{N-1} e^{-\frac{2\pi i}{N}kn} e^{\frac{2\pi i}{N}ln} = \delta_{k-l \bmod N}$$

where $\delta_{k-l} \pmod N = 1$ if and only if $k \equiv l \pmod N$, else 0. It is saying that when k is congruent to l, equivalent to $k - l = mN$, where the difference between k and l is divisible by some integer m. The $\delta_{k-l \bmod N}$ term is a Kronecker delta of $k \equiv l \pmod N$, where the result is 1 if $k \equiv l \pmod N$ holds and 0 in the case of $k \not\equiv l \pmod N$. If we define $\omega = e^{i\frac{2\pi}{N}}$ as a primitive Nth root of unity (satisfying $\omega^N = 1$), we have the derivation

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1} e^{-\frac{2\pi i}{N}kn}\, e^{\frac{2\pi i}{N}ln} &= \frac{1}{N}\sum_{n=0}^{N-1} \omega^{-kn}\omega^{ln} \\
&= \frac{1}{N}\sum_{n=0}^{N-1} \omega^{n(l-k)} \\
&= \begin{cases} \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} 1 = 1, & \text{if } l \equiv k \pmod N, \\ \dfrac{1}{N}\dfrac{1-\omega^{(l-k)N}}{1-\omega^{(l-k)}} = 0, & \text{if } l \not\equiv k \pmod N \end{cases} \\
&= \delta_{k-l \bmod N},
\end{aligned}$$

where we used the fact that $\omega^{n^N} = \big(\omega^N\big)^n = 1^n = 1$ for $n \in \mathbb{N}$.

3.4 Advanced Operations

It's probably best to illustrate more advanced operations on complex numbers with the help of some examples

Example. Evaluating $\sqrt{i}$ or $\sqrt{\sqrt{1}}$ gives:

$$\sqrt{i} = \left(e^{\frac{\pi i}{2}}\right)^{\frac{1}{2}} = e^{\frac{\pi i}{4}} = \cos\frac{\pi}{4} + i\sin\frac{\pi}{4} = \frac{1}{\sqrt{2}}(1+i)$$

The inverse is given by

$$\left(\frac{1}{\sqrt{2}}(1+i)\right)^2 = \frac{1}{2}(1+2i+i^2) = \frac{(2i)}{2} = i$$

Example. Evaluating

$$\left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)^{50}$$

gives

$$\left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)^{50} = \left(e^{\frac{\pi i}{3}}\right)^{50} = e^{\frac{50\pi i}{3}} = e^{\left(16+\frac{2}{3}\right)\pi i} = e^{\frac{2\pi i}{3}} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i.$$

Example. Evaluating

$$2^{3+4i}$$

gives us

$$2^{3+4i} = 2^3 \cdot 2^{4i} = 8 \cdot e^{4\ln(2)\,i} = 8\cos\big(4\ln 2\big) + i\,8\sin\big(4\ln 2\big).$$

Example. Evaluating

$$\cos(3+4i)$$

gives us

$$\begin{aligned}
\cos(3+4i) &= \frac{1}{2}\big(e^{i(3+4i)} + e^{-i(3+4i)}\big) \\
&= \frac{1}{2}\big(e^{-4+3i} + e^{4-3i}\big) \\
&= \frac{1}{2}e^{-4}(\cos 3 + i\sin 3) + \frac{1}{2}e^{4}(\cos 3 - i\sin 3) \\
&= \frac{1}{2}(e^{-4} + e^{4})\cos 3 + i\,\frac{1}{2}(e^{-4} - e^{4})\sin 3
\end{aligned}$$

Example. If we have the equation $z^5 = \frac{1}{2} + \frac{\sqrt{3}}{2}i$, solving for z gives

$$z_k = e^{\frac{\pi i}{15}} e^{\frac{2k\pi i}{5}} \quad k = 0, 1, 2, 3, 4.$$

It is worth noting that there exists a way to express the trigonometric functions sin and cos as a function of euler's number. We know that

$$e^{i\theta} = \cos\theta + i\sin\theta \qquad e^{-i\theta} = \cos\theta - i\sin\theta$$

from this, we can derive that

$$\begin{aligned} e^{i\theta} + e^{-i\theta} &= \cos\theta + i\sin\theta + \cos\theta - i\sin\theta \\ e^{i\theta} + e^{-i\theta} &= \cos\theta + \cos\theta \\ \cos\theta &= \frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right) \end{aligned}$$

and that

$$\begin{aligned} e^{i\theta} + e^{-i\theta} &= \cos\theta + i\sin\theta - (\cos\theta - i\sin\theta) \\ e^{i\theta} - e^{-i\theta} &= i\sin\theta - (i\sin\theta) \\ \sin\theta &= \frac{1}{2i}\left(e^{i\theta} - e^{-i\theta}\right) \end{aligned}$$

Another way of expressing this is by saying

$$\cos x = \Re(e^{i\theta}) = \frac{e^{ix} + e^{-ix}}{2}$$
$$\sin x = \Im(e^{i\theta}) = \frac{e^{ix} - e^{-ix}}{2i}$$

One final yet very important item to rememebr throughout the curriculum is that powers for complex numbers are **rotations**.

4 Sets, Groups, and Functions

This chapter is mainly going to fly over the various mathmatical concepts that make up the backbone of many mathmatical fields, including those relevant to quantum computing.

4.1 Sets

The concepts of sets are fundamental to many areas of mathmatics. A set is a well-defined collection of distinct objects, which is also an object in its own right.

Definition 4.1 (Set)**.** A set is an (unordered) collection of objects, which are said to be elements or members of the set.

Let $A = \{2, 4, 6, 8\}$ be the set of even numbers less than 10. The elements of this set are 2, 4, 6, and 8. We can also write sets using set-builder notation, such as $B = \{x \mid x \text{ is a vowel in the English alphabet}\} = \{a, e, i, o, u\}$.

Definition 4.2 (Tuple)**.** A tuple (or sequence) is an ordered list of elements.

Consider the tuple $T = (3, 1, 4, 1, 5)$. This is an ordered sequence where the first element is 3, the second is 1, the third is 4, the fourth is 1, and the fifth is 5. Note that the order matters and repetition is allowed, so $(3, 1, 4, 1, 5) \neq (1, 1, 3, 4, 5)$.

Definition 4.3 (Cardinality)**.** The cardinality of a set A, denoted $|A|$, is the number of elements in A.

For the set $A = \{2, 4, 6, 8\}$, the cardinality is $|A| = 4$ since there are 4 elements. For the set $B = \{x, y, z\}$, we have $|B| = 3$. The empty set has cardinality $|\emptyset| = 0$.

We can categorize sets based on their cardinality into finite, countably infinite, or uncountably infinite. A set is said to be countably infinite if it can be bijectively mapped to the set of natural numbers $\mathbb{N}$ such as the set of integers $\mathbb{Z}$, while the set is said to be uncountably infinite if there is no one-to-one to the set of natural numbers, such as the set of real numbers $\mathbb{R}$.

Example. Finite set: $A = \{1, 2, 3, 4, 5\}$ has $|A| = 5$. Countably infinite set: The set of integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ can be put in one-to-one correspondence with the natural numbers $\mathbb{N}$, so $|\mathbb{Z}| = |\mathbb{N}| = \aleph_0$. Uncountably infinite set: The set of real numbers $\mathbb{R}$ cannot be put in one-to-one correspondence with $\mathbb{N}$, so $|\mathbb{R}| > \aleph_0$.

Definition 4.4 (Subset and Superset)**.** We call B a subset of A, denoted $B \subseteq A$ if $\forall b \in B, b \in A$. In this case, A is a superset of B, denoted $B \supseteq A$. If B is a subset of A but not equal to A, then B is called a proper subset of A, denoted $B \subset A$.

Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{2, 4\}$. Then $B \subseteq A$ since every element of B is also in A. We can also say $A \supseteq B$. Since $B \neq A$, we have $B \subset A$ (B is a proper subset of A). Additionally, $A \subseteq A$ since every set is a subset of itself.

Definition 4.5 (Union)**.** The union of two sets A and B, denoted by $A \cup B$, is the set containing all the elements in A, B, or both.

Let $A = \{1, 3, 5, 7\}$ and $B = \{2, 3, 6, 7, 8\}$. Then $A \cup B = \{1, 2, 3, 5, 6, 7, 8\}$, which contains all elements that appear in either set A or set B (or both).

Definition 4.6 (Intersection)**.** The intersection of two sets A and B, denoted by $A \cap B$, is the set containing all the elements in both A and B.

Using the same sets $A = \{1, 3, 5, 7\}$ and $B = \{2, 3, 6, 7, 8\}$, we have $A \cap B = \{3, 7\}$, which contains only the elements that appear in both sets.

Corollary 4.7 (Disjoint Sets)**.** Two or more sets are said to be disjoint if they have no elements in common, that is, their intersection is the empty set: $A \cap B = \emptyset$

Let $A = \{1, 3, 5\}$ and $B = \{2, 4, 6\}$. These sets are disjoint because $A \cap B = \emptyset$ - they share no common elements.

Definition 4.8 (Difference)**.** The difference of two sets A and B, denoted by $A - B$ or $A \setminus B$, is the set containing all the elements in A but not in B.

Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{3, 4, 6, 7\}$. Then $A - B = \{1, 2, 5\}$, which contains the elements in A that are not in B. Similarly, $B - A = \{6, 7\}$.

Definition 4.9 (Universal Set)**.** The universal set, denoted by U, is the set that contains all elements under consideration, usually in relation to a particular problem or discussion. Every other set in that context is a subset of the universal set U.

In a problem about students at a university, the universal set might be $U =$ all students at the university. If we're discussing card games, the universal set could be $U =$ all 52 cards in a standard deck.

Definition 4.10 (Complement)**.** The complement of a set A, denoted by $\overline{A}$ or A^c, is the set of all elements in the universal set U that are not in set A.

Let $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ be the universal set and $A = \{2, 4, 6, 8, 10\}$. Then the complement of A is $\overline{A} = \{1, 3, 5, 7, 9\}$, containing all elements in U that are not in A.

Definition 4.11 (Cartesian Product)**.** The cartesian product of two sets A and B, denoted $A \times B$, is the set of all ordered pairs $(a, b), a \in A, b \in B$.

Let $A = \{1, 2\}$ and $B = \{x, y, z\}$. Then $A \times B = \{(1, x), (1, y), (1, z), (2, x), (2, y), (2, z)\}$. The Cartesian product contains all possible ordered pairs where the first element comes from A and the second from B. Note that $|A \times B| = |A| \cdot |B| = 2 \cdot 3 = 6$.

Definition 4.12 (Set Partitions)**.** A partition of a set A is a collection of disjoint subsets of A such that every element in A is included in exactly one subset. These subsets are called blocks of the partition.

Let $A = \{1, 2, 3, 4, 5, 6\}$. One partition of A could be $\{A_1, A_2, A_3\}$ where $A_1 = \{1, 2\}$, $A_2 = \{3, 5\}$, and $A_3 = \{4, 6\}$. These subsets are disjoint ($A_i \cap A_j = \emptyset$ for $i \neq j$) and their union gives the original set ($A_1 \cup A_2 \cup A_3 = A$).

Theorem 4.13 (Set Partition)**.** If $\{A_1, A_2, \ldots, A_n\}$ is a partition of set A, and $B \subseteq A$, then $\{A_1 \cap B, A_2 \cap B, \ldots, A_n \cap B\}$ is a partition of set B.

Let $A = \{1, 2, 3, 4, 5, 6\}$ with partition $\{A_1, A_2, A_3\}$ where $A_1 = \{1, 2\}$, $A_2 = \{3, 5\}$, and $A_3 = \{4, 6\}$. If $B = \{2, 3, 4, 5\} \subseteq A$, then $\{A_1 \cap B, A_2 \cap B, A_3 \cap B\} = \{\{2\}, \{3, 5\}, \{4\}\}$ forms a partition of B.

Definition 4.14 (Power Set)**.** The power set of A is the set of all subsets of A, denoted by $\mathcal{P}(A)$ is the set of all subsets of A.

Let $A = \{1, 2\}$. The power set $\mathcal{P}(A)$ contains all possible subsets of A: $\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$ Note that $|\mathcal{P}(A)| = 2^{|A|} = 2^2 = 4$. For any set with n elements, its power set has 2^n elements. For set operations in general, there are a few laws worth noting.

Identity Laws:

$$A \cup \varnothing = A$$

$$A \cap U = A.$$

Domination Laws:

$$A \cup U = U$$
$$A \cap \varnothing = \varnothing.$$

Idempotent Laws:

$$A \cup A = A$$
$$A \cap A = A.$$

Absorption Laws:

$$A \cup (A \cap B) = A$$
$$A \cap (A \cup B) = A.$$

Complement Laws:

$$A \cup \overline{A} = U$$
$$A \cap \overline{A} = \varnothing$$
$$\overline{\overline{A}} = A.$$

Commutative Laws:

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A.$$

Associative Laws:

$$A \cup (B \cup C) = (A \cup B) \cup C$$
$$A \cap (B \cap C) = (A \cap B) \cap C.$$

Distributive Laws:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

De Morgan's Laws:

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$
$$\overline{A \cup B} = \overline{A} \cap \overline{B}.$$

Definition 4.15 (Totally Ordered Set)**.** A totally ordered set is a set in which every pair of elements is comparable; for any two elements a and b, either $a \leq b$ or $b \leq a$ holds.

Examples include the real numbers and the integers with their usual orders.

Definition 4.16 (Partially Ordered Set (Poset))**.** A partially ordered set (poset) is a set equipped with a relation $\leq$ that is reflexive, antisymmetric, and transitive; in a poset, not every pair of elements must be comparable, so some pairs may be incomparable. For example, the power set $\mathcal{P}(\mathbb{Z})$ ordered by inclusion ($\subseteq$) contains subsets such as $\{1, 2\}$ and $\{2, 3\}$ that are not comparable, and the positive integers ordered by divisibility have incomparable primes like 5 and 7.

Definition 4.17 (Supremum)**.** Let S be a nonempty subset of a partially ordered set P. An element $u \in P$ is the supremum of S, denoted by $u = \sup S$, if:

1. **Upper bound:** every $s \in S$ satisfies $s \leq u$.
2. **Least upper bound:** if v is any upper bound of S, then $u \leq v$.

For instance, $\sup\{-1, -2, -3, \dots\} = -1$, and $\sup\{\sin x : x \in [0, \pi]\} = 1$.

Definition 4.18 (Infimum)**.** Let S be a nonempty subset of a partially ordered set P. An element $\ell \in P$ is the infimum of S, denoted by $\ell = \inf S$, if:

1. **Lower bound:** every $s \in S$ satisfies $\ell \leq s$.
2. **Greatest lower bound:** if v is any lower bound of S, then $v \leq \ell$.

For example, $\inf\{e^{-x} : x > 0\} = 0$, even though 0 is not an element of the set.

4.2 Groups

Groups, rings, and fields lay the foundation in mathmatics that build upon sets with additional operations in algebra. These strucutres are ubiquitous in mathmatics and physics and lay the foundation for quantum computing as well.

A group is a set equipped with a single binary operation that exhibits certain properties, much like addition and multiplication. A ring expands on this by incorporating two operations, typically referred to as multiplication and addition. A field is a more stringent structure where the set is a group under both operations, with multiplication also being commutative, and every non-zero element having a multiplicative inverse.

Definition 4.19. A group is a set G which is *closed* under an operation $\cdot$ (that is $\forall x, y \in G, x \cdot y \in G$) and satisfies the following properties:

1. **Identity:** $\exists e \in G$ where $\forall x \in G, x \cdot e = x = e \cdot x$. We define e to be the identity element.

2. **Inverse:** $\forall x \in G, \exists y \in G$ such that $x \cdot y = e = y \cdot x$, where e is the identity element identified above.

3. **Associativity:** The operation $\cdot$ is associative for every $x, y, z \in G$, i.e.

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z$$

In mathmatical contexts, it is common to omit the symbol $\cdot$ or $*$ for the group operation and simply write $x \cdot y$ as xy. Some common exmaples include the gorup of integers $\mathbb{Z}$ under addition, denoted by $(\mathbb{Z}, +)$, with $e = 0$, and the inverse of x being $-x$. Another example would be the group of integers modulo n, denoted $\mathbb{Z}/n\mathbb{Z}$, with closure (the sum of any two integers modulo n also forming a group under modulo n), identity, inverse, and associativity all holding under the subspace of $\mathbb{Z}/n\mathbb{Z}$.

Definition 4.20 (Abelian Group)**.** A group is said to be *abelian* if the operation $\cdot$ is commutative $\forall x, y \in G$, that is,

$$x \cdot y = y \cdot x$$

While the examples so fare are all abelian groups, there are some of groups that are not abelian. For exmaple, the set of all symmetries of an equilateral triangle, known as the dihedral group D_3 is not abelian. As you can see, the table of operations is not symmetrical, rendering the group as a non-abelian group. However, the identity, inverse, and associative elements/properties are all present upon verification.

	e	a	b	c	r	s
e	e	a	b	c	r	s
a	a	e	r	s	b	c
b	b	s	e	r	c	a
c	c	r	s	e	a	b
r	r	c	a	b	s	e
s	s	b	c	a	e	r

However, in the context of quantum computing, there are a few symmetry groups worth noting. They are

1. $\mathbf{SO}(N)$: The orthogonal group in N dimensions consists of all $N \times N$ orthogonal matrices with determinant 1, representing rotations in $\mathbb{R}^N$. These rotations preserver distance sand the orientation of objects in question. SO(2) corresponds to a circle, and SO(3) to a sphere.

2. $\mathbf{SU}(2)$: The special unitary group of degree 2 comprises of all 2×2 unitary matrices with determinant 1. It is closely related to SO(3) and is commonly used to describe spins and qubit states. Each rotation in SO(3) corresponds to two points in SU(2).

3. $\mathbf{SU}(N)$: This represents the special unitary group of degree N, extending the concepts of SU(2). These groups are useful in the study of quantum entanglement in quantum computing concerning N-level quantum systems.

Definition 4.21 (Subgroups)**.** A subgroup H of group G is a subset of G that is a group in by itself, with the same group operation in G. You can think of this as a reduced version of G where still

1. The identity element of G is in H.
2. $\forall h_1, h_2 \in H, \quad h_1 \cdot h_2 \in H$.
3. $\forall h \in H, \quad h^{-1} \in H$

Theorem 4.22 (Lagrange's Theorem)**.** For any finite group G and any subgroup H of G, the order of H divides the order of G, i.e.

$$|G| \equiv 0 \mod |H|$$

Definition 4.23 (Coset)**.** Given a group G and a subgroup H of G, the **left coset** of H with representative $g \in G$ is the set $gH = \{gh | h \in H\}$. Similarly, the right coset is $Hg = \{hg | h \in H\}$.

Theorem 4.24 (Parition Theorem for Cosets)**.** The collection of all left cosets of a subgroup H forms a partition of the group G. This means

1. Every element of G belongs to exactly one coset of H.
2. Cosets are disjoint and have no elements in common.

Definition 4.25 (Normal Subgroup)**.** A subgroup N of a group G is called a normal subgroup if it is invariant under conjugation by elements of G. This means that $\forall n \in N, \quad g \in G, \quad gng^{-1} \in N$. In notation, $N \triangleleft G$ if

$$gNg^{-1} = \{gng^{-1} | n \in N\} \subseteq N \forall g \in G.$$

Now, note that this is not the same as an abelian group. While the abelian group imples that every subgroup is a normal subgroup, the latter does not imply the former.

Example (Quaternion Group Q_8)**.** Let

$$Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$$

with relations

$$i^2 = j^2 = k^2 = ijk = -1.$$

This group is non-abelian since, for instance, $ij = k$ while $ji = -k$, so $ij \neq ji$. Nevertheless, every subgroup of Q_8 is normal; the subgroups are $\{1\}$, $\{\pm 1\}$, $\langle i \rangle = \{\pm 1, \pm i\}$, $\langle j \rangle = \{\pm 1, \pm j\}$, $\langle k \rangle = \{\pm 1, \pm k\}$, and Q_8 itself, and each is invariant under conjugation by any element of Q_8. Thus Q_8 is a non-abelian group in which all subgroups are normal (a Hamiltonian group).

Also, it is worth noting that the left and right cosets of a normal subgroup N are the same, allowing the group operations on cosets to be well-defined, which brings us to quotient groups:

Definition 4.26 (Quotient Group)**.** Let G be a group, and $N \triangleleft G$. The quotient group G/N is the set of cosets of $N \in G$ with the group operations defined by:

$$(gN)(hN)(gh)N \quad \forall g, h \in G$$

Example. Consider the group $\mathbb{Z}$ of integers under addition and the subgroup $2\mathbb{Z}$ consisting of all even integers. The quotient group $\mathbb{Z}/2\mathbb{Z}$ is the set of cosets of $2\mathbb{Z}$ in $\mathbb{Z}$:

- The coset $0 + 2\mathbb{Z} = \{\ldots, -4, -2, 0, 2, 4, \ldots\}$ represents the even integers.
- The coset $1 + 2\mathbb{Z} = \{\ldots, -3, -1, 1, 3, 5, \ldots\}$ represents the odd integers.

Thus, $\mathbb{Z}/2\mathbb{Z}$ has two elements: $0 + 2\mathbb{Z}$ and $1 + 2\mathbb{Z}$, corresponding to the even and odd integers, respectively. The group operation is addition modulo 2.

This quotient group is *isomorphic* to $\mathbb{Z}_2 = \{0, 1\}$ under addition modulo 2, denoted as $\mathbb{Z}/2\mathbb{Z} \cong \mathbb{Z}_2$. Two groups are isomorphic if they have the same structure, meaning there is a one-to-one correspondence between the elements of the two groups that preserves the group operation.

Definition 4.27 (Cyclic Group)**.** A group G is cyclic if $\exists g \in G$ such that $\forall x \in G$, x can be expressed as powers (repeated operations) of g.

A good example of this would be D_3, which also has multiple generators.

Definition 4.28 (Ring)**.** A ring is a set R equipped with two operations $+$ and $\times$ satisfying the following properties:

1. $(R, +)$ forms an abelian group.
2. $(R, \times)$ is associative, that is $\forall a, b, c \in R$,
$$a \times (b \times c) = (a \times b) \times c$$
3. The distributive properties hold, that is $\forall a, b, c \in R$,
$$a \times (b + c) = (a \times b) + (a \times c)$$
$$(b + c) \times a = (b \times a) + (c \times a)$$

For example, both the set of integers $\mathbb{Z}$ and the group of integers modulo n, $\mathbb{Z}/n\mathbb{Z}$, are both rings. However, multiplication notably does not have an inverse under most groups of integers. For this to hold, we need to have fields.

Definition 4.29 (Field)**.** A field is a set F with two operations $+$ and $\times$, where

1. $(R, +)$ forms an abelian group.
2. $(R - \{0\}, \times)$ forms an abelian group.
3. The distributive properties holds as in rings.

Some common examples include the set of rational numbers $\mathbb{Q}$, the set of real numbers $\mathbb{R}$, and the set of complex numbers $\mathbb{C}$. Interestingly, $\mathbb{Z}/p\mathbb{Z}$ where p is a prime is also a field.

4.3 Functions

Functions serve as mappings from one set to another. They assign each element in the domain to exactly one element in the codomain, and are essential for describing relationships using math.

Definition 4.30 (Function)**.** Let f be a function from set A to set B. *A function from A to B*, denoted $f : A \to B$, is an assignment of exactly one element of B to each element of A.

We write $f(a) = b$ to denote the assignment of b to an element a of A by the function f.

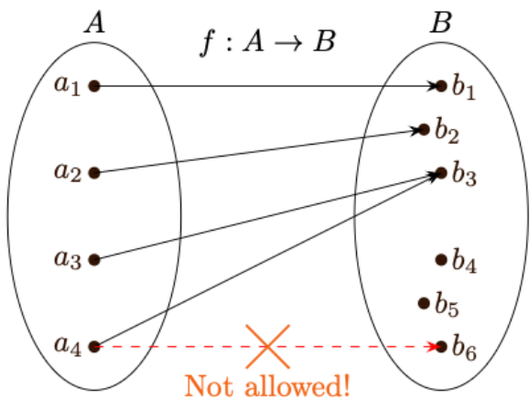
Example. $f(x) = x^2$ is a function $f : \mathbb{R} \to \mathbb{R}$. $\quad r = \sqrt{x^2 + y^2}$ is a function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

Definition 4.31 (Image, Range, and Domain)**.** Let f be a function from A to B.

- We say that A is the *domain* of f and B is the *codomain* of f.
- If $f(a) = b$, b is the *image* of a and a is a *pre-image* of b.
- The *range* of f (a subset of B) is the set of all images of elements of A.

- Let S be a subset of A. The *image of* S is a subset of B that consists of the images of the elements of S. We denote the image of S by $f(S)$, so that

$$f(S) = \{f(s) \mid s \in S\}$$



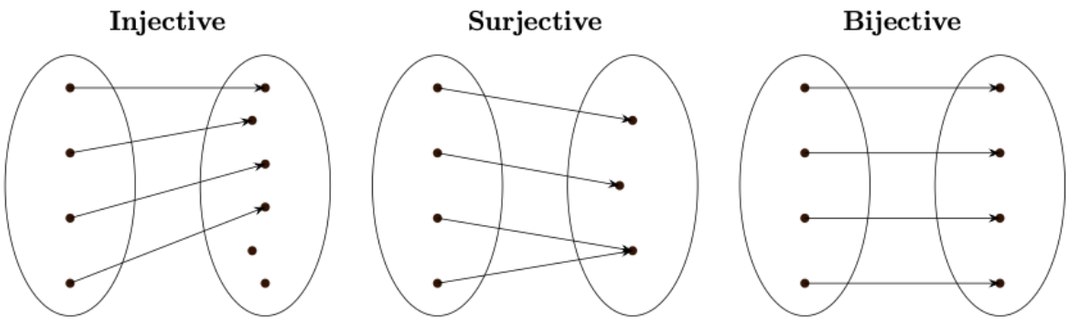
There are also a few common function types that we will define:

Definition 4.32 (Injective)**.** A function $f : A \to B$ is injective if $\forall a_1, a_2 \in A$, we have $f(a_1) = f(a_2) \implies a_1 = a_2$.

Definition 4.33 (Surjective)**.** A function $f : A \to B$ is surjective if the whole codomain is covered, meaning that $\forall b \in B, \quad \exists a \in A$ such that $f(a) = b$.

Sometimes we call surjective functions *onto* functions.

Definition 4.34 (Bijective)**.** A function is bijective if it is both *injective* and *surjective*.



Only with bijectivity established can we define the inverse of a function.

Definition 4.35 (Inverse Function)**.** If we take the function $f : A \to B$, it's domain A and codomain B, we can define an inverse function $f^{-1} : B \to A$ such that $\forall b \in B, \quad f^{-1}(b) = a$ if and only if $f(a) = b$.

Put plainly, $\forall a \in A, \quad f^{-1}(f(a)) = a$ and $\forall b \in B, \quad f(f^{-1}(b)) = b$.

4.4 Common Functions and Asymptotic Behavior

This section will be a brief review of real functions ($f : \mathbb{R} \to \mathbb{R}$) commonly used in quantum computing.

Power Functions

Power functions take the form $f(x) = x^p, \quad x \geq 0, p \in \mathbb{R}$.

The behavior of the function varies significantly with the exponent p:

- For $p > 0$, $f(x)$ increases as x increases. $f(x)$ exhibits a more rapid growth with a larger p.

- For $p < 0$, $f(x)$ decreases as x increases.
- When $p = 0$, $f(x) = 1$, regardless of x (excluding $x = 0$), which is a constant function.
- For $p = 1$, $f(x) = x$, representing a linear relationship.

Key properties of power functions include the rules for exponentiation:

- Multiplying powers with the same base: $x^a \cdot x^b = x^{a+b}$.
- Dividing powers with the same base: $x^a / x^b = x^{a-b}$.

Polynomial Functions

A polynomial function is a sum of terms $a_i x^i$, where i is a non-negative integer:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n.$$

Its behavior for large x values is predominantly determined by its highest power term, x^n, where n is the degree of the polynomial.

An n-th degree polynomial has n complex roots (counting multiplicities), and according to Vieta's formulas, the sum of these roots is equal to $-a_{n-1}/a_n$ and their product $(-1)^n a_0/a_n$.

Example. Consider the polynomial function $f(x) = x^3 - 7x^2 + 14x - 8$. It can be factored as:

$$f(x) = (x - 1)(x - 2)(x - 4).$$

The roots of this polynomial are $x = 1, x = 2$, and $x = 4$, which can be found by solving the equations $(x - 1) = 0, (x - 2) = 0$, and $(x - 4) = 0$. According to Vieta's formulas, the sum of the roots is:

$$1 + 2 + 4 = 7 = -\frac{-7}{1},$$

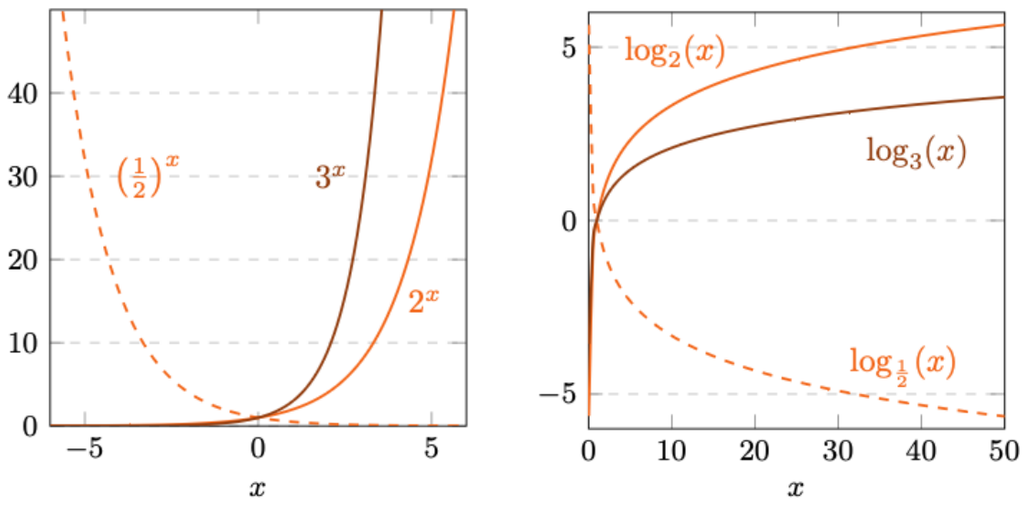
and the product is:

$$1 \cdot 2 \cdot 4 = 8 = (-1)^3 \frac{-8}{1}.$$

Exponential Functions

Exponential functions are defined as $f(x) = b^x$, where b is a positive constant (called the base) and $b \neq 1$. The variable x is the exponent. The key characteristic is that the variable is in the exponent. Some important notes:

- Growth and Decay:
  - If $b > 1$, $f(x)$ exhibits exponential growth – increasing rapidly as x increases. Larger bases lead to faster growth.
  - If $0 < b < 1$, $f(x)$ shows exponential decay – decreasing towards zero as x increases.
- Always Positive: Exponential functions are always positive for any real-valued input x.
- Horizontal Asymptote: They approach zero for one direction of x (negative infinity for growth, positive infinity for decay).
- Base e: The natural exponential function with $b = e$ (Euler's number, ≈ 2.718), i.e, e^x, also denoted as $\exp(x)$, has special significance across mathematics.



Logarithmic Functions

Logarithmic functions are the inverses of exponential functions. They are defined as $f(x) = \log_b(x)$, where b is a positive constant ($b \neq 1$) and $x > 0$. Some key points:

- Reversing Exponentiation: If $b^y = x$ then $\log_b(x) = y$.
- Growth and Behavior
  - For $b > 1$, $\log_b(x)$ increases as x increases, but very slowly.
  - For $0 < b < 1$, $\log_b(x)$ decreases as x increases.
- Vertical Asymptote: Logarithmic functions have a vertical asymptote at $x = 0$.
- Logarithms of 1 and the Base: $\log_b(1) = 0$ and $\log_b(b) = 1$.
- The natural logarithm, written as $\ln(x)$ has the base e.

Key Properties:

- The Product Rule: $\log_b(xy) = \log_b(x) + \log_b(y)$
- Logarithms "Break" Exponents: $\log_b(x^y) = y \cdot \log_b(x)$
- Changing Bases: $\log_b(x) = \log_a(x)/\log_a(b)$

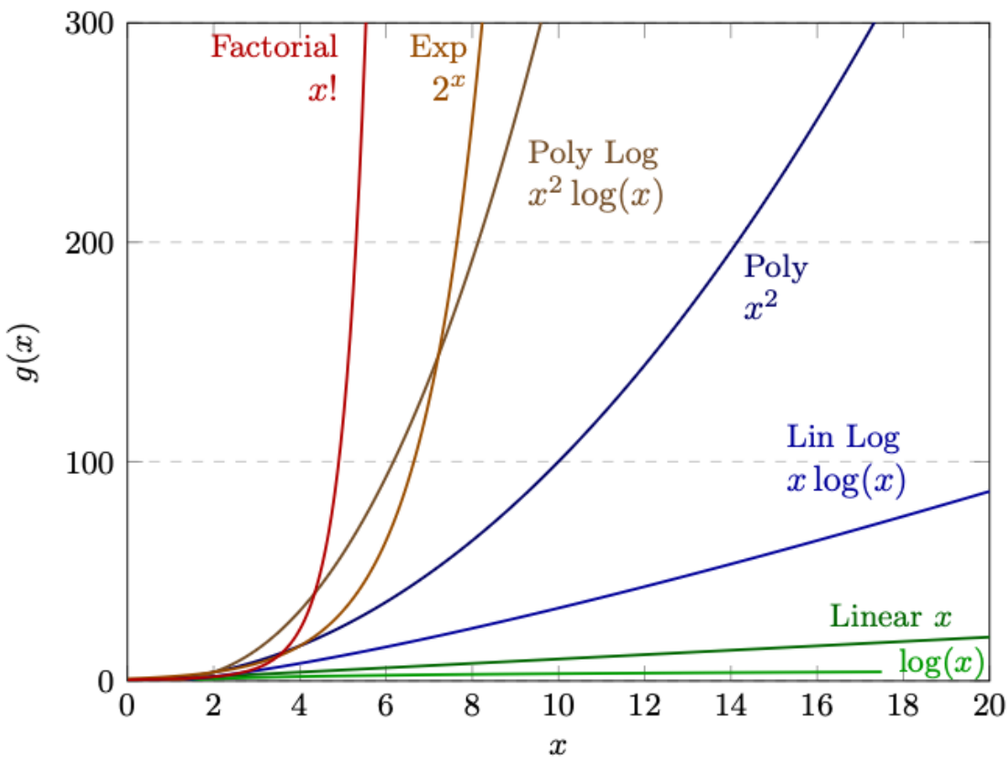
Scaling behavior, especially in the context of data structures and algorithm efficiency, is a significant topic for any computing related fields. It concerns itself with witht he evolution of function curves as the response becomes significantly large. We use the *Big O notation* to describe the behavior of a function $f(x), \quad x \to \infty$. We say that $f(x)$ is $O(g(x))$ meaning that there is some positive constant c such that the upper bound for complexity growth does not increase faster than $c \cdot g(x)$ for a sufficiently large x.

Example. For example, consider

$$f(x) = 6x^3 + 2x + 1,$$

we say that $f(x)$ is $O(x^3)$ as $x \to \infty$, as X^3 is the dominant term.

Most of the common limiting functions are illustrated in the following figure:



1. Log-log: $g(x) = \log\log(x)$
   - Exhibits extremely slow growth. Algorithms within this complexity class increase their running time at a negligible rate with input size escalation.
   - Applications include specialized computational geometry problems.
2. Log: $g(x) = \log(x)$
   - Denotes high efficiency. The execution time grows much slower than the input size.
   - Examples include binary search in sorted arrays and operations on certain balanced tree data structures.
3. Sublinear: $g(x) = x^p, 0 < p < 1$
   - Exhibits growth slower than linear but faster than logarithmic.
   - Common examples include the Grover's search algorithm in quantum computing, which has a complexity of approximately $O(\sqrt{x})$, and some algorithms that utilize probabilistic methods to achieve faster-than-linear performance on average.
4. Linear: $g(x) = x$
   - Indicates direct proportionality. Doubling the input size doubles the running time.
   - Common examples are searching in unsorted lists and identifying max/min elements in a list.
5. Polynomial: $g(x) = x^p, p > 1$
   - The growth rate is influenced by the exponent p. Higher values lead to rapid increases in running time with input size.
   - Examples: Bubble sort and insertion sort (quadratic complexity), matrix multiplication algorithms (cubic complexity or better).
6. Poly-log: $g(x) = x^p \log(x), p \geq 1$
   - Less efficient than the corresponding poly (or linear for $p = 1$) but still considered scalable.

- Fast Fourier Transform (FFT) algorithms are a prime example of algorithms with linear-log complexity. Some fast sorting algorithms also approach this performance.

7. Exponential: $g(x) = b^x, b > 1$

   - Characterized by rapid growth. Algorithms in this class quickly become impractical for moderate input sizes.
   - Examples: Brute-force approaches to the Traveling Salesman Problem. Currently known classical algorithms for integer factorization.

8. Factorial: $g(x) = x!$

   - Exhibits extremely rapid growth, surpassing even exponential functions in rate. Practical for only very small input sizes.
   - Example: Generating all permutations of a set.

9. Hyper-exponential: $g(x) = x^x, g(x) = b^{a^x}, g(x) = b^{x!}$, and $g(x) = b^{x^x}$, etc., where $a, b > 1$

   - Exhibits growth that is even more rapid than factorial functions.
   - Example: Modeling scenarios with extremely high growth rates, beyond combinatorial complexity.

5 Vectors and Vector Spaces

Vectors and the study of vector spaces are fundamental to the study of quantum computing, where they provide the mathmatical framework for representing and manipulating quantum states, commonly used to describe superpositions, entanglement, and other phenomena integral to quantum computing. This section serves as a systematic introduction to the realm of such concepets to lay the foundation for lies ahead.

5.1 Real Vectors and Complex Vectors

A vector is an ordered sequence of numbers. For now, we will focus on vectors with numeric elements.

Definition 5.1 (Ordered n-Tuple)**.** An ordered sequence of n numbers $(v_1, v_2, \ldots, v_n)$ is called an ordered n-tuple.

Definition 5.2 (Euclidean Space)**.** The set comprising all n-tuples is called n-space or Euclidean space, and denoted as $\mathbb{R}^n$ space. The complex Euclidean space is denoted by $\mathbb{C}^n$ and comprises of all complex n-tuples.

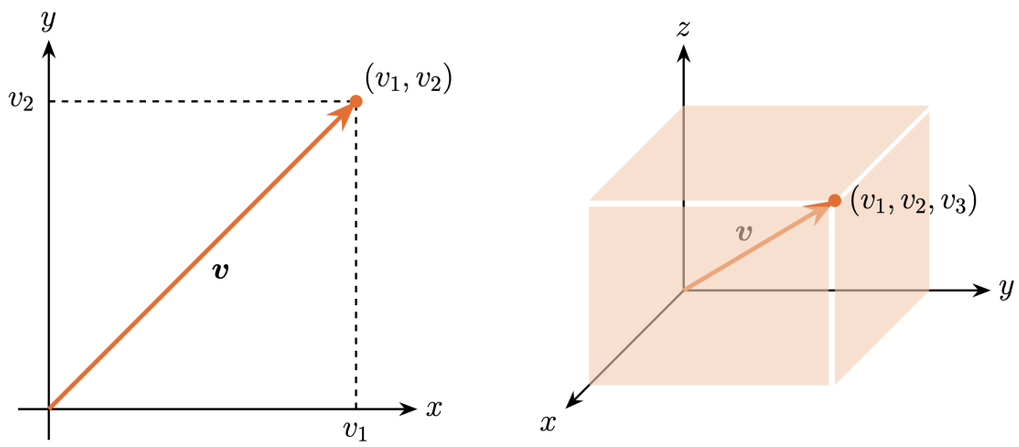
Now in the context of quantum computing, n is typically finite as we are usually considering finite quantum systems. A real vector is in $\mathbb{R}^n$, while a complex vector is in $\mathbb{C}^n$. Much like $\{\mathbf{0}\}$, we will begin with the defiition of the zero vector.

Definition 5.3 (Zero Vector)**.** The zero vector, in either $\mathbb{R}^n$ or $\mathbb{C}^n$, denoted by $\{\mathbf{0}\}$, is defined as the vector where all components are zero.

$$\{\mathbf{0}\} = (0, 0, \ldots, 0)$$

While it is difficult to visually represent vectors in higher dimensions, vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ are easier to represent. Generally, vectors both posess magnitude and direction, conviniently representing force, velocity, and heat flow. They carry weight and direction.

In linear algebra, vectors with identical length and direction are considered equivalent, unlike vector fields which also consider position, which means that we generally don't draw a distinction between *collinearity* and *parallelism*.



While seemingly simple, the intuition from representing vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ extends to $\mathbb{R}^n, \quad n > 3$ and even $\mathbb{C}^n$, for which geometric applications become exceedingly difficult or near impossible.

5.2 Basic Vector Algebra

A few basic operations on vectors are defined in this subsection in order to establish the foundation of vector manipulation.

Definition 5.4 (Vector Equality)**.** Given two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$, the two vectors are equal if and only if they are equal element wise. This is also denoted as $\mathbf{u} = \mathbf{v}$.

For each complex vector, there exists a unique complex conjugate that is defined as follows:

Definition 5.5 (Complex Vector Conjugate)**.** The complex conjugate of a complex vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ is denoted as $\mathbf{v}^*$ and is given by:

$$\mathbf{v}^* = ({v_1}^*, {v_2}^*, \ldots, {v_n}^*)$$

Where each element is taken to its complex conjugate.

Some other common operations also exist for vectors.

Definition 5.6 (Vector Sum)**.** The sum of two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$ is given by element-wise addition:

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$$

While not immediately obvious why this requires a proof, we have

Theorem 5.7. Vector addition satisfies the commutative, associative, and identity properties. Mathmatically,

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$

$$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$$

$$\mathbf{v} + \mathbf{0} = \mathbf{v}$$

Definition 5.8 (Vector Negation)**.** The negation of a vector $\mathbf{u}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$ is defined as the element wise negation of $\mathbf{u}$, denoted as $-\mathbf{u}$ and defined as:

$$-\mathbf{u} = (-v_1, -v_2, \ldots, -v_n)$$

Which allows us to define vector subtraction.

Definition 5.9 (Vector Subtraction)**.** Vector subtraction is defined as adding the negated version of the second vector, also defined as

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$$

Scalar multiplication is also very intuitive, defined as

Definition 5.10. Given $\mathbf{v}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$ and a scalar k, the scalar product is denoted by $k\mathbf{u}$, is defined as:

$$k\mathbf{v} = (kv_1, kv_2, \ldots, kv_n)$$

Note that the scalar in this case will usually be from the space in which the vector $\mathbf{v}$ resides. When we combine what we just defined, we have the linear combination.

Definition 5.11 (Linear Combination)**.** Consider a vector $\mathbf{w}$ and a set of r vectors $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}\}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$, we say that $\mathbf{w}$ is a linear combination of $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}\}$ if $\mathbf{w}$ can be expressed as:

$$\mathbf{w} = k_1\mathbf{v_1} + k_2\mathbf{v_2} + \ldots + k_r\mathbf{v_r} = \sum_{i=1}^{r} k_i\mathbf{v_i}$$

where each k_i is a scalar, referred to as the **coefficients** of the linear combination.

5.3 Vector Spaces, Subspaces, and Span

After laying the groundwork for basic vector operations, we're now focusing on vector spaces, which are a collection of vectors satisfying specific axioms. This forms a foundational concept in linear algebra, and provides a structured way to handle and manipulate dobjects in multiple dimensions.

Definition 5.12 (Vector Space)**.** A vector space (or linear space) is the combination of a certain set V combined with a scalar field F (either in $\mathbb{R}$ or $\mathbb{C}$) such that the following two operations are defined

1. *Vector Addition:* Vector addition in V remain in V.
2. *Scalar Multiplication:* Scalar multiplications remain in V for scalars from F.

These operations must satisfy the following properties:

- Addition is associative and commutative as defined above. Addition also contains an identity element as well as an inverse.
- Multiplication is distributive, both for the scalar and the vector.
- $(ab)\mathbf{v} = a(b\mathbf{v})$.
- The multiplicative identity is $1 \in F$.

We define item 1 and 2 from the definition as **closure under addition** and **closure under multiplication**, which are two important properties to satisfy for the whole spiel to work. Unless otherwise specified, discussions over the vector space $\mathbb{R}^n$ assume field $\mathbb{R}$, discussions over vector space $\mathbb{C}^n$ assume field $\mathbb{C}$. It is evident that $\mathbb{C}^n$ is also a vector space over $\mathbb{R}$, depending on the choice of the scalar field. in the context of quantum computing, we consistently assume that $\mathbb{C}^n$ is treated as a complex vector space.

While most of these definitions may be redundant, they are nevertheless necessary in order for later definitions.

Example. Let V be a vector space over the field F. Given a vector v in V and a scalar $k \in F$, use the previous axioms to prove that $k\mathbf{0} = \mathbf{0}$.

Proof. Using the distributive property for scalar multiplication, we have:

$$k(\mathbf{0} + \mathbf{0}) = k\mathbf{0} + k\mathbf{0}.$$

From the existence of the additive identity, we know $\mathbf{0} + \mathbf{0} = \mathbf{0}$, so:

$$k\mathbf{0} = k\mathbf{0} + k\mathbf{0}.$$

By the additive inverse property, there exists an additive inverse $-k\mathbf{0} \in V$. Adding $-k\mathbf{0}$ to both sides gives:

$$k\mathbf{0} + (-k\mathbf{0}) = (k\mathbf{0} + k\mathbf{0}) + (-k\mathbf{0}).$$

The left-hand side simplifies to $\mathbf{0}$, and using the fact that addition is commutative and contains an inverse on the right-hand side gives:

$$\mathbf{0} = k\mathbf{0} + [k\mathbf{0} + (-k\mathbf{0})] = k\mathbf{0} + \mathbf{0} = k\mathbf{0}.$$

This completes the proof. Note that the properties of associativity and identity in addition are also used. □

Next we will move on to subspaces. Subspaces of vector spaces are subsets that form a vector space of its own under the same vector operations. For example, the set of vectors in $\mathbb{R}^2$ with a positive x component violate the properties of a subspace as it is not closed when multiplied with a negative scalar. More rigorously

Definition 5.13 (Subspace)**.** A non-empty subset W of a vector space V is called a subspace of V if W is a vector space under the *same scalar field* F and the same operations of vector addition and scalar multiplication as in V.

While the properties listed out in the bullet points are easy to verify as those properties are inherently preserved when defining a subspace. In order to verify that a subset is a subspace, it suffices to verify whether the points 1. and 2. hold.

Theorem 5.14. If W is a non-empty subset of a vector space V, then W is a subspace of V if and only if W satisfies closure under addition and scalar multiplication.

Example. Prove that the empty set is a subspace of $\mathbb{R}^n/\mathbb{C}^n$.

Proof. We only need to verify that addition and scalar multiplication holds in the subspace. Since

$$\mathbf{0} + \mathbf{0} = \mathbf{0}, \qquad k\mathbf{0} = \mathbf{0}$$

Therefore, $\{\mathbf{0}\}$ is a subspace. □

Definition 5.15 (Zero Subspace)**.** The subset $W = \{\mathbf{0}\}$ is called the zero subspace of a vector space V where $\mathbf{0}$ is the zero vector in V.

An informal corrolary is that any hyperplane spanned by vectors that does not pass through the origin will not form a subspace.

Example. Show that $S = \{\mathbf{x} | \mathbf{x} = (z, z^*), z \in \mathbb{C}\}$ is not a subspace of $\mathbb{C}^2$

Proof. Since this is the space of the vectors spanned by a complex number z and its inverse z^*, we see that it conflicts with a scalar multiplication is defined as

$$k\mathbf{v} = (kz, kz^*).$$

The second element in the vector should be the complex conjugate of the first element. However, for every $\mathbf{v} \in S$ scaled by $k \in \mathbb{C}$, it becomes evident that the element is no longer in the set. Mathmatically:

$$(kz)^* = k^* z^* \neq kz^* \Rightarrow k\mathbf{v} \notin S.$$

Which violates closure under multiplication and proves that S is not a subspace of $\mathbb{C}^2$. □

While the previous definitions can be called foundational or elementary, we are interested to see how subspaces can be constructed using a set of vectors rather than from the top down, i.e., from the reduction of another subspace.

Theorem 5.16. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a set of vectors from vector space V, then all possible linear combinations of the vectors in S form a subspace of V.

Proof. Let W be the set of all possible linear combinations of the vectors in S and take any two vectors $\mathbf{a}, \mathbf{b} \in W$. Then we can express $\mathbf{a}$ and $\mathbf{b}$ as a linear combination of all vectors in S with coefficients $a_i, b_i \in F$. Then, for any scalar $k \in F$, we see that addition and scalar multiplication hold under S, making W a subspace of V. □

The formal definition also is as follows:

Definition 5.17 (Span)**.** Given a set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ from the vector space V, and let W be the subspoace of V that contains all possible linear combinations of S. Here, we define W as the span of S, and write

$$W = \text{span}(S) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}.$$

In other words, S **spans** W.

5.4 Linear Independence, Basis, and Dimension

We commonly use basis to describe the fundamental units on the cartesian coordiante and commonly use $\hat{\mathbf{i}} = (1, 0)$ and $\hat{\mathbf{j}} = (0, 1)$ in order to form a basis for $\mathbb{R}^2$. It is also easy to see that any vector in $\mathbb{R}^2$ can be expressed as a linear combination of these two vectors. However, it is not immediately mathmatically obvious why we must chose these to vectors. to form the basis of these vectors.

The first prerequisite for some combination of vectors to span a space, is for them to be linearly independent. This ensures that no vectors are redundant, which brings us to an associated property of subspaces

Definition 5.18 (Linear Dependence)**.** Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a set of vectors from a vector space V. The set S is said to be linearly dependent if there exists scalars $c_1, c_2, \ldots, c_n$ not all zero such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n = \mathbf{0}.$$

If not such scalars exists, such that $\forall c_i = 0, i \in [1, n]$, then the set is linearly independent.

A minimal set of vectors that spans a vector space V is a basis. Of course, there is also a rigorous definition.

Definition 5.19 (Basis)**.** A set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ from a vector space V forms the basis for V if and only if

- $\text{span}(S) = V$
- S is linearly independent

These two properties together construct the following theorem:

Theorem 5.20 (Uniqueness of Basis Representation)**.** If S is a basis of a vector space V, then every vector $\mathbf{w} \in V$ has a unique representation as a linear combination of S.

Proof. Assuming that $\text{span}(S) = V$ and take

$$\mathbf{w} = \sum_{i=1}^{n} k_i \mathbf{v}_i$$

and another representation of the same vector,

$$\mathbf{w} = \sum_{i=1}^{n} m_i \mathbf{v}_i$$

then we have

$$\mathbf{0} = \sum_{i=1}^{n} (k_i - m_i)\mathbf{v}_i$$

since S is linearly independent, we must have

$$k_i - m_i = 0, \quad \forall k_i, m_i \Rightarrow k_i = m_i$$

which renders the representation unique. □

From this, we can see that there is a well-defined and unique set of coefficients for every vector $\mathbf{w} \in V$ expressed as a linear combination of vector set S, where $\text{span}(S) = V$. There is also a special term that we assign to the coefficients:

Definition 5.21 (Coordinate)**.** Take the vector $\mathbf{w}$ expressed in the terms of a basis S for from a vector space V over a field $\mathbb{R}/\mathbb{C}$:

$$\mathbf{w} = \sum_{i=1}^{n} k_i \mathbf{v}_i$$

The unique vector $(k_1, k_2, \ldots, k_n)$ formed from the scalar coefficients in $\mathbb{R}^n/\mathbb{C}^n$ is said to be the coordinate vector, or the coordinate of $\mathbf{w}$ relative to S, denoted as

$$(\mathbf{w})_S = (k_1, k_2, \ldots, k_n)$$

Example. Any vector $(a, b) \in \mathbb{R}^2/\mathbb{C}^2$ in the standard basis $\hat{\mathbf{i}} = (1, 0)$ and $\hat{\mathbf{j}} = (0, 1)$ gives

$$(a, b) = a(1, 0) + b(0, 1)$$

which in turn gives us the coordinates for (a, b), that is

$$(a, b)_{\{\hat{\mathbf{i}},\ \hat{\mathbf{j}}\}} = (a, b)$$

This example goes to show that the coordinates in a Cartesian coordinate system uses the unit vectors along the x and y axis as its coordinate system. While this sentence sounds like a broken record, we can see that, while we prefer to use the unit vectors as the basis vectors to represent vectors in lower dimensional space, we can use alternative axes for representing the same points in n-dimensional space, and we'll introduce dimensions right off the bat with a theorem:

Theorem 5.22. For a finite-dimensional vector space, all bases posess the same number of vectors.

While the formal proof of this theorem requires extensive background in linear systems and matrix algebra, we nevertheless wish to develop an intuitive understanding. Consider a basis S for a finite-dimensional vector space V. If we add an extra vector to this, the set would be linearly dependent. Conversely, removing any vector would reduce the span and fail to cover all of V. However, this theorem confirms that the dimension of a vector space is well defined.

Definition 5.23 (Dimension)**.** The dimensions of a finite-dimensional vector space V is the number of vectors in its basis, denoted as $\dim(V)$.

6 Inner Product Spaces

What lays the foundation for representing n-qubit systems is being able to mathmatically represent them as state vectors residing in $\mathbb{C}^{2^n}$, which must be orthonormal basis – vectors mutually orthogonal of length 1. In order to rigorously define inner product spaces, we require a well-defined inner product – a mathmatical structure that generalizes length and orthogonality to higher dimensions. This, in turn, serves as the foundation for Hilbert spaces. We will also introduce dirac (or bra-ket) notation in this chapter which presents itself as an elegant method to represent quantum states and operations.

6.1 Dirac Notation Basics

Definition 6.1 (Matrix)**.** A matrix is a rectangular array of numbers, which can be either real or complex. A matrix with m rows and n columns is represented as:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m \times n}$$

where the entries $a_{11}, a_{12}, \ldots, a_{mn}$ are called **elements**. The matrix is referred to as an $m \times n$ matrix, indicating its size.

When a matrix contains only one column (or one row), it is called a column vector (or a row vector):

Definition 6.2 (Column and Row Vectors)**.** A matrix with only one column ($n \times 1$ in size) is called a column vector, denoted as:

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

Similarly, a matrix with only one row ($1 \times m$ in size) is called a row vector, denoted as:

$$\begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix}.$$

The process of matrix transposition, which interchanges the rows and columns of a matrix, is a fundamental operation in matrix algebra, defined as follows.

Definition 6.3 (Transpose)**.** Given an $m \times n$ matrix A as given in Eq. 6.1, the transpose of A, denoted as A^T, is an $n \times m$ matrix defined as:

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}_{n \times m}$$

resulting from interchanging the rows and columns of A.

From here on out, vectors such as $\mathbf{v}$ will represent column vectors and the transpose thereof will represent row vectors. In dirac notation, column vectors in $\mathbb{C}^n$ are usually represented by a symbol called ket.

Definition 6.4 (Ket)**.** Given a general vector $v = (v_1, v_2, \ldots, v_n)$ in $\mathbb{C}^n$, we use $|v\rangle$ to denote its column vector form:

$$|v\rangle \equiv \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

where $|v\rangle$ is referred to as a ket.

There are a few general and commmon methods of representing a state vector ket, which include:

- $|\psi\rangle$ for a general state vector;
- $|\lambda_i\rangle$ for eigenvectors corresponding to the eigenvalue λ_i;
- $|j\rangle, |k\rangle$ with $j, k \in \{0, 1, 2, \ldots, n-1\}$ to denote the computational basis vectors in $\mathbb{C}^n$.

Therea realso certain symbols reserved for specific vectors of significance, much like how we reserve π and e in common arithmatic. These include

- $|0\rangle, |1\rangle$ for the computational basis vectors in single qubit systems;
- $|V\rangle, |H\rangle$ for rectilinear polarization states of a photon;
- $|\Psi^+\rangle, |\Psi^-\rangle, |\Phi^+\rangle, |\Phi^-\rangle$ for Bell states of two qubits.

the standard basis vectors of $\mathbb{C}^n$ are indispensable in quantum computing, so it is necessary to define the computational basis for it

Definition 6.5 (Computational Basis)**.** For a single-qubit system in $\mathbb{C}^2$, the standard basis vectors, commonly referred to as the **computational basis**, are defined as the following ket vectors:

$$|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle \equiv \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

For a one-qudit system (with d distinct levels) in $\mathbb{C}^d$, the computational basis includes vectors:

$$|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad |1\rangle \equiv \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad |d-1\rangle \equiv \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Therefore, a general vector can be expressed as

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \alpha |0\rangle + \beta |1\rangle, \quad \forall \alpha, \beta \in \mathbb{C}$$

We conversely also have the hermitian adjoint of any vector $|\psi\rangle$, which is defined as

Definition 6.6 (Hermitian Adjoint)**.** Consider a column vector in $\mathbb{C}^n$:

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix},$$

its Hermitian adjoint, or simply *adjoint*, is defined as its conjugate transpose, denoted as:

$$v^\dagger \equiv (v^*)^T = (v^T)^* = \begin{bmatrix} v_1^* & v_2^* & \cdots & v_n^* \end{bmatrix}.$$

We have a convinient method of encapsulating the concept of adjoint vectors using a bra in dirac notation, defined as:

Definition 6.7 (Bra)**.** The bra, denoted as $\langle v|$, represents the Hermitian adjoint of $|v\rangle$, formally defined as:

$$\langle v| \equiv |v\rangle^\dagger.$$

Theorem 6.8. Given vectors $\langle u|$, $\langle v|$ in $\mathbb{C}^n$ and scalars $\alpha, \beta \in \mathbb{C}$, bras exhibit the following conjugate-linear properties:

$$\langle u + v| = \langle u| + \langle v| \tag{1}$$
$$\langle \alpha v| = \alpha^* \langle v| \tag{2}$$
$$\langle \alpha u + \beta v| = \alpha^* \langle u| + \beta^* \langle v|. \tag{3}$$

Note that $\langle \psi|$ and $|\psi\rangle$ are fundamentally not in the same dimension, so that operations on them will not be very straightforward.

6.2 Norm and Unit Vectors

We know that the squared distance of a line can be calculated by adding the squared values of their x and y componenets. In three dimensional space, it means adding the squared value of the z component as well. This allows us to generalize to the nth dimension, where we have the following, rather intuitive, definition for norm.

Definition 6.9 (Norm of a Real Vector)**.** For any real vector $\mathbf{v} \in \mathbb{R}^n$, its length is denoted $\|\mathbf{v}\|$ and defined by:

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

Theorem 6.10. The norm of a vector in $\mathbb{R}^n/\mathbb{C}^n$ exhibits the following properties:

1. $\|\mathbf{v}\| \geq 0$
2. $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = 0$
3. $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$

For complex numbers, we just need to use the complex conjugate into the equation.

Definition 6.11 (Norm of a Complex Number)**.** For any complex vector $\mathbf{v} \in \mathbb{C}^n$, the norm is defined by the formula:

$$\|\mathbf{v}\| = \sqrt{v_1 {v_1}^* + v_2 {v_2}^* + \cdots + v_n {v_n}^*}$$

We can also derive the following theorem:

Theorem 6.12. For a complex number $\mathbf{v} \in \mathbb{C}$, the following identity holds:

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

Proof. Since we know that $|z| = \sqrt{z^* z}$, we can square both sides of the equation to give us:

$$\|\mathbf{v}\|^2 = z_1^* z_1 + z_2^* z_2 + \cdots + z_n^* z_n = |v_1|^2 + |v_2|^2 + \cdots + |v_n|^2$$

□

On a side note, we have the following equivalencies for a vector in dirac notation:

$$\|v\| \equiv \||v\rangle\| = \|\langle v|\|$$

In quantum mechanics, most of the vectors must be represented as complex vectors of norm 1.

Theorem 6.13. For any non-zero vector $\mathbf{v} \in \mathbb{R}$, the following formula will give the normalized vector in the direction of $\mathbf{v}$, denoted by $\hat{\mathbf{v}}$:

$$\hat{\mathbf{v}} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$$

Proof.

$$\|\hat{\mathbf{v}}\| = \left\|\frac{1}{\|\mathbf{v}\|}\mathbf{v}\right\| = \left|\frac{1}{\|\mathbf{v}\|}\right|\|\mathbf{v}\| = \frac{1}{\|\mathbf{v}\|}\|\mathbf{v}\| = 1$$

□

You might realize by this point, that the unit vectors reside on some version of a unit hypersphere. In 2 dimensions, this would be the unit circle. Consequently, rotations in hyperspace keep the unit vector as a unit vector. Distance is defined as:

Definition 6.14 (Distance)**.** Given two vectors $\mathbf{v}$ and $\mathbf{u}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$, the distance between $\mathbf{v}$ and $\mathbf{u}$ is defined as:

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

6.3 Complex Inner Product Spaces

A vector space will only become an inner product space when it contains the inner product as a valid operation. In $\mathbb{R}^n$, this operation is simply the dot product. Generally speaking, the inner product is the operation defined for a vector space V, scalar field F, as

$$\langle \cdot, \cdot \rangle : V \times V \to F$$

The more familiar form of the inner product in $\mathbb{R}^n$, defined as:

Definition 6.15 (Real Dot Product)**.** For vectors $\mathbf{v}$ and $\mathbf{u}$ in $\mathbb{R}^n$, their dot product is deinfed as:

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

Which is also known as the Euclidian inner product.

The norm of a real vector and the distance between two vectors can also be expressed as a dot product as:

Theorem 6.16. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we have

$$\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$$
$$\|\mathbf{u} - \mathbf{v}\|^2 = (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})$$

Proof. Write the equation out and group it together using the definition of the real dot product. □

At the same time, we have a few more intuitive and convinient algebraic properties of the dot product.

Theorem 6.17. For $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $k \in \mathbb{R}$, we have the following algebraic identities:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \qquad \text{(Commutative Law)}$$
$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} \qquad \text{(Left Distributive Law)}$$
$$(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} \qquad \text{(Right Distributive Law)}$$
$$(k\mathbf{u}) \cdot \mathbf{v} = k(\mathbf{u} \cdot \mathbf{v}) \qquad \text{(Left Homogeneity Law)}$$
$$\mathbf{u} \cdot (k\mathbf{v}) = k(\mathbf{u} \cdot \mathbf{v}) \qquad \text{(Right Homogeneity Law)}$$

In the complex plane, things get more interesting.

Definition 6.18 (Complex Dot Product)**.** Given $\mathbf{v}, \mathbf{u} \in \mathbb{C}^n$, we have the complex dot product:

$$\mathbf{u} \cdot \mathbf{v} = u_1^* v_1 + u_2^* v_2 + \cdots + u_n^* v_n = \sum_n^{i=1} u_i^* v_i$$

Generally speaking, the complex dot product will produce a complex scalar. In some other contexts, one might find the definition of the complex dot product to have the second element be the conjugate element. However, the previous definition is in line with the conventions of quantum mechanics and computing. The use of the complex conjugate is also designed to ensure that the norm defined previously is the dot product of a complex number with itself.

Theorem 6.19. Given a complex number $\mathbf{v}$, it's norm can be expressed as:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} \quad \iff \quad \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$$

Complex inner product spaces are complex vector spaces "equipped" with a complex inner prodcut. You can think of them as spaces where the dot product (a resulting scalar from two vector multiplications) is a valid operation. We like using $\mathbb{R}^n/\mathbb{C}^n$ as examples of common inner product spaces, and in quantum computing we really like $\mathbb{C}^n$ as it is the norm for representing quantum systems.

Definition 6.20 (Complex Inner Product Space)**.** A complex inner product space is a complex vector space V equipped with an inner product, defined as: $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$. The inner product, for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V, \quad k \in \mathbb{C}$, must satisfy:

1. $\langle \mathbf{u} \cdot \mathbf{v} \rangle = \langle \mathbf{v} \cdot \mathbf{u} \rangle^*$
2. $\langle \mathbf{u} \cdot \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u} \cdot \mathbf{v} \rangle + \langle \mathbf{u} \cdot \mathbf{w} \rangle$
3. $\langle \mathbf{u} \cdot k\mathbf{v} \rangle = k \langle \mathbf{u} \cdot \mathbf{v} \rangle$
4. $\langle \mathbf{v} \cdot \mathbf{v} \rangle \in \mathbb{R}$
5. $\langle \mathbf{v} \cdot \mathbf{v} \rangle = 0 \quad \iff \quad \mathbf{v} = 0$

If we put the dot product and the complex vector space together, we have:

Theorem 6.21. $\mathbb{C}^n$ is a complex inner product space when the complex dot product is used as the inner product.

Proof. Take $V \equiv \mathbb{C}^n$. In order to prove that the dot product qualifies as an inner product space, we need to make sure that it satisfies the properties laid out in the previous definition.

1. $$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} {u_i}^* v_i = \sum_{i=1}^{n} (u_i {v_i}^*)^* = \left(\sum_{i=1}^{n} u_i {v_i}^* \right)^* = (\mathbf{v} \cdot \mathbf{u})^*$$
2. $$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \sum_{i=1}^{n} {u_i}^* (v_i + w_i) = \sum_{i=1}^{n} {u_i}^* v_i + \sum_{i=1}^{n} {u_i}^* w_i = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$$
3. $$\mathbf{u} \cdot (k\mathbf{v}) = \sum_{i=1}^{n} {u_i}^* (k v_i) = k \sum_{i=1}^{n} {u_i}^* v_i = k(\mathbf{u} \cdot \mathbf{v})$$
4. $$\mathbf{v} \cdot \mathbf{v} = \sum_{i=1}^{n} {v_i}^* v_i = \sum_{i=1}^{n} |v_i|^2 \geq 0, \qquad \forall \mathbf{v}$$
5. $$\forall i \in [1, n], \qquad |v_i| = 0 \quad \Rightarrow \quad v_i = 0 \quad \Rightarrow \quad \mathbf{v} = \mathbf{0}$$

□

In addition to the previously defined identities, we also have:

Theorem 6.22. Given a complex inner product space V with an inner product $\langle \cdot \cdot \cdot \rangle$, with $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V, \quad k \in \mathbb{C}$, the following identities hold:

$$\begin{aligned} \langle \mathbf{0} \cdot \mathbf{v} \rangle &= \langle \mathbf{v} \cdot \mathbf{0} \rangle = 0 \\ \langle \mathbf{u} + \mathbf{v} \cdot \mathbf{w} \rangle &= \langle \mathbf{u} \cdot \mathbf{w} \rangle + \langle \mathbf{v} \cdot \mathbf{w} \rangle \\ \langle k\mathbf{u} \cdot \mathbf{v} \rangle &= k^* \langle \mathbf{u} \cdot \mathbf{v} \rangle \end{aligned}$$

Proof. Considering that the conjugate is taken for the left-handed element in the inner product:

$$\begin{aligned} \langle k\mathbf{u} \cdot \mathbf{v} \rangle &= \langle \mathbf{v} \cdot k\mathbf{u} \rangle^* \\ &= (k \langle \mathbf{v} \cdot \mathbf{u} \rangle)^* \\ &= k^* \langle \mathbf{v} \cdot \mathbf{u} \rangle^* \\ &= k^* \langle \mathbf{u} \cdot \mathbf{v} \rangle \end{aligned}$$

□

Now as a subset of complex inner product spaces, we have the real inner product space.

Definition 6.23 (Real Inner Product Space)**.** A real inner product space is a vector space V equipped with the binary function (inner product) $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$. It must, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V, \quad k \in \mathbb{R}$, satisfy:

1. $\langle \mathbf{u} \cdot \mathbf{v} \rangle = \langle \mathbf{v} \cdot \mathbf{u} \rangle$

2. $\langle \mathbf{u} \cdot \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u} \cdot \mathbf{v} \rangle + \langle \mathbf{u} \cdot \mathbf{w} \rangle$
3. $\langle \mathbf{u} \cdot k\mathbf{v} \rangle = k \langle \mathbf{u} \cdot \mathbf{v} \rangle$
4. $\langle \mathbf{v} \cdot \mathbf{v} \rangle \in \mathbb{R}$
5. $\langle \mathbf{v} \cdot \mathbf{v} \rangle = 0 \iff \mathbf{v} = 0$

The only key difference lies in the first property. We also have the following theorem for the real inner product:

Theorem 6.24. The following identities hold for the real inner product space:

$$\langle \mathbf{u} \cdot k\mathbf{v} + m\mathbf{w} \rangle = k \langle \mathbf{u} \cdot \mathbf{v} \rangle + m \langle \mathbf{u} \cdot \mathbf{w} \rangle$$
$$\langle k\mathbf{u} + m\mathbf{v} \cdot \mathbf{w} \rangle = k \langle \mathbf{u} \cdot \mathbf{w} \rangle + m \langle \mathbf{v} \cdot \mathbf{w} \rangle$$

Which, conviniently, leads us to the definition for norm and distance using the inner product.

Definition 6.25 (Norm and Distance)**.** For any $\mathbf{v}, \mathbf{u} \in V$, the length (norm) is defined by

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v} \cdot \mathbf{v} \rangle}$$

The distance between two vectors is defined by:

$$d(\mathbf{v}, \mathbf{u}) = \|\mathbf{v} - \mathbf{u}\| = \sqrt{\langle \mathbf{v} - \mathbf{u} \cdot \mathbf{v} - \mathbf{u} \rangle}$$

Using dirac notation, representing inner products become much simpler and straightforward.

Definition 6.26 (Product of Row and Column Vectors)**.** Given a row vector $\mathbf{r}$ and a column vector $\mathbf{c}$ in $\mathbb{C}^n$:

$$\mathbf{r} = \begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}$$

their matrix product is defined as:

$$\mathbf{rc} = \sum_{i=1}^{n} r_i c_i$$

Now, noting that we can simplify the product betweenn a vector $|\phi\rangle$ and a hermitian adjoint ($\psi^\dagger = \langle\psi|$) as $\langle\psi| \, |\phi\rangle \equiv \langle\psi|\phi\rangle$, we have the following definition:

Definition 6.27 (Dirac Notation of Inner Product)**.** Given two vectors $|u\rangle = \mathbf{u}, |v\rangle = \mathbf{v}$, we have the inner product defined as:

$$\langle u|v\rangle \equiv \langle \mathbf{u} \cdot \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}.$$

For convinience, we have the following theorem:

Theorem 6.28. For vectors $|u\rangle, |v\rangle, |w\rangle \in \mathbb{C}^n, \alpha \in \mathbb{C}$, the following identities hold:

$$\langle u|v\rangle = \langle v|u\rangle^*$$
$$\langle u + v|w\rangle = \langle u|w\rangle + \langle v|w\rangle$$
$$\langle u|v + w\rangle = \langle u|v\rangle + \langle u|w\rangle$$
$$\langle u|\alpha v\rangle = \alpha \langle u|v\rangle$$
$$\langle \alpha u|v\rangle = \alpha^* \langle u|v\rangle$$
$$\|\mathbf{v}\|^2 = \langle v|v\rangle$$

We also have the following theorem for any unit vector:

Theorem 6.29. Any unit vector $|\phi\rangle$ satisfies:

$$\langle\phi|\phi\rangle = 1$$

Proof. Since $|\phi\rangle$ is a unit vector, $\|\phi\| = 1$, which means that $\langle\phi|\phi\rangle = \|\phi\|^2 = 1$ □

There are a host of computational examples that can be found in the textbook and it is highly recommended to look over them and compute for yourself. Now it is worth mentioning a quick note on **Hilbert Spaces**, which is the mathmatical framework under which quantum states reside. While such spaces are often infinite-dimensional, the complex inner product spaces $\mathbb{C}^n$ are often special cases of finite-dimensional Hilbert spaces.

Infinite-Dimensional Complex Vector Spaces $\mathbb{C}^\infty$

We first start with $\mathbb{C}^\infty$ where columns consist of infinite sequences of complex numbers. This forms the foundation for infinite dimensional Hilbert spaces. An inner product here is defined as:

$$\langle \mathbf{u} \cdot \mathbf{v} \rangle = \sum_{i=1}^{\infty} {u_i}^* v_i$$

Notably, for the inner product to be well defined, the series must converge, which leads to the concept of square-summable sequences (l^2).

Functional Inner Spaces

Functional Spaces, which are defined as the spaces of functions, consist of functions with specific properties. If we equip them with an inner product, these functional spaces become (you guessed it) functional inner spaces. For the inner product of functions $f(x), g(x)$ over the interval $[a, b]$, we have:

$$\langle f \cdot g \rangle = \int_a^b f^*(x) g(x)\, dx.$$

One of the most important examples of functional inner spaces is $L^2([a, b])$, is the space of all square-intergrable functions on $[a, b]$, defined as:

$$L^2([a,b]) = \left\{ f : [a,b] \longrightarrow \mathbb{C} | \int_a^b |f(x)|^2\, dx < \infty \right\}$$

The corresponding norm, (unironically) called the L^2 norm, is defined as:

$$\|f\| = \sqrt{\int_a^b |f(x)|^2\, dx}$$

One will see later how concepts like orthogonality and projection naturally extends to functions. In the textbook, there is a brief tangent that connects this to the fourier series $\phi_k(x)$, which will not be covered at this time.

Hilbert Space

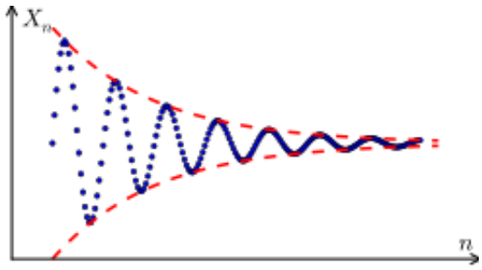
Building on the previous concepts of $\mathbb{C}^n$ and functional spaces, we are ready to define Hilber spaces. These are generalization of inner product spaces that is *complete*.

Definition 6.30 (Hilbert Space)**.** A Hilber space is an **inner product space** that is *complete* with respect to the norm induced by its inner product. Completeness means that every Cauchy sequence in the space converges to a point within the space.

So what is a cauchy sequence? It is what Augustin Cauchy proposed as the solution to the question: how do you decide when to consider a sequence as convergent?

Definition 6.31 (Cauchy Sequence)**.** A $\{s_n\} = (s_1, s_2, \ldots)$ is called a **Cauchy Sequence** if:

$$\forall N \in \mathbb{N}, \quad \epsilon > 0, \quad n > N \qquad \Longrightarrow \qquad |s_n - s_{n+1}| < \epsilon$$



In relation to the Hilbert space, it means that every sequence in the Hilbert space will converge to a point in the Hilbert space, that it is complete. In the context of quantum computing and representing information, the n-qubit state space is a finite system constructed from an inner product space with dimension 2^n. This is formalized by the following theorem:

Theorem 6.32. Every finite-dimensional real or complex inner product space is a Hilbert space.

In quantum mechanics, Hilbert spaces will often extend to infinite dimensions. It ensures that any linear combination or superposition remains a valid quantum state, providing consistency to the space. In other words, this space is **complete**.

6.4 Orthogonality and Projection

We define orthogonality using the inner product, where, in close conjunction with the norm and establishing orthonormal bases, is a key concept. It entertains a close relationship with the projection problem, playing a crucial role in determining the measurement probabilities for different outcome states. From a geometric standpoint, the angle between two non-zero vectors in $\mathbb{R}^2/\mathbb{R}^3$ is defined as the intuitive angle between the two vectors.

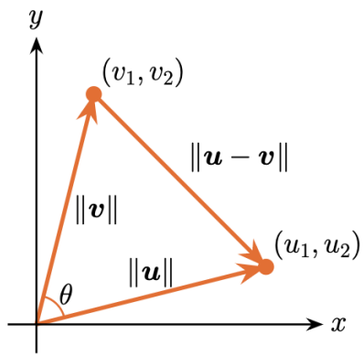
When we talk about the dot product, it is closely related to the two. Applying the law of cosines:

$$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$$

and rearranging this equation gives us:

$$\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta = \frac{1}{2}(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2)$$

We can see that the right hand side is equal to the dot product of $\mathbf{u}$ and $\mathbf{v}$, bringing us to the following theorem and definition.



Theorem 6.33. For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2/\mathbb{R}^3$, we have the following identity:

$$\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$$

where θ is the geometric angle between $\mathbf{u}$ and $\mathbf{v}$.

In fact, while we say that this is true for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2/\mathbb{R}^3$, it is only because we mention that this is for the geometric angle θ. It is very difficult to visualize θ in hyperspace, yet the following general definition holds for non-zero vectors in $\mathbb{R}^n$.

Definition 6.34 (Angle Between Two Real Vectors)**.** For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the angle is said to lie within the interval $[0, \pi]$ and is given by:

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$$

Note that for the previous definition to hold, the computed cosine value must lie within the interval of $[-1, 1]$. This condition is guaranteed to be satisfied given the **Cauchy-Schwarz Inequality**, stating that:

$$-\|\mathbf{u}\|\|\mathbf{v}\| \leq \mathbf{v} \cdot \mathbf{u} \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

For complex numbers, we say that $\mathbf{u} \cdot \mathbf{v}$ is generally a complex number. As a result, the previous definition for $\cos\theta$ no longer holds since the right side is generally complex. However, the condition $\mathbf{v} \cdot \mathbf{u} = 0$ still holds regardless and remains of great use even for complex vectors.

Definition 6.35 (Orthogonality in $\mathbb{C}^n$)**.** Given two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$, we say that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $\mathbf{v} \cdot \mathbf{u} = \mathbf{u} \cdot \mathbf{v} = 0$.

Definition 6.36 (Orthogonality in Inner Product Spaces)**.** Generally, for $\mathbf{u}, \mathbf{v} \in V$, where V is an inner product space, we say that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal when $\langle \mathbf{u} \cdot \mathbf{v} \rangle = \langle \mathbf{v} \cdot \mathbf{u} \rangle = 0$

Definition 6.37 (Generalized Pythagorean Theorem)**.** For $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$, if $\mathbf{u}$ and $\mathbf{v}$ are orthognal, then

$$\|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2$$

Proof. Since we know that orthogonal vectors $\mathbf{u}, \mathbf{v}$ have the property $\langle \mathbf{u} \cdot \mathbf{v} \rangle = 0$, then:

$$\langle \mathbf{u} \cdot \mathbf{v} \rangle = (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \|\mathbf{u}\|^2 + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \|\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

□

There is an intriguing property of orthogonal vectors that can be illustrated with the following example:

Example. Given $\mathbf{u} = (1, 0)$, find a unit vector $\mathbf{v} \in \mathbb{C}^2$ that is orthogonal to $\mathbf{u}$. If we take $\mathbf{v} = (v_1, v_2)$, we know that $\langle \mathbf{u} \cdot \mathbf{v} \rangle = 0$, which means that $v_1 = 0$, and it will not matter what v_2 is equal to, as long as it is of length one. Therfore, we have

$$\mathbf{v} = (0, e^{i\phi}), \quad \phi \in \mathbb{R}$$

This example illustrates a simple concept about complex vectors, whcih is that there exists an infinite number of unit vectors that are orthognoal to $(1, 0)$ in $\mathbb{C}^2$. This comes from the fact that multiplying, or scaling, by a factor $i\phi$ only maintains the magnitude of a complex number z. Therefore, we have the the following theorem:

Theorem 6.38. Given two unit vectors $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$ that are orthogonal, there are an infinitely many unit vecros in the span of $\mathbf{w}$ which are ortogonal to $\mathbf{z}$ in the form of $e^{i\phi}\mathbf{w}, \phi \in \mathbb{R}$.

Proof. We rewrite $\mathbf{w}' = e^{i\phi}\mathbf{w}, \phi \in \mathbb{R}$ as $\mathbf{w}' = \lambda\mathbf{w}, \lambda \in \mathbb{C}$. The orthogonality operation then becomes:

$$\mathbf{z} \cdot \mathbf{w}' = \mathbf{z} \cdot (\lambda\mathbf{w}) = \lambda(\mathbf{z} \cdot \mathbf{w}) = \lambda\mathbf{0} = \mathbf{0}$$

Furthermore, since $\|\mathbf{w}'\| = 1$, we see that:

$$\|\mathbf{w}'\| = \|\lambda\mathbf{w}\| = |\lambda| \cdot \|\mathbf{w}\| = |\lambda|$$

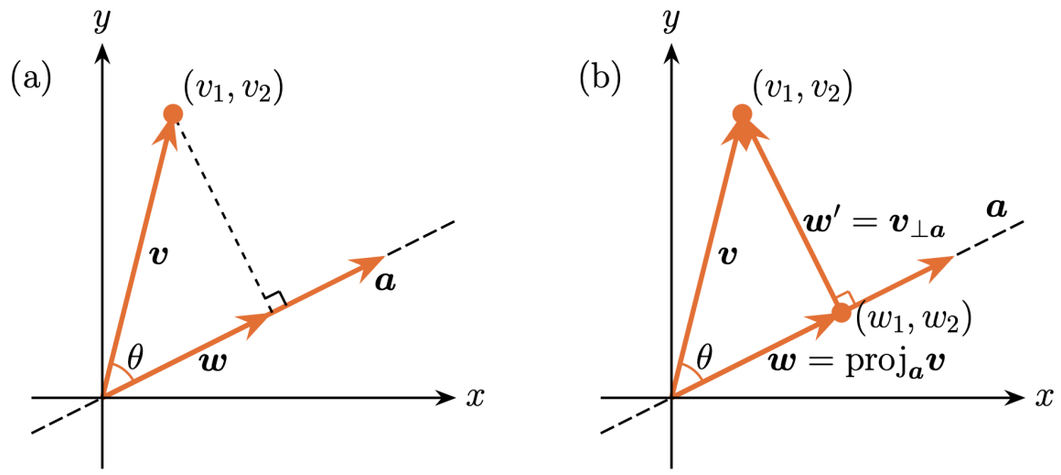
Implying that λ is of the form $e^{i\phi}$. □

Since quantum states are represented by unit vectors in $\mathbb{C}^n$, we can see that multiplying any unit vector by a factor of $e^{i\phi}$ preserves its modulus, hence preserving it as a unit vector. We therefore call these vectors **global phase factors**, which highlights its effect across the quantum state.

Theorem 6.39. Given two orthogonal vectors $|\phi\rangle, |\psi\rangle \in \mathbb{C}^n$, they satisfy:

$$\langle\phi|\psi\rangle = \langle\psi|\phi\rangle = 0$$

Orthogonaliy is closely related to the concept of projections in the sense that it is commonly known as an orthogonal projection. An example of this in $\mathbb{R}^2$ is illustrated by the dashed line in the figure below.



Once you understand this conceptually in two dimensions, it becomes easier to see how $\mathbf{a}$ can be extended to a plane in three dimensions, and so on, bringing us to the general projection theorem in $\mathbb{R}^n/\mathbb{C}^n$.

Theorem 6.40 (Projection Theorem)**.** Given $\mathbf{v}, \mathbf{a} \in \mathbb{R}^n/\mathbb{C}^n$, there is an *unique* way to decompose $\mathbf{v}$ as $\mathbf{v} = \mathbf{w} + \mathbf{w}'$ where $\mathbf{w}$ is in the span of $\mathbf{a}$ and $\mathbf{w}'$ is orthogonal to $\mathbf{a}$.

Here, we denote $\mathbf{w}$ as the **orthogonal projection of v onto a** and denote it as $\text{proj}_{\mathbf{a}}\mathbf{v}$ while $\mathbf{w}'$ as the **complement of v orthogonal to a**, denoted as $\mathbf{v}_{\perp \mathbf{a}}$. They are given by the following formulas:

$$\text{proj}_{\mathbf{a}}\mathbf{v} = \mathbf{v}_{\parallel \mathbf{a}} = \frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{a}\|^2}\mathbf{a}$$
$$\mathbf{v}_{\perp \mathbf{a}} = \mathbf{v} - \text{proj}_{\mathbf{a}}\mathbf{v}$$

Proof. Since $\mathbf{w}$ is in the span of the vector $\mathbf{a}$, it can be writen as a scaled version of $\mathbf{a}$, i.e. $\mathbf{w} = \lambda\mathbf{a}, \lambda \in \mathbb{C}$. It is thus possible to expand the equation into:

$$\mathbf{a} \cdot \mathbf{v} = \mathbf{a} \cdot (\mathbf{w} + \mathbf{w}') = \mathbf{a} \cdot \mathbf{w} + \mathbf{a} \cdot \mathbf{w}' = \mathbf{a} \cdot \mathbf{w} = \mathbf{a} \cdot (\lambda\mathbf{a}) = \lambda(\mathbf{a} \cdot \mathbf{a})$$

Solving for λ gives us:

$$\lambda = \frac{\mathbf{a} \cdot \mathbf{v}}{\mathbf{a} \cdot \mathbf{a}} = \frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{a}\|^2}$$

We therefore have:

$$\mathbf{w} = \lambda\mathbf{a} = \frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{a}\|^2}\mathbf{a}, \qquad \mathbf{w}' = \mathbf{v} - \mathbf{w} = \mathbf{v} - \frac{\mathbf{a} \cdot \mathbf{v}}{\|\mathbf{a}\|^2}\mathbf{a}$$

□

Since in quantum computing the state vectors are unit vectors, the projection between unit complex vectors are of vital importance. We thus have:

Theorem 6.41 (Projection Theorem Between Complex Unit Vectors)**.** Consider $|\phi\rangle, |\psi\rangle \in \mathbb{C}^n$. The projection of $|\phi\rangle$ onto $|\psi\rangle$ is given by the formula:

$$\text{proj}_{|\psi\rangle} |\phi\rangle = \langle\psi|\phi\rangle |\psi\rangle$$

where $|\langle\psi|\phi\rangle| \leq 1$.

Now in order to prove this, we need an adaptation of the Cauchy-Schwarz Inequality for Dirac notation

$$|\langle\phi|\psi\rangle|^2 \leq \langle\phi|\psi\rangle \cdot \langle\psi|\psi\rangle \quad \iff \quad |\langle\phi \cdot \psi\rangle| \leq \|\psi\|\|\phi\|$$

.

Proof. If we rewrite this equation into Dirac notation, it gives us

$$\text{proj}_{|\phi\rangle} |\psi\rangle = \frac{\langle\psi|\phi\rangle}{\langle\phi|\phi\rangle} |\psi\rangle$$

Since $|\phi\rangle$ is a unit vector, $\langle\phi|\phi\rangle = \|\phi\|^2 = 1$, simplifying to the equation in the theorem above. In order to prove that $\langle\phi|\psi\rangle \leq 1$, we apply the Cauchy-Schwarz inequality and get:

$$|\langle\phi|\psi\rangle|^2 \leq \langle\phi|\phi\rangle \langle\psi|\psi\rangle = 1$$

since both $\langle\phi|\phi\rangle = \langle\psi|\psi\rangle = 1$. □

This is important for properties such as the Born Rule, which states that upon measuring an obvservable M on a quantum state $|\psi\rangle$, the probability of obtaining each eigenvalue λ_i is given by $|\langle\lambda_i|\psi\rangle|^2$, where $|\lambda_i\rangle$ is the corresponding eigenvector. Since the aforementioned theorem ensures that $|\langle\lambda_i|\psi\rangle|^2 \leq 1$, we can ensure that it is a valid probability. It is now necessary to reintroduce the general form of the following theorem:

Theorem 6.42 (Cauchy-Schwarz Inequality)**.** Consider any two vectors $\mathbf{u}, \mathbf{v} \in V$, where V is an inner product space.Then we have the following inequality:

$$|\langle\mathbf{u} \cdot \mathbf{v}\rangle|^2 =\leq \langle\mathbf{u} \cdot \mathbf{u}\rangle \langle\mathbf{v} \cdot \mathbf{v}\rangle .$$

This theorem can be adapted into several other forms for various inner product spaces, such as $\mathbb{R}^n$, $\mathbb{C}^n$, etc.

6.5 Orthonormal Bases

We know that bases of vector spaces are a set of linearly independent vectors that span a given space. In this subsection, we will extend this generalization to inner product spaces (which have additional properties such as norm and orthogonality). In these spaces, orthonormal bases are formed by sets of vectors that are both orthognoal and normalized, are an intuitive and funcamental concept for both quantum mechanics and quantum computing, simplifying the representation of quantum states and illustrating key concepts such as superposition, entanglement, state evoluytion, and measurement. Here, Dirac notation is the norm.

Definition 6.43 (Orthonormal Basis)**.** A basis $S = \{|b_1\rangle, |b_2\rangle, \ldots, |b_n\rangle\}$ for an n-dimensional complex inner product space V is said to be orthonormal when they have norm 1 and are all orthogonal to each other. That is:

$$\|b_1\| = \|b_2\| = \ldots = \|b_n\| = 1$$

$$\langle b_i|b_j\rangle = 0 \quad \forall i \neq j$$

This is very much intuitive and analogous to the definition of any orthonormal basis. However, extending this to complex inner product spaces, we have the following theorem:

Theorem 6.44 (Fundamental Property of Orthonormal Bases)**.** Given a basis $S$of a complex inner product space, S is orthonormal if and only if all inner products between pairs of basis vectors satisfy:

$$\langle b_i|b_j\rangle = \delta_{ij}, \qquad \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

When expressing any unit vector in terms of orthonormal basses, the following key property holds true.

Theorem 6.45. Let $|\psi\rangle \in \mathbb{C}^n$ be a *unit vector*, and take a set of orthonormal basis $\{|b_i\rangle\}$. Suppose we can express $|\psi\rangle$ as a linear combination of the basis vectors as:

$$|\psi\rangle = c_1 |b_1\rangle + c_2 |b_2\rangle + \cdots + c_n |b_n\rangle$$

with $\{c_i\} \in \mathbb{C}$, then we have the squared magnitudes of the coefficients satisfying the following constraint:

$$\sum_{i=1}^{n} |c_i|^2 = 1$$

Proof. Using rules of the complex dot product and the properties of an orthonormal basis, we can see that this holds:

$$\langle\psi|\psi\rangle = (c_1^* \langle b_1| + c_2^* \langle b_2| + \cdots c_n^* \langle b_n|)(c_1 |b_1\rangle + c_2 |b_2\rangle + \ldots c_n |b_n\rangle) \tag{4}$$

$$= \sum_{i=1}^{n} c_i^* c_i \langle b_i|b_i\rangle + \sum_{i\neq j} c_i^* c_j \langle b_i|b_j\rangle \tag{5}$$

$$= \sum_{i=1}^{n} c_i^* c_i 1 + \sum_{i\neq j} c_i^* c_j \mathbf{0} \tag{6}$$

$$= \sum_{i=1}^{n} |c_i|^2 \tag{7}$$

Since $|\psi\rangle$ is a unit vector, $\langle\psi|\psi\rangle = \|\psi\|^2 = 1$, so $\langle\psi|\psi\rangle = \sum_{i=1}^{n} |c_i|^2 = 1$. □

In quantum computing, we would take $|\phi\rangle$ to be a **superposition state** within the basis $\{|b_i\rangle\}$, and when measured, the probability that this quantum state would reduce itself to $\{|b_i\rangle\}$ is $|c_i|^2$. The previous theorem simply proves that the summation of these states is equal to one, consistent with the definition of probability. Now note that, starting at this point in the notes, that **all complex vectors expressed in Dirac notation are unit vectors**, unless stated otherwise. This is standard practice for quantum computing literature.

In quantum computing, state vectors $|\phi\rangle$ are typically expressed as superpositions within an orthonormal basis, rather than the standard basis themselves. For example, take the state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, rather than taking the matrix form $|+\rangle = \begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}^T$. This greatly simplifies operations such as outer and tensor products, introduced in the later chapters.

Theorem 6.46. Given $|\psi\rangle$ and an orthonormal basis S in an n-dimensional complex inner product space, $|\psi\rangle$ can be decomposed as a superposition of basis vectors:

$$|\psi\rangle = c_1 |b_1\rangle + c_2 |b_2\rangle + \cdots + c_n |b_n\rangle$$

where each coefficient c_i is given by:

$$c_i = \langle b_i|\psi\rangle$$

Proof. Since we are expressing $|\psi\rangle$ as a linear combination of the basis vectors $\{|b_i\rangle\}$,

$$|\psi\rangle = c_1 |b_1\rangle + c_2 |b_2\rangle + \cdots + c_n |b_n\rangle$$

then in order to find c_i, we only need to take

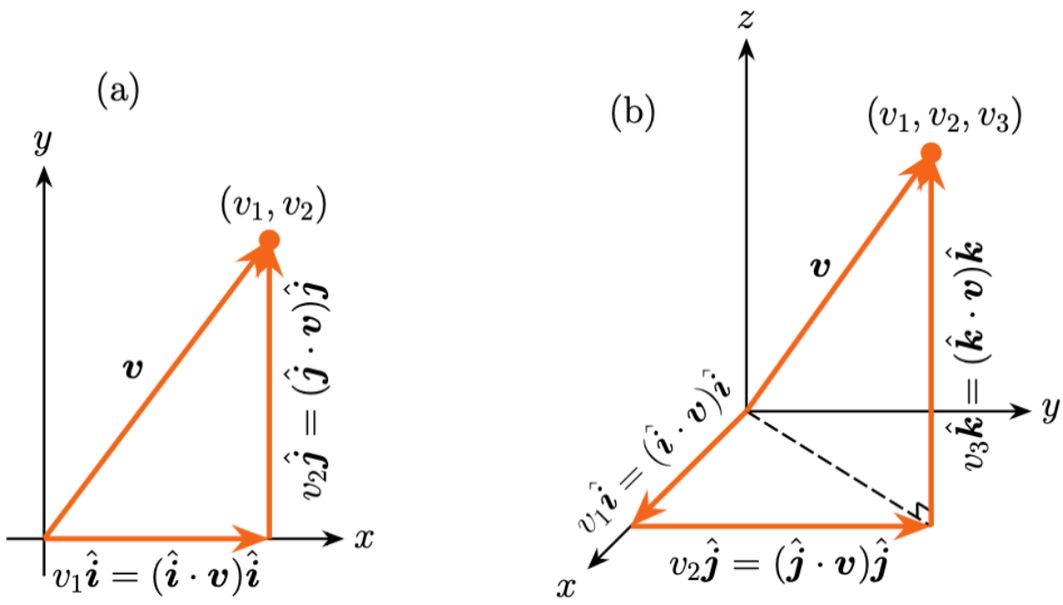
$$\begin{aligned}\langle b_i|\psi\rangle &= \langle b_i|(c_1 |b_1\rangle + c_2 |b_2\rangle + \cdots + c_n |b_n\rangle) \\ &= c_1 \langle b_i|b_1\rangle + c_2 + \cdots + c_n \langle b_i|b_n\rangle \\ &= c_1 \cdot 1 + \sum_{i\neq j} c_j \cdot 0 \\ &= c_i\end{aligned}$$

□

This can also be rewritten as

$$|\psi\rangle = \langle b_1|\psi\rangle |b_1\rangle + \langle b_2|\psi\rangle |b_2\rangle + \cdots + \langle b_n|\psi\rangle |b_n\rangle = \sum_{i=1}^{n} \langle b_i|\psi\rangle |b_i\rangle$$

where each component $\langle b_i|\phi\rangle |b_i\rangle$ is just the projection of $|\phi\rangle$ onto $|b_i\rangle$ or $\text{proj}_{|b_i\rangle}|\psi\rangle$.



Now we know that generally vectors can be decomposed into some naturally existing subspaces and vectors defined in real linear algebra. This translates well to complex inner product spaces, where $c_i = \langle b_i|\psi\rangle$ represents the i-th coordinate of $|\psi\rangle$ relative to the orthonormal basis $S = \{|b_i\rangle\}$, and obtained by projecting $|\psi\rangle$ onto the corresponding basis vector $|b_i\rangle$.

We know that all finite-dimensional inner product spaces have an orthonormal basis. Actually, they have infinite orthonormal bases, and in some situations it is critical to create an orthonormal basis from a non-orthonormal one. The gram-schmidt process defines a methodical algorithm for htis task, which is especially beneficial when working with degenerate eigenvalues.

In order to fully understand this processs, we must first introduce a few theorems that lay the foundation for this.

Theorem 6.47 (Basis from Linearly Independent Vectors)**.** Given a set S with exactly n vectors from a vector space V with $\dim(V) = n$, S forms a basis for V if and only if S is linearly independent.

Proof. Assume that S forms a basis yet is linearly independent. Then it must be possible to remove at least one vector from S without changing it's span. However, this means that the number of vectors will now be strictly less than n, which is in direct contradiction to the previous theorem that states the minimum number of vectors required to span a space is equal to its dimension. Therefore, S must be linearly independent. □

Theorem 6.48 (Orthogonal Basis)**.** Take $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a set of nonzero vectors in an inner product space $V, \dim(V) = n$, and that these vectors are orthogonal to each other s.t. $\langle \mathbf{v}_i|\mathbf{v}_j\rangle = 0, \forall i \neq j$. Thus, S constitues an orthogonal basis of V.

Proof. Since S is already defined as an orthogonal basis, we only need to demonstrate that S is linearly independent. Consider the following equation:

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n = \mathbf{0}$$

If we take the left dot product for each $\mathbf{v}_i$ with this equation, then we have

$$\begin{aligned}\langle \mathbf{v}_i|k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n\rangle &= k_1\langle \mathbf{v}_i|\mathbf{v}_1\rangle + k_2\langle \mathbf{v}_i|\mathbf{v}_2\rangle + \cdots + k_n\langle \mathbf{v}_i|\mathbf{v}_n\rangle \\ &= k_i\langle \mathbf{v}_i|\mathbf{v}_i\rangle\end{aligned}$$

We are left with:

$$k_i\langle \mathbf{v}_i|\mathbf{v}_i\rangle = 0$$

Since $\langle \mathbf{v}_i|\mathbf{v}_i\rangle = \|\mathbf{v}_i\|^2 \neq 0$ for all non-zero vectors, it follows that $k_i = 0$. Hence, S is linearly independent. □

Normalizing each basis vector in an orthogonal basis naturally leads to an orthonormal basis. This transformation is formalized in:

Theorem 6.49. Given an orthogonal basis S for an inner product space V, an orthonormal basis can be obtained by normalizing each basis vector in S:

$$\hat{\mathbf{v}_1} = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \hat{\mathbf{v}_2} = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, \ldots, \hat{\mathbf{v}_n} = \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|}$$

The Gram Schmidt process,

Theorem 6.50 (The Gram-Schmidt Process)**.** The following computational steps, known as the Gram-Schmidt process, transform any basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ of an inner product space V into an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$:

1. $\mathbf{v}_1 = \mathbf{u}_1$
2. $\mathbf{v}_2 = \mathbf{u}_2 - \text{proj}_{\mathbf{v}_1}\mathbf{u}_2$
3. $\mathbf{v}_3 = \mathbf{u}_3 - \text{proj}_{\mathbf{v}_1}\mathbf{u}_3 - \text{proj}_{\mathbf{v}_2}\mathbf{u}_3$
4. $\mathbf{v}_4 = \mathbf{u}_4 - \text{proj}_{\mathbf{v}_1}\mathbf{u}_4 - \text{proj}_{\mathbf{v}_2}\mathbf{u}_4 - \text{proj}_{\mathbf{v}_3}\mathbf{u}_4$

n. $\mathbf{v}_n = \mathbf{u}_n - \sum_{i=1}^{n-1} \text{proj}_{\mathbf{v}_i}\mathbf{u}_n$

Proof. In this proof, we utilize the complex dot product as the inner product. Given that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a basis for V, and hence $\dim(V) = n$, we aim to demonstrate that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ forms an orthogonal basis. This follows from the previous theorem where we showed that $\mathbf{v}_i \cdot \mathbf{v}_j = 0, \forall i \neq j$.

We establish the following relationships: that the inner product of a vector with another one onto itself is just the dot product with that vector, and the dot product of two vectors that are already orthogonal, then the projection of another vector onto the orthogonal vector is simply zero. Mathmatically:

$$\mathbf{v}_i \cdot \text{proj}_{\mathbf{v}_i}\mathbf{u}_j = \mathbf{v}_i \cdot \left(\frac{\mathbf{v}_i \cdot \mathbf{u}_j}{\mathbf{v}_i \cdot \mathbf{v}_i}\mathbf{v}_i\right) = \frac{\mathbf{v}_i \cdot \mathbf{u}_j}{\mathbf{v}_i \cdot \mathbf{v}_i}\mathbf{v}_i \cdot \mathbf{v}_i = \mathbf{v}_i \cdot \mathbf{u}_j, \tag{8}$$

$$\mathbf{v}_i \cdot \text{proj}_{\mathbf{v}_k}\mathbf{u}_j = \mathbf{v}_i \cdot \left(\frac{\mathbf{v}_k \cdot \mathbf{u}_j}{\mathbf{v}_k \cdot \mathbf{v}_k}\mathbf{v}_k\right) = \frac{\mathbf{v}_k \cdot \mathbf{u}_j}{\mathbf{v}_k \cdot \mathbf{v}_k}\mathbf{v}_i \cdot \mathbf{v}_k = 0 \quad \text{if } \mathbf{v}_i \cdot \mathbf{v}_k = 0. \tag{9}$$

With these formulas, we proceed with the proof using mathematical induction.

Base Case:

For $n = 2$, the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are given by:

$$\mathbf{v}_1 = \mathbf{u}_1, \quad \mathbf{v}_2 = \mathbf{u}_2 - \text{proj}_{\mathbf{v}_1}\mathbf{u}_2.$$

Compute the inner product $\mathbf{v}_1 \cdot \mathbf{v}_2$:

$$\begin{aligned}\mathbf{v}_1 \cdot \mathbf{v}_2 &= \mathbf{v}_1 \cdot (\mathbf{u}_2 - \text{proj}_{\mathbf{v}_1}\mathbf{u}_2) \\ &= \mathbf{v}_1 \cdot \mathbf{u}_2 - \mathbf{v}_1 \cdot \text{proj}_{\mathbf{v}_1}\mathbf{u}_2 \\ &= \mathbf{v}_1 \cdot \mathbf{u}_2 - \mathbf{v}_1 \cdot \mathbf{u}_2 \\ &= 0.\end{aligned}$$

Thus, $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal.

Inductive Step:

Assume that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are mutually orthogonal. We prove that $\mathbf{v}_{k+1}$ is orthogonal to all previous $\mathbf{v}_i$ for $i = 1, 2, \ldots, k$. By definition:

$$\mathbf{v}_{k+1} = \mathbf{u}_{k+1} - \sum_{i=1}^{k} \text{proj}_{\mathbf{v}_i}\mathbf{u}_{k+1}.$$

Compute the inner product $\mathbf{v}_j \cdot \mathbf{v}_{k+1}$ for $j = 1, 2, \ldots, k$:

$$\begin{aligned}\mathbf{v}_j \cdot \mathbf{v}_{k+1} &= \mathbf{v}_j \cdot \left(\mathbf{u}_{k+1} - \sum_{i=1}^{k} \text{proj}_{\mathbf{v}_i} \mathbf{u}_{k+1} \right) \\ &= \mathbf{v}_j \cdot \mathbf{u}_{k+1} - \sum_{i=1}^{k} \mathbf{v}_j \cdot \text{proj}_{\mathbf{v}_i} \mathbf{u}_{k+1}.\end{aligned}$$

For $i \neq j$, $\mathbf{v}_j \cdot \text{proj}_{\mathbf{v}_i} \mathbf{u}_{k+1} = 0$ because $\mathbf{v}_j \cdot \mathbf{v}_i = 0$ by the induction hypothesis. For $i = j$:

$$\mathbf{v}_j \cdot \text{proj}_{\mathbf{v}_j} \mathbf{u}_{k+1} = \mathbf{v}_j \cdot \mathbf{u}_{k+1}.$$

Thus:

$$\mathbf{v}_j \cdot \mathbf{v}_{k+1} = \mathbf{v}_j \cdot \mathbf{u}_{k+1} - \mathbf{v}_j \cdot \mathbf{u}_{k+1} = 0.$$

By induction, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}$ are mutually orthogonal. □

An illustration of the Gram-Schmidt Process in $\mathbb{R}^2$ and $\mathbb{R}^3$ is depicted in the above figure a few pages back. We can generally define the projection of a vector onto a subspace as:

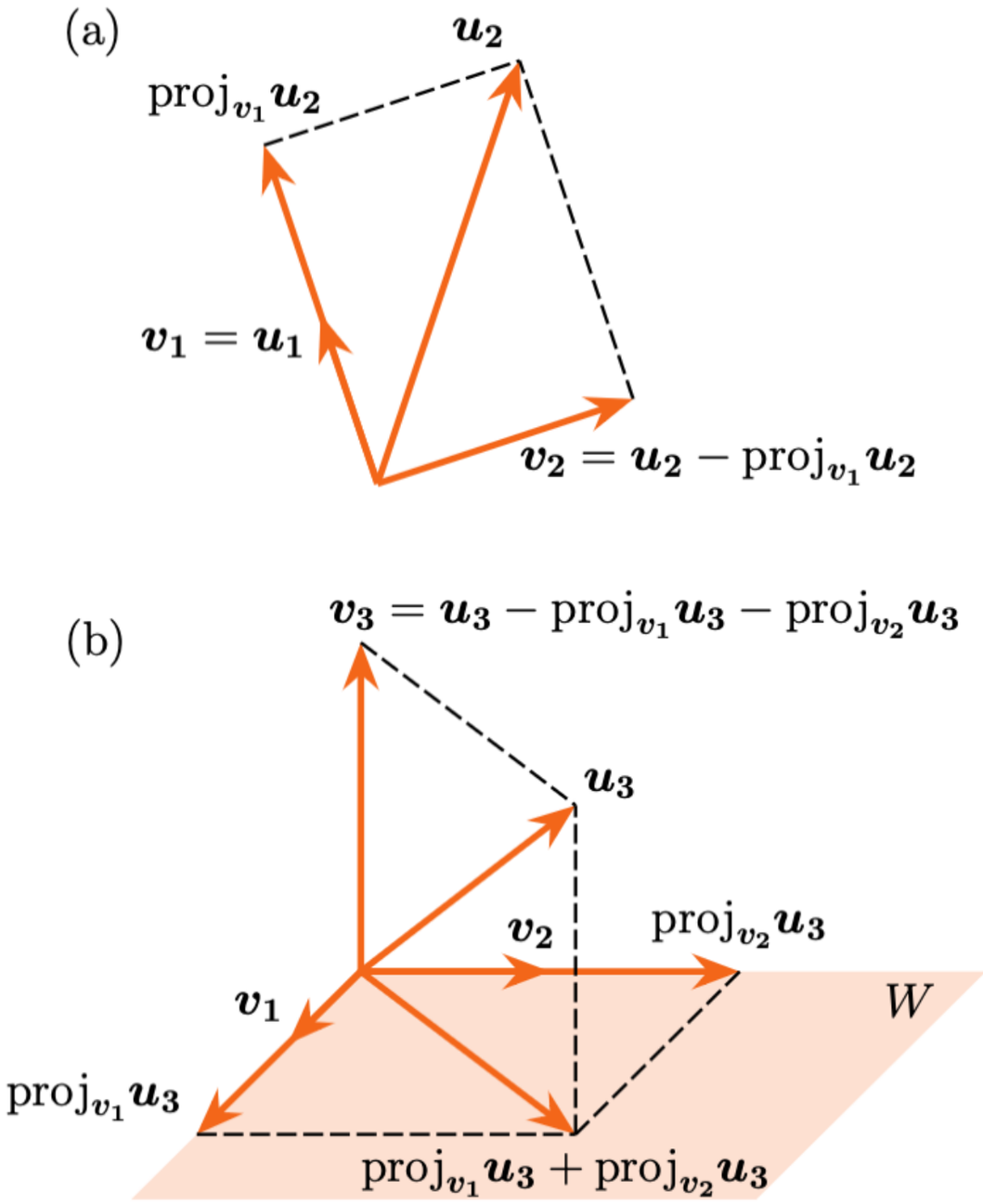
Definition 6.51 (Projection)**.** Let V be an inner product space and $\mathbf{u} \in V$, and take $W \subseteq V$. If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for W, then we define $\text{proj}_W \mathbf{u}$ as:

$$\text{proj}_W \mathbf{u} = \sum_{i=1}^{r} \text{proj}_{\mathbf{v}_i} \mathbf{u} = \sum_{i=1}^{r} \frac{\langle \mathbf{v}_i \cdot \mathbf{u} \rangle}{\langle \mathbf{v}_i \cdot \mathbf{v}_i \rangle} \mathbf{v}_i$$

The component of $\mathbf{u}$ orthogonal to W, denoted by $\text{proj}_{W^\perp} \mathbf{u}$, is:

$$\text{proj}_{W^\perp} \mathbf{u} = \mathbf{u} = \text{proj}_W \mathbf{u}$$

Here, we take $W^\perp$ to be the **orthogonal complement** of W, where the vectors in $W^\perp$ are orthogonal to every vector in W.



Following the previous definition, the Gram-Schmidt Process can be succinctly summarized:

1. $\mathbf{v}_1 = \mathbf{u}_1$,
2. $\mathbf{v}_2 = \text{proj}_{W_1^\perp} \mathbf{u}_2$, where $W_1 = \text{span}\{\mathbf{v}_1\}$,
3. $\mathbf{v}_3 = \text{proj}_{W_2^\perp} \mathbf{u}_3$, where $W_2 = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$,
4. $\mathbf{v}_4 = \text{proj}_{W_3^\perp} \mathbf{u}_4$, where $W_3 = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$,

$\vdots$

n. $\mathbf{v}_n = \text{proj}_{W_{n-1}^\perp} \mathbf{u}_n$, where $W_{n-1} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}\}$,

where W_i represents the subspace spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i$.

Practical Implementation in Python

Standard "textbook" implementations of Gram-Schmidt often fail in quantum computing contexts because they do not correctly handle complex conjugation (using `np.dot` instead of `np.vdot`) or suffer from numerical instability. Below are three verified approaches for quantum states.

1. The Production Approach: NumPy QR

Recommended for actual work. This approach is significantly faster, numerically stable, and utilizes optimized C routines.

```
import numpy as np

def quantum_gram_schmidt(state_vectors):
    # Orthonormalizes a set of quantum state vectors using QR
    decomposition.
    # 'reduced' returns Q with dimensions matching the input columns
    Q, _ = np.linalg.qr(state_vectors, mode='reduced')
    return Q
```

2. The Pedagogical Approach: Modified GS

Recommended for understanding. This is the correct manual implementation for quantum computing. It explicitly handles complex conjugation ($\langle\psi|\phi\rangle$) and uses the **Modified** algorithm to stabilize against rounding errors.

```
def modified_gram_schmidt(A):
    # Ensure complex type to prevent casting errors
    A = A.astype(complex)
    n_rows, n_cols = A.shape
    Q = np.zeros((n_rows, n_cols), dtype=complex)

    for j in range(n_cols):
        v = A[:, j]
        # Subtract projections onto ALL previous vectors
        for i in range(j):
            # np.vdot conjugates the first argument: <Q_i | v>
            projection = np.vdot(Q[:, i], v)
            v = v - projection * Q[:, i]

        norm = np.linalg.norm(v)
        if norm < 1e-10:
            Q[:, j] = np.zeros_like(v)
        else:
            Q[:, j] = v / norm
    return Q
```

3. The Robust Approach: Handling Linear Dependence

Recommended for redundant sets. If your input vectors might be linearly dependent (common when generating overcomplete bases), this version explicitly checks for dependence and skips redundant vectors to avoid division by zero.

```
def robust_gram_schmidt(vectors, tol=1e-10):
    basis = []
    for v in vectors:
        w = np.array(v, dtype=complex)
        for b in basis:
            w -= np.vdot(b, w) * b # Orthogonalize

        norm = np.linalg.norm(w)
        if norm > tol:
            basis.append(w / norm)

    return np.array(basis).T # Return as columns
```

7 Fundamentals of Matrix Algebra

Matrices play a fundamental role across various fields, such as physics, engineering, data science, and especially quantum computing. Here, they represent quantum measurements, evolution, and mixed quantum states. A solid understanding of matrix algebra as well as the relationship between vectors and matrices is profound and essential for advanced topics in quantum computing. This chapter introduces the fundamental principles of matrix algebra, which the proceeding chapters will build upon.

7.1 Matrix Basics

After introducing matrices and vectors in the previous sections, we will provide a thorough exploration of all fundamental concepts essential for the subsequent sections and chapters.

Definition 7.1 (Matrix)**.** A matrix A is defined as a rectangular array with m rows and n columns, represented as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m \times n},$$

where the numbers in the matrix are called entries or elements. The matrix is said to be an $m \times n$ matrix, or $m \times n$ in size.

In linear algebra, matrices are conventionally denoted by capitalized Latin or Greek letters, such as $A, B, U, V, \Sigma, \Delta$. This convention largely holds in quantum computing with a notable exception: the density matrix ρ. Additionally, the notation $[a_{ij}]$ might be used to denote the matrix A, and $(A)_{ij}$ is used to refer specifically to the entries of A, with i, j referring to the individual entries of A.

Definition 7.2 (Column and Row Vectors)**.** A matrix with only one column ($n \times 1$ in size) is called a column vector, denoted as

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

Similarly, a matrix with only one row ($1 \times m$ in size) is called a row vector, denoted as

$$\mathbf{v}^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix}.$$

In matrix algebra, a scalar can be considered as a 1×1 matrix when this interpretation fits the context of the operations. This allows scalars to be seamlessly incorporated into matrix operations, especially when dealing with matrices of varying dimensions.

Definition 7.3 (Transpose)**.** Given an $m \times n$ matrix A, the transpose of A, denoted by A^T, is the $n \times m$ matrix

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}_{n \times m},$$

obtained by interchanging the rows and columns of A.

Definition 7.4 (Conjugate)**.** For a complex matrix A, its conjugate is the matrix A^* obtained by taking the complex conjugate of each entry:

$$A^* = \begin{bmatrix} a_{11}^* & a_{12}^* & \cdots & a_{1n}^* \\ a_{21}^* & a_{22}^* & \cdots & a_{2n}^* \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^* & a_{m2}^* & \cdots & a_{mn}^* \end{bmatrix}.$$

We remind that the complex conjugate is obtained through inverting the sign of the complex component of the individual entries.

Definition 7.5 (Adjoint)**.** For a complex matrix A, its Hermitian adjoint, or adjoint, denoted by $A^\dagger$, is defined as the conjugate transpose:

$$A^\dagger = (A^*)^T = (A^T)^* = \begin{bmatrix} a_{11}^* & a_{21}^* & \cdots & a_{m1}^* \\ a_{12}^* & a_{22}^* & \cdots & a_{m2}^* \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n}^* & a_{2n}^* & \cdots & a_{mn}^* \end{bmatrix}_{n\times m}.$$

For a real matrix, it's complex conjugate is just the matrix itself, while the adjoint is equal to the transpose. Just some matrix manipulation here, nothing else. While some textbooks might use $\overline{A}$ as the conjugate of a matrix, A^* as the conjugate transpose (adjoint) of the matrix.

Theorem 7.6. For any matrix A, the following property holds:

$$(A^T)^T = A, \quad (A^\dagger)^\dagger = A$$

Definition 7.7 (Square Matrix)**.** A matrix that has an equal number of rows and columns is termed a square matrix, or an $n \times n$ matrix.

Definition 7.8 (Main Diagonal)**.** The main diagonal of a square matrix consists of the entries where the row and column indices are the same. In an $n \times n$ matrix A, the entry a_{ij} is on the main diagonal if $i = j$.

Definition 7.9 (Matrix Equality)**.** Two matrices of the same size, $A_{m\times n}$ and $B_{m\times n}$, are said to be equal if all corresponding entries are identical:

$$a_{ij} = b_{ij}, \quad \text{for all } i = 1, 2, \ldots, m \text{ and } j = 1, 2, \ldots, n.$$

If two matrices A and B are identical in this sense, this relationship is denoted simply by

$$A = B.$$

Definition 7.10 (Symmetric Matrix)**.** A square matrix A is termed a symmetric matrix if $A = A^T$.

While symmetric matrices are more imporant in the analysis of real matrices, Hermetian matrices are more pertient especially as they remain invariant when taking the Hermitian adjoint.

Definition 7.11 (Hermitian Matrix)**.** A square complex matrix A is said to be Hermitian or a Hermitian matrix if $A = A^\dagger$.

Hermitian matrices are of crucial importance in quantum mnechanics and computing since they represent operators for quantum measurements. A matrix is Hermitian if it is square with entries that are conjugate symmetric about the main diagonal. Any real symmetric matrix is Hermitian.

Example. The following matrix is a Hermitian Matrix:

$$A = \begin{bmatrix} 0 & a - ib & c - id \\ a + ib & 1 & m - in \\ c + id & m + in & 2 \end{bmatrix}$$

Now similar to vectors, the fundamental algebraic operations for matrices include addition and scalar multiplication, which are defined as follows.

Definition 7.12 (Matrix Addition)**.** Matrix addition is defined only for two matrices of the same size, $A_{m\times n}$ and $B_{m\times n}$. Their sum, denoted by $A + B$, is an $m \times n$ matrix computed entrywise as

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}.$$

Definition 7.13 (Scalar Multiplication of Matrices)**.** Let $A_{m\times n}$ be a matrix and $k \in \mathbb{R}$ a scalar. The scalar multiple kA is the $m \times n$ matrix obtained by multiplying each entry of A by k:

$$kA = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{bmatrix}.$$

Definition 7.14 (Negative of a Matrix)**.** The negative of a matrix $A_{m\times n}$, denoted by $-A$, is defined as the scalar multiple $(-1)A$. Explicitly,

$$-A = \begin{bmatrix} -a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & -a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{m1} & -a_{m2} & \cdots & -a_{mn} \end{bmatrix}.$$

Definition 7.15 (Matrix Subtraction)**.** For two matrices $A_{m\times n}$ and $B_{m\times n}$, matrix subtraction $A - B$ is defined as the addition of A and the negative of B:

$$A - B = A + (-B) = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \cdots & a_{1n} - b_{1n} \\ a_{21} - b_{21} & a_{22} - b_{22} & \cdots & a_{2n} - b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} - b_{m1} & a_{m2} - b_{m2} & \cdots & a_{mn} - b_{mn} \end{bmatrix}.$$

While these are visually appealing methods for notation, they are quite cumbersome in notation.

Theorem 7.16. Given two matrices $A_{m\times n} = [a_{ij}]$ and $B_{m\times n} = [a_{ij}]$ of the same size and a scalar k, the basic algebraic operations can be computed element wise as follows:

$$\begin{aligned} (A + B)_{ij} &= a_{ij} + b_{ij} \\ (kA)_{ij} &= ka_{ij} \\ (-A)_{ij} &= -a_{ij} \\ (A - B)_{ij} &= a_{ij} - b_{ij} \end{aligned}$$

For transpose and Hermitian of matrices, the element-wise formula is as follows:

Theorem 7.17. Given $A_{m\times n} = [a_{ij}]$, the entries of A^T and $A^\dagger$ are given element-wise by:

$$\begin{aligned} (A^T)_{ji} &= a_{ij} \\ (A^\dagger)_{ji} &= a^*_{ij} \end{aligned}$$

Since matrices adhere to the same fundamental operations as vectors, they naturally form vecor spaces. In order to complete the definition, we first need to define the zero matrix.

Definition 7.18 (Zero Matrix)**.** Among all matrices of a specific size $m \times n$, the unique zero matrix, denoted by 0, is defined as the matrix whose entries are all zero:

$$0_{m\times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{m\times n}$$

or in element-wise notation,

$$(0)_{ij} = 0, \quad \forall i = 1, 2, \ldots, m \text{ and } j = 1, 2, \ldots, n.$$

Theorem 7.19. The following properties hold true for any matrices $A, B, C, 0$ of the same size and scalars k, m:

$$\begin{aligned} A + B &= B + A \\ A + 0 &= 0 + A = A \\ (A + B) + C &= A + (B + C) \\ k(A + B) &= kA + kB \\ (k + m)A &= kA + mA \\ (km)A &= k(mA) \\ 0A &= 0 \\ 1A &= A \end{aligned}$$

While this might seem familiar to you, it is worth reiterating given that there are some key differences between vectors and matrices. At the same time, all of the aformentioned properties remain true if addition is replaced with subtraction. It's rather easy to verfiy that matrices of a specific size $m \times n$ satisfy the all of the axioms required for a vector spaces given in the previous theorem. Therefore, we have:

Theorem 7.20. All matrices of size $m \times n$ constitute a vector space. For real matrices over $\mathbb{R}$, this is defined as $\mathbb{R}^{m\times n}$, nad for complex matrices over $\mathbb{C}$, this is denoted as $\mathbb{C}^{m\times n}$.

For the matrix space $\mathbb{R}^{2\times 2}$, we have the following basis:

$$B_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad B_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

In general, there for a vector space in $\mathbb{R}^{m\times n}/\mathbb{C}^{m\times n}$, there are going to be mn basis matrices in the mn dimensional space. This warrants the introduction of a more broader concept, namely tensors. If we take vectors to be one-dimensional arrays in $\mathbb{C}^n$, and matrices to be two-dimensional arrays in $\mathbb{C}^{n\times n}$, and three dimensional arrays in $\mathbb{C}^{n\times n\times n}$, the mathmatical term given to such objects in higher dimensions is defined as tensors. By this definition, vectors are 1D tensors, matrices are 2D tensors, and then generally k-dimensional tensors will reside in $\mathbb{C}^{n\times n\times \cdots \times n}$.

7.2 Matrix Multiplication

While the previous subsection established the foundational principles of matrix algebra, behavior of matrices largely seemed like vectors. However, matrix multiplication significantly differs from vector multiplication, which you will be able to see in this chapter. At the same time, we note that this will mainly focus on square matrices as they are of particular importance in quantum computing.

Before diving into matrix multiplication, it is crucial to understand how to partition a matrix into its row and column vectors, which are the two fundamental partitions of a matrix.

Definition 7.21 (Partition into Row and Column Vectors)**.** Given a matrix $A_{m\times n}$, it can be partitioned into its row vectors as

$$A = \left[\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \hline a_{21} & a_{22} & \cdots & a_{2n} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline a_{m1} & a_{m2} & \cdots & a_{mn} \end{array}\right] = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_m \end{bmatrix},$$

where $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$ represent the row vectors of A. The matrix can also be partitioned into its column vectors as

$$A = \left[\begin{array}{c|c|c|c} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{array}\right] = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix},$$

where $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ are the column vectors of A.

Matrix partitioning will be essential to what is to come in a moment. We will adopt the convention that $\mathbf{r}_i$ and $\mathbf{c}_j$ represent row and column vectors of a given matrix. We previously introduced the dot product as:

$$\mathbf{rc} = \begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = r_1c_1 + r_2c_2 + \cdots + r_nc_n = \mathbf{r} \cdot \mathbf{c}.$$

Now we will the dot product to matrix products. For the matrices $A_{m\times p}$ and $B_{p\times n}$, note that the number of columns in A must be equal to the number of rows in B. While you are multiplying the rows in A with the columns in B, the entries in a row from A and entries in a column from B must be the same. Therefore, the elements in a column of A and the number of elements in a row of B must add up.

Definition 7.22 (Matrix Product)**.** The matrix product AB is defined when the number of columns in $A_{m\times p}$ matches the number of rows in $B_{p\times n}$. Partition A into its row vectors and B into its column vectors:

$$A = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_m \end{bmatrix}, \qquad B = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix}.$$

The product AB is consequently defined as an $m \times n$ matrix, computed as

$$(AB)_{m\times n} = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{c}_1 & \mathbf{r}_1 \cdot \mathbf{c}_2 & \cdots & \mathbf{r}_1 \cdot \mathbf{c}_n \\ \mathbf{r}_2 \cdot \mathbf{c}_1 & \mathbf{r}_2 \cdot \mathbf{c}_2 & \cdots & \mathbf{r}_2 \cdot \mathbf{c}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{r}_m \cdot \mathbf{c}_1 & \mathbf{r}_m \cdot \mathbf{c}_2 & \cdots & \mathbf{r}_m \cdot \mathbf{c}_n \end{bmatrix}.$$

This relationship can be succinctly expressed using element-wise notation:

Theorem 7.23. Given two matrices $A_{m\times p}$ and $B_{p\times n}$, and $C_{m\times n} = AB$, the entries of C are given by the row–column rule for matrix multiplication:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj},$$

where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

$$AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mp} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pj} & \cdots & b_{pn} \end{bmatrix}$$

Now this brings up an imporant aspect of matrix multiplication, which is that matrix multiplication generally does not commute, or that $AB \neq BA$. This takes place when either one is not defined, or that they differ in size, or that they are simply not equal. This property stands in sharp contrast with scalar multiplication and the dot product of vectors, which are commutative or conjugate commutative. In quantum computing, the special case of two matrices commuting holds significant importance. To formally express this relationship, we have the following definition:

Definition 7.24 (Commutator)**.** Given two square matrices of equal dimension, their commutator is denoted as $[A, B]$ and defined by:

$$[A, B] = AB - BA$$

When $[A, B] = 0$ or equivalently $AB = BA$, the matrices A and B are said to commute.

Building on this concept of commutation, where matrix multiplication is symmetric, we explore its counterpart, namely the anti-commutator. This operator provides an alternative:

Definition 7.25 (Anti-Commutator)**.** Given two square matrices of the same size, their anti-commutator is denoted as A, B and defined by:

$$A, B = AB + BA$$

When $A, B = 0$ or equivalently $AB = -BA$, the matrices A and B are said to anticommute.

Commutation and anticommutation relations are particularly relevant when working with Pauli matrices, which are going to be thouroughly covered in Section 12. Despite the non-commutativity of matrix multiplication, it shares many properties with scalar multiplication. For instance:

Theorem 7.26. Assuming all matrix multiplications are valid, the following identities hold:

$$\begin{aligned} A(B+C) &= AB + AC \\ (A+B)C &= AC + BC \\ (AB)C &= A(BC) \end{aligned}$$

Proof. For the operation $A(B+C)$ to be valid, consider matrices $A_{m\times p} = [a_{ik}]$, $B_{p\times n} = [b_{kj}]$, and $C_{p\times n} = [c_{kj}]$ for $i = 1, 2, \ldots, m$, $k = 1, 2, \ldots, p$, and $j = 1, 2, \ldots, n$.

Define $D = A(B + C) = [d_{ij}]$ and $F = AB + AC = [f_{ij}]$ (both are $m \times n$ matrices). Compute the entries of D and F using element-wise notation for matrix multiplication:

$$d_{ij} = \sum_{k=1}^{p} a_{ik}(b_{kj} + c_{kj}), \qquad f_{ij} = \sum_{k=1}^{p} a_{ik}b_{kj} + \sum_{k=1}^{p} a_{ik}c_{kj}.$$

By applying the distributive laws of scalars within the summation,

$$d_{ij} = \sum_{k=1}^{p} a_{ik}(b_{kj} + c_{kj}) = \sum_{k=1}^{p}(a_{ik}b_{kj} + a_{ik}c_{kj}) = \sum_{k=1}^{p} a_{ik}b_{kj} + \sum_{k=1}^{p} a_{ik}c_{kj} = f_{ij},$$

for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Hence, by matrix equality, $D = F$. □

The proofs of the two other properties follow a similar pattern and are omitted for brevity. At the same time, when we are dealing with transpose or adjoint operations, the following identities are very useful:

Theorem 7.27. Assuming all matrix multiplications are valid, the following identities hold:

$$\begin{aligned} (AB)^* &= A^*B^* \\ (AB)^T &= B^TA^T \\ (AB)^\dagger &= B^\dagger A^\dagger \end{aligned}$$

Proof. To prove the part of the above theorem, consider $A_{m\times p} = [a_{ik}]$ and $B_{p\times n} = [b_{kj}]$ for $i = 1, 2, \ldots, m$, $k = 1, 2, \ldots, p$, and $j = 1, 2, \ldots, n$.

Let $C_{n\times m} = (AB)^T = [c_{ji}]$ and $D_{n\times m} = B^TA^T = [d_{ji}]$. Using the row–column rule, compute the entries of C and D as follows:

$$c_{ji} = \left((AB)^T\right)_{ji} = (AB)_{ij} = \sum_{k=1}^{p} a_{ik}b_{kj},$$

$$d_{ji} = (B^TA^T)_{ji} = \sum_{k=1}^{p}(B^T)_{jk}(A^T)_{ki} = \sum_{k=1}^{p} b_{kj}a_{ik}.$$

Since multiplication of scalar entries is commutative, $a_{ik}b_{kj} = b_{kj}a_{ik}$ for all i, j, k, it follows that $c_{ji} = d_{ji}$. Therefore, $C = D$, proving that

$$(AB)^T = B^TA^T.$$

The third equation builds upon the previous logic, where

$$(AB)^\dagger = \left((AB)^*\right)^T = (A^*B^*)^T = (B^*)^T(A^*)^T = B^\dagger A^\dagger.$$

□

Now, matrix multiplication can also be performed using a technique known as the column-row expansion. This is an alternative way of computing the product of matrices by expanding it into a sum of individual matrices.

Theorem 7.28 (Column–Row Expansion)**.** Consider matrices $A_{m\times p}$ and $B_{p\times n}$. Partition A into its column vectors and B into its row vectors:

$$A = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_p\end{bmatrix}, \qquad B = \begin{bmatrix}\mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_p\end{bmatrix}.$$

Then, the product AB can be calculated by summing the products of these vectors:

$$AB = \mathbf{c}_1\mathbf{r}_1 + \mathbf{c}_2\mathbf{r}_2 + \cdots + \mathbf{c}_p\mathbf{r}_p = \sum_{i=1}^{p} \mathbf{c}_i\mathbf{r}_i.$$

Note that in this each product term become full-sized matrices, and thus the entire product is express as a sum of such matrices. However, we can't have a little fun without Dirac notation, and in the field of quantum computing, the matrices are primarily square matrices and column vectors. In this subsection, we will delve into some of the most frequently encountered matrix multiplication scenarios in quantum computing utilizing bra-ket notation. For example, we know that the inner product (essentially a dot product) effectively produces a scalar and is defined as:

$$\langle \mathbf{u}\cdot\mathbf{v}\rangle = \langle \mathbf{u}|\mathbf{v}\rangle = \begin{bmatrix}u_1^* & u_2^* & \cdots & u_n^*\end{bmatrix}\begin{bmatrix}v_1 \\ v_2 \\ \vdots \\ v_n\end{bmatrix} = u_1^*v_1 + u_2^*v_2 + \cdots + u_n^*v_n.$$

Hereforth, we also define a matrix-vector multiplication as multiplying an $n \times n$ matrix by an $n \times 1$ column vector, yielding another $n \times 1$ column vector. In dirac notation, this operation is expressed as:

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

In this context, A is referred to as a transformation matrix as it transforms the vector $|u\rangle$ into $|v\rangle$. The matrix representation of this operation is as follows:

$$\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}\begin{bmatrix}u_1 \\ u_2 \\ \vdots \\ u_n\end{bmatrix} = \begin{bmatrix}v_1 \\ v_2 \\ \vdots \\ v_n\end{bmatrix}.$$

This operation is fundamental in quantum computing, particularly in the context of quantum gates, as we can simply take the vector $|u\rangle$ as the input quantum gate, and $|v\rangle$ as the output quantum gate. A embodies the matrix representation of the quantum gate, which is mathmatical candor at its finest. Next we will introduce the outer product of two vectors, which is defined as follows:

Definition 7.29 (Outer Product of Two Vectors)**.** Given two vectors $|u\rangle$ and $|v\rangle$ in $\mathbb{C}^n$, their outer product is an $n \times n$ matrix defined as

$$|u\rangle\langle v| = \begin{bmatrix}u_1 \\ u_2 \\ \vdots \\ u_n\end{bmatrix}\begin{bmatrix}v_1^* & v_2^* & \cdots & v_n^*\end{bmatrix} = \begin{bmatrix}u_1v_1^* & u_1v_2^* & \cdots & u_1v_n^* \\ u_2v_1^* & u_2v_2^* & \cdots & u_2v_n^* \\ \vdots & \vdots & \ddots & \vdots \\ u_nv_1^* & u_nv_2^* & \cdots & u_nv_n^*\end{bmatrix}.$$

A pretty useful formula for outer products is as follows:

$$|v\rangle\langle u| = (|u\rangle\langle v|)^\dagger$$

This can be demonstrated by applying the conjugate transpose prooperty:

$$(|u\rangle\langle v|)^\dagger = (\langle v|)^\dagger(|u\rangle)^\dagger = |v\rangle\langle u|$$

Example. Compute $|0\rangle\langle 0|$, $|0\rangle\langle 1|$, $|1\rangle\langle 0|$, and $|1\rangle\langle 1|$ in matrix representation.

Using the computational basis, we find:

$$|0\rangle\langle 0| = \begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \qquad |0\rangle\langle 1| = \begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}0 & 1\end{bmatrix} = \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix},$$

$$|1\rangle\langle 0| = \begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix} = \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \qquad |1\rangle\langle 1| = \begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}0 & 1\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}.$$

Now in this example, we see that it is possible to represent a general matrix $A_{2\times 2}$ in terms of outer products. This approach can be generalized to any matrix using the following theorem.

Theorem 7.30 (Matrix Basis Decomposition)**.** Let $\{|0\rangle, |1\rangle, \ldots, |d-1\rangle\}$ be the computational basis of $\mathbb{C}^d$. For a general matrix $A \in \mathbb{C}^{d\times d}$ given by

$$A = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0(d-1)} \\ a_{10} & a_{11} & \cdots & a_{1(d-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{(d-1)0} & a_{(d-1)1} & \cdots & a_{(d-1)(d-1)} \end{bmatrix},$$

it can be decomposed into a sum of outer products:

$$A = \sum_{i,j} a_{ij} |i\rangle\langle j|, \qquad \text{for } i, j = 0, 1, \ldots, d-1.$$

Example. Given the matrix A as described above, demonstrate that:

$$A|k\rangle = \sum_i a_{ik} |i\rangle$$

Proof.

$$\begin{aligned} A|k\rangle &= \left(\sum_{i,j} a_{ij} |i\rangle\langle j|\right)|k\rangle \\ &= \sum_{i,j} a_{ij} |i\rangle\langle j|k\rangle \\ &= \sum_{i,j} a_{ij} |i\rangle \delta_{jk} \\ &= \sum_i a_{ij} |i\rangle \end{aligned}$$

□

Where the equailty $\langle j|k\rangle = \delta_{jk}$ (the Koronecker delta) simplifies the expression by collapsing the sum over j to only the term where $j = k$. Next we will consider basis change matrices based on the previous example. In quantum computing, basis changes are often expressed as sums of outer products between two basis vectors. For example, consider changing from the set of vectors $|0\rangle, |1\rangle$ to the basis $|+\rangle, |-\rangle$. We take the matrix U that performs this mapping to be

$$U = |+\rangle\langle 0| + |-\rangle\langle 1|$$

We can verify that U completes this mapping by showing, for example, that $(|+\rangle\langle 0|)|0\rangle = |+\rangle\langle 0|0\rangle$. This result can be generalized to the following formula:

Theorem 7.31. Given a matrix U as an outer product of $U = |u\rangle\langle v|$ and a vector $|w\rangle$, the product $U|w\rangle$ is computed as:

$$U|w\rangle = \langle v|w\rangle |u\rangle$$

which is just a scalar multiple of $|u\rangle$

Proof. We have

$$U|w\rangle = (|u\rangle\langle v|)|w\rangle = |u\rangle(\langle v|w\rangle) = |u\rangle\langle v|w\rangle = \langle v|w\rangle|u\rangle$$

Since $\langle v|w\rangle$ is a scalar, the result is just a scalar multiple of $|u\rangle$. □

This provides you with an effective yet clean change of basis method. A more comprehensive discussion on changing basis will be presented in the next section. Sometimes, it is also necessary to compute the inner product between two vectors, wherein one or both vectors are the results of a matrix-vector product. For instance in classical linear algebra, we have:

$$\langle Ax|y\rangle = (A\mathbf{x})\cdot\mathbf{y}$$
$$\langle x|Ay\rangle = \mathbf{x}\cdot(A\mathbf{y})$$
$$\langle Ax|Ay\rangle = (A\mathbf{x})\cdot(A\mathbf{y})$$

For brevity, we denote $|Ax\rangle \equiv A|x\rangle$ and $\langle Ax| \equiv (A|x\rangle)^\dagger$, and adopt the following identities when dealing with these forms:

Theorem 7.32. The inner product on the left-hand side can be expressed as matrix products on the right-hand side:

$$\langle x|Ay\rangle = \langle x|A|y\rangle$$
$$\langle Ax|y\rangle = \langle x|A^\dagger|y\rangle$$
$$\langle Ax|Ay\rangle = \langle x|A^\dagger A|y\rangle$$

Proof. These relationships can be defined using the rules of bra-ket notation:

$$\langle x|Ay\rangle = \langle x|\ |Ay\rangle = \langle x|A|y\rangle$$
$$\langle Ax|y\rangle = \langle Ax|\ |y\rangle = (|Ax\rangle)^\dagger|y\rangle = |x\rangle^\dagger A^\dagger|y\rangle = \langle x|A^\dagger|y\rangle$$
$$\langle Ax|Ay\rangle = \langle Ax|\ |Ay\rangle = (|Ax\rangle)^\dagger A|y\rangle = \langle x|A^\dagger A|y\rangle$$

□

For the given matrix A, the expression $\langle x|A|y\rangle$ defines the mapping $\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$, which is linear in the second argument $|y\rangle$ and conjugate linear in the first argument $|x\rangle$, which refers to $\langle x|A|y\rangle$ as something in sesquilinear form, a generalization of inner products in complex vector spaces. What does this mean?

When we say the mapping is linear in $|y\rangle$, we mean it behaves exactly as you would expect standard multiplication to behave. It preserves addition and scalar multiplication without changing the scalar. Mathematically, if you replace $|y\rangle$ with a linear combination $c_1|y_1\rangle + c_2|y_2\rangle$ (where c_1, c_2 are complex numbers), the expression splits cleanly:

$$\langle x|A(c_1|y_1\rangle + c_2|y_2\rangle) = c_1\langle x|A|y_1\rangle + c_2\langle x|A|y_2\rangle$$

Conjugate linear (sometimes called antilinear) is where complex spaces differ from real ones. It means that the mapping preserves addition, but when you pull out a scalar, it comes out as its complex conjugate.

$$(c_1|x_1\rangle + c_2|x_2\rangle)^\dagger A|y\rangle = c_1^*\langle x_1|A|y\rangle + c_2^*\langle x_2|A|y\rangle$$

Recall that in Dirac notation (bra-ket notation), the "bra" $\langle x|$ corresponds to the Hermitian conjugate (adjoint) of the column vector x. If $|x\rangle$ is the column vector x, then $\langle x|$ is the row vector $x^\dagger$ (conjugate transpose). If you scale the vector x by a complex number c, the conjugate transpose scales by c^*.

$$(cx)^\dagger = c^* x^\dagger$$

So therefore the term Sesquilinear literally means "one-and-a-half linear." It is "one" linear because of the second argument ($|y\rangle$). It is "half" linear (antilinear) because of the first argument ($\langle x|$).

Sesquilinearity ensures that the norm of a vector (the inner product of a vector with itself) is always a real number. If it were linear in both arguments, $\langle cx|cx\rangle$ would equal $c^2 \langle x|x\rangle$, which could be a complex number. Because it is conjugate linear in the first slot, we get:

$$\langle cx|cx\rangle = c^* c \langle x|x\rangle = |c|^2 \langle x|x\rangle$$

This guarantees that lengths/probabilities remain real and non-negative. Now going along with the previous theorems, if A is Hermitian and both vectors on either side are equal, we call this a **Hermitian quadratic form** for the expression:

$$\langle\psi| M |\psi\rangle$$

where M is Hermitian ($M = M^\dagger$). This form is critical in quantum computing as it represents the statistical average of an observable M on the state $|\psi\rangle$ always yielding a real scalar. This is explored further below.

Theorem 7.33. Given a Hermitian matrix $M \in \mathbb{C}^{n\times n}$ and a complex vector $|\psi\rangle \in \mathbb{C}^n$, $\langle\psi| M |\psi\rangle$ is always real.

Proof. Taking $|x\rangle = |y\rangle = |\psi\rangle$ from previous equations, we have:

$$\langle\psi| M^\dagger |\psi\rangle = (\langle\psi| M |\psi\rangle)^*$$

Since M is Hermitian, it holds that $M = M^\dagger$, and thus:

$$\langle\psi| M |\psi\rangle = \langle\psi| M^\dagger |\psi\rangle$$

This implies that:

$$\langle\psi| M |\psi\rangle = (\langle\psi| M |\psi\rangle)^* \quad \Longrightarrow \quad \langle\psi| M |\psi\rangle \in \mathbb{R}$$

□

7.3 Matrix Inverses

Very similar to how real numbers have inverses, matrices have them too.

Definition 7.34 (Identity Matrix)**.** The identity matrix, denoted by I_n, is an $n \times n$ square matrix characterized by ones on its main diagonal and zeros in all off-diagonal positions:

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

The identity matrix can also be defined using the Kronecker delta function as:

$$I_n = [\delta_{ij}], \qquad \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Theorem 7.35. For any matrices $A_{m\times n}$, the identity matrix I_n and I_m satisfies the properties $AI_n = A$ and $I_m A = A$.

Proof. Let $A = [a_{ij}]$ and the product $C = AI_n = [c_{ij}]$:

$$c_{ij} = \sum_{k=1}^{n} a_{ik}\delta_{kj}.$$

The only non-zero term in the summation occurs when $k = j$, yielding:

$$c_{ij} = a_{ij}\delta_{jj} = a_{ij}, \qquad \forall i, j,$$

affirming that $AI_n = A$. We can apply a similar method for validating $I_m A = A$. □

Note that we can omit the subscript when the context clearly identifies the size of the identity matrix I. Next, we will define the inverse of a matrix, which is conceptually analogous to the inverse of a scalar, but entails a few more hoops that you need to jump through.

Definition 7.36 (Inverse of a Matrix)**.** Consider a square matrix $A_{n\times n}$. If there exists a matrix $B_{n\times n}$ such that $BA = AB = I$, then we consider A **invertible**, B as the **inverse** of A, denoted by A^{-1}. If there exists no such B, then A is described as **singular**.

We know that only square matrices are invertible. However, we also know that the inverse of a matrix is unique.

Theorem 7.37. The inverse of a matrix, if it exists, is unique.

Proof. Assume B and C are the inverse of a matrix A. By definition:

$$AB = BA = I = AC = CA.$$

So if we take

$$BAC = B(AC) = BI = B$$

which is by defnition equal to

$$BAC = (BA)C = IC = C$$

it is easy to see that $B = C$ and that the inverse of a matrix is unique. □

Theorem 7.38. For two square matrices A and B,

$$AB = I \Leftrightarrow BA = I \quad \Longleftrightarrow \quad A = B^{-1} \Leftrightarrow B = A^{-1}.$$

It is out of the scope of this text to compute the inverse of matrices. However, not all matrices are invertible. The zero matrix is an obvious example of a singular matrix. In general, you can think of singular matrices having some form of information reduced from it so that compression and expansion is not able to recreate in inverted image of it. More formally:

Theorem 7.39. A square matrix A is singular if its row vectors or column vectors are linearly independent.

Proof. Consider a square matrix A with n linearly dependent column vectors. Let $\text{span}(S) = A$. Since the column vectors are linearly dependent, that implies that $\dim(S) < n$. Next, we consider a matrix B partitioned into its column vectors. If we define

$$C = AB = \sum_{i=1}^{n} A\mathbf{b}_i,$$

then each product is going to be linearly dependent since A is linearly dependent, implying that

$$\forall \mathbf{b}_i, \quad A\mathbf{b}_i \in \text{span}(S) \Longrightarrow C \neq I.$$

It is easy to see that this holds true for all matrices B such that C will never be the identity matrix I, and hence A must be singular. A similar method can be used to prove that the same holds for linearly independent row vectors. □

As a brief note, while invertibility is a fundamental property of square matrices and it does come up time and time again, the need to examine invertibility is not necessary in quantum computing. This is primarily so becuase unitary matrices, the mathmatical representation for quantum gates, are by definition invertible. As a result, the emphasis lies on unitary matrices and their properties, which are central to quantum computing.

In practice, it is often more desirable to represent solutions of problems as a vector. Such problems are often formulated using matrix equations.

Theorem 7.40. Given an invertible matrix $A \in \mathbb{C}^{n\times n}, |\phi\rangle \in \mathbb{C}^n$, the matrix equation

$$A\,|\phi\rangle = |\psi\rangle$$

admits a unique solution:

$$|\phi\rangle = A^{-1}\,|\psi\rangle\,.$$

Proof. Multiplying the inverse on the right hand side of the equation yields:

$$A^{-1}A\left|\phi\right\rangle = A^{-1}\left|\psi\right\rangle \qquad \Rightarrow \qquad I\left|\phi\right\rangle = A^{-1}\left|\psi\right\rangle \qquad \Rightarrow \qquad \left|\phi\right\rangle = A^{-1}\left|\psi\right\rangle$$

□

Theorem 7.41. Given a series of invertible matrices $A_1, A_2, \ldots, A_{n-1}, A_n$ of the same dimensions, the following identity holds true:

$$(A_1A_2\cdots A_{n-1}A_n)^{-1} = A_n^{-1}A_{n-1}^{-1}\cdots A_2^{-1}A_1^{-1}.$$

Proof. We verify the identity by showing that the product of the sequence of matrices and their inverses results in the identity matrix, I:

$$\begin{aligned}(A_n^{-1}A_{n-1}^{-1}\cdots A_2^{-1}A_1^{-1})(A_1A_2\cdots A_{n-1}A_n) &= A_n^{-1}A_{n-1}^{-1}\cdots A_2^{-1}IA_2\cdots A_{n-1}A_n\\ &= A_n^{-1}A_{n-1}^{-1}\cdots A_3^{-1}IA_3\cdots A_{n-1}A_n\\ &= \cdots\\ &= A_n^{-1}A_{n-1}^{-1}A_{n-1}A_n\\ &= A_n^{-1}IA_n = A_n^{-1}A_n = I.\end{aligned}$$

Therefore we conclude that

$$(A_1A_2\cdots A_{n-1}A_n)^{-1} = A_n^{-1}A_{n-1}^{-1}\cdots A_2^{-1}A_1^{-1}.$$

□

Theorem 7.42. If A is an invertible matrix, then it's adjoint ($A^\dagger$) is also invertible. The inverse of $A^\dagger$ is given by:

$$\left(A^\dagger\right)^{-1} = (A^{-1})^\dagger$$

Proof. Multiplying the left side of the equation with $A^\dagger$ yields:

$$I = A^\dagger(A^{-1})^\dagger \qquad \Rightarrow \qquad I^\dagger = AA^{-1} = I$$

□

Next we will move on to diagonal matrices, defined as square matrices with non-zero entries only on the main diagonal, play a signifcant role as a source of convinience in linear algebra.

Definition 7.43 (Diagonal Matrix)**.** A diagonal matrix is a square matrix where all off-diagonal entries are zero. A general $n \times n$ diagonal matrix is represented as

$$\Lambda = \begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ 0 & 0 & \ddots & 0\\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}.$$

While you might be used to using D to represent diagonal matrices from real linear algebra, Λ is used just the same. We also like to use the notation $\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ to represent diagonal matrices. A key advantage of using diagonal matrices is that matrix algebra operations are exceptionally straightforward.

Theorem 7.44. Given an $n \times n$ diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and an $m \times n$ matrix A partitioned into column vectors $A = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n\end{bmatrix}$, the matrix product $A\Lambda$ is obtained by multiplying each column of A by the corresponding λ_i in Λ:

$$A\Lambda = \begin{bmatrix}\lambda_1\mathbf{c}_1 & \lambda_2\mathbf{c}_2 & \cdots & \lambda_n\mathbf{c}_n\end{bmatrix}.$$

Similarly, for an $n \times p$ matrix B partitioned into row vectors $B = \begin{bmatrix}\mathbf{r}_1\\ \mathbf{r}_2\\ \vdots\\ \mathbf{r}_n\end{bmatrix}$, the matrix product ΛB is obtained by multiplying each row vector of B by the corresponding λ_i in Λ:

$$\Lambda B = \begin{bmatrix}\lambda_1\mathbf{r}_1\\ \lambda_2\mathbf{r}_2\\ \vdots\\ \lambda_n\mathbf{r}_n\end{bmatrix}.$$

Theorem 7.45. A diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is invertible if and only if $\lambda_i \neq 0$ for all $i = 1, 2, \ldots, n$. The inverse of such a matrix is given by:

$$\Lambda^{-1} = \operatorname{diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n}\right) = \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{bmatrix}.$$

Theorem 7.46. Let $\{|0\rangle, |1\rangle, \ldots, |d-1\rangle\}$ be the computational basis of $\mathbb{C}^d$. A general diagonal matrix $\Lambda = \operatorname{diag}(k_0, k_1, \ldots, k_{d-1})$ can be expressed as

$$\begin{bmatrix} k_0 & 0 & \cdots & 0 \\ 0 & k_1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & k_{d-1} \end{bmatrix} = k_0 |0\rangle\langle 0| + k_1 |1\rangle\langle 1| + \cdots + k_{d-1} |d-1\rangle\langle d-1|.$$

Definition 7.47 (Matrix Power)**.** For a square matrix A, non-negative integer powers are defined as follows:

$$A^0 = I, \qquad A^n = \underbrace{AA\cdots A}_{n \text{ factors}}.$$

For invertible matrices, negative integer powers can also be defined:

Definition 7.48 (Negative Power)**.** For an invertible matrix A, negative integer powers are defined as

$$A^{-n} = \underbrace{A^{-1}A^{-1}\cdots A^{-1}}_{n \text{ factors}}.$$

The two above definitions yields the following properties:

Theorem 7.49. For any matrix A, the following properties hold:

$$A^p A^q = A^{p+q}, \qquad (A^p)^q = A^{pq}$$

provided all powerers involved are defined.

A special type of matrix related to matrix powers are idempotent matrices, derived from the latin words "idem," meaning "the same," and "potent," meaning "power."

Definition 7.50 (Idempotent Matrix)**.** A square matrix A is called idempotent if $A^2 = A$.

Idenpotent matrices frequently arise in the context of projection matrices.

Definition 7.51 (Projection Matrix)**.** A projection matrix onto a $n \times 1$ vector $|\psi\rangle$ is defined as:

$$P_\psi = |\psi\rangle\langle\psi|,$$

such that P_ψ projects any $n \times 1$ vector onto $|\psi\rangle$. In this context, $|\psi\rangle$ and $|\phi\rangle$ are normalized vectors.

Proof. Consider the matrix multiplication:

$$P_\psi |\phi\rangle = |\psi\rangle\langle\psi|\phi\rangle = \langle\psi|\phi\rangle |\psi\rangle$$

which precisely follows the definition from the previous section. □

Theorem 7.52. The projection matrix $P_\psi = |\psi\rangle\langle\psi|$ is idempotent.

Proof. To verify that P_ψ is idempotent, we can simply compute P_ψ^2:

$$P_\psi^2 = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| = P_\psi,$$

considering that $\langle\psi|\psi\rangle = 1$ given that they are normalized vectors. □

Conceptually, this follows since repeated powers of P_ψ is simply the same projection over and over again.

Theorem 7.53. Given a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, its d-th power is given by:

$$\Lambda^d = \begin{bmatrix} \lambda_1^d & 0 & \cdots & 0 \\ 0 & \lambda_2^d & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n^d \end{bmatrix},$$

where $d \in \mathbb{Z}$. If d is negative, Λ must be invertible.

Another interesting note before the end of this subsection is that we can also have polynomials for matrices.

Definition 7.54 (Matrix Polynomial)**.** Given a polynomial function

$$p(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n,$$

the polynomial function of a square matrix A is defined as:

$$p(A) = c_0 I + c_1 A + c_2 A^2 + \cdots + c_n A^n,$$

where I is the identity matrix of the same size as A, and $p(A)$ is a square matrix of the same size as A.

Since matrices and their powers are associative under multiplication, matrix polynomials, also following the same rules. For instance, algebraic identities such as

$$1 - x^n = (1 - x)(1 + x + x^2 + \cdots + x^{n-1})$$

translates directly over to any square matrix A such as

$$I - A^n = (I - A)(I + A + A^2 + \cdots + A^{n-1}).$$

Matrix polynomials are a powerful tool for analzying square matrices, particularly in the context of spectral decomposition, diagonalization, and the Cayley-Hamilton theorem, covered in section 11.

7.4 Trace and Determinant