

QCI402 - Mathematical 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 Mathematics; 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 mathematical underpinnings of quantum computing that the course surrounds itself with. Mathematical Foundations for Quantum Computing takes a scaffolding approach designed to efficiently convey the required theoretical understanding of mathematics 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	3
1.1	Summation over a single Variable	3
1.2	Products and other Notations	5
1.3	Summation over Multiple Variables	5
2	Trigonometry	7
2.1	Definitions	7
2.2	Basic Properties and Inverse Functions	7
2.3	Special Angles and Function Values	9
2.4	Trigonometric Identities	10
2.5	The Spherical Coordinate System	11
3	Complex Numbers	13
3.1	Cartesian Form	13
3.2	Exponential Form	14
3.3	Basic Operations	15
3.4	Advanced Operations	17
4	Sets, Groups, and Functions	19
4.1	Sets	19
4.2	Groups	22
4.3	Functions	24
4.4	Common Functions and Asymptotic Behavior	25
5	Vectors and Vector Spaces	30
5.1	Real Vectors and Complex Vectors	30
5.2	Basic Vector Algebra	30
5.3	Vector Spaces, Subspaces, and Span	31
5.4	Linear Independence, Basis, and Dimension	33

6	Inner Product Spaces	35
6.1	Dirac Notation Basics	35
6.2	Norm and Unit Vectors	37
6.3	Complex Inner Product Spaces	38
6.4	Orthogonality and Projection	42
6.5	Orthonormal Bases	45
7	Fundamentals of Matrix Algebra	54

1 Summation and Product Notations

This section primarily focuses on the common notations applied across mathematics 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 $n \rightarrow \infty$, $\frac{1}{2^n} \rightarrow 0$. Consequently, $S_n \rightarrow \frac{1}{1 - \frac{1}{2}} = 2$. A rigorous proof of the second example requires extensive calculus and is not immediately obvious. While any mathematical 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,

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

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

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

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.2 Products and other Notations

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

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

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

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

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, irrespective of the order. The product of two sums can be expanded into a double sum as follows:

$$\begin{aligned} \left(\sum_{i=1}^m a_i \right) \left(\sum_{j=1}^n b_j \right) &= (a_1 + a_2 + \dots + a_m)(b_1 + b_2 + \dots + b_n) \\ &= a_1 b_1 + a_1 b_2 + a_1 b_3 + a_1 b_4 + \dots + a_1 b_n \\ &\quad + a_2 b_1 + a_2 b_2 + a_2 b_3 + a_2 b_4 + \dots + a_2 b_n \\ &\quad + a_3 b_1 + a_3 b_2 + a_3 b_3 + a_3 b_4 + \dots + a_3 b_n \\ &\quad + \dots \\ &\quad + a_m b_1 + a_m b_2 + a_m b_3 + a_m b_4 + \dots + a_m b_n \\ &= \sum_{i=1}^m \sum_{j=1}^n a_i b_j = \sum_{i=1}^m a_i \sum_{j=1}^n b_j \end{aligned}$$

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

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

where the term $\sum_{1 \leq j \leq n} a_{i,j}$ 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 i th element then aggregate while the second notation sums each column starting from the j th 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

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

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 $\sum_{1 \leq i_1 < \dots < i_k \leq n}$ 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 we see that the equation actually simplifies to a subset of the binomial theorem

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

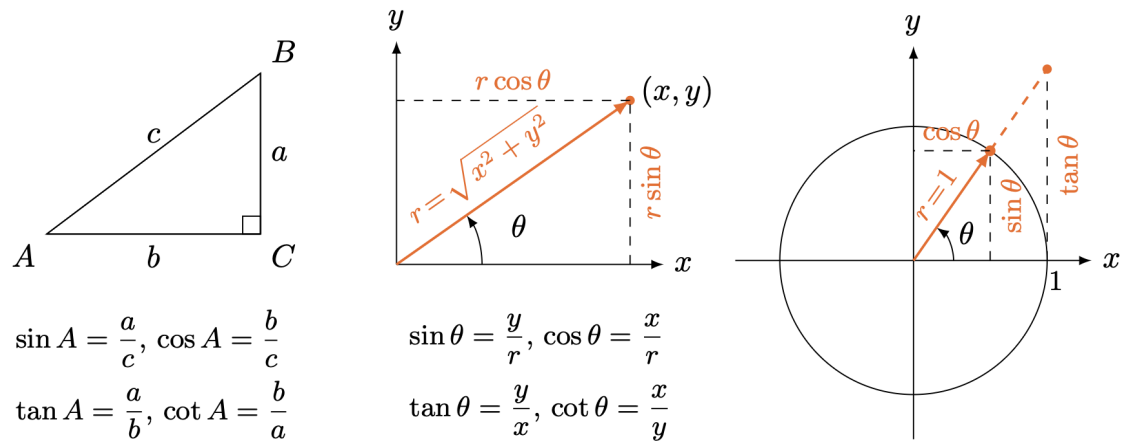
where $\binom{n}{k}$ is the binomial coefficient representing the number of ways to choose k elements from a set of n distinct elements.

2 Trigonometry

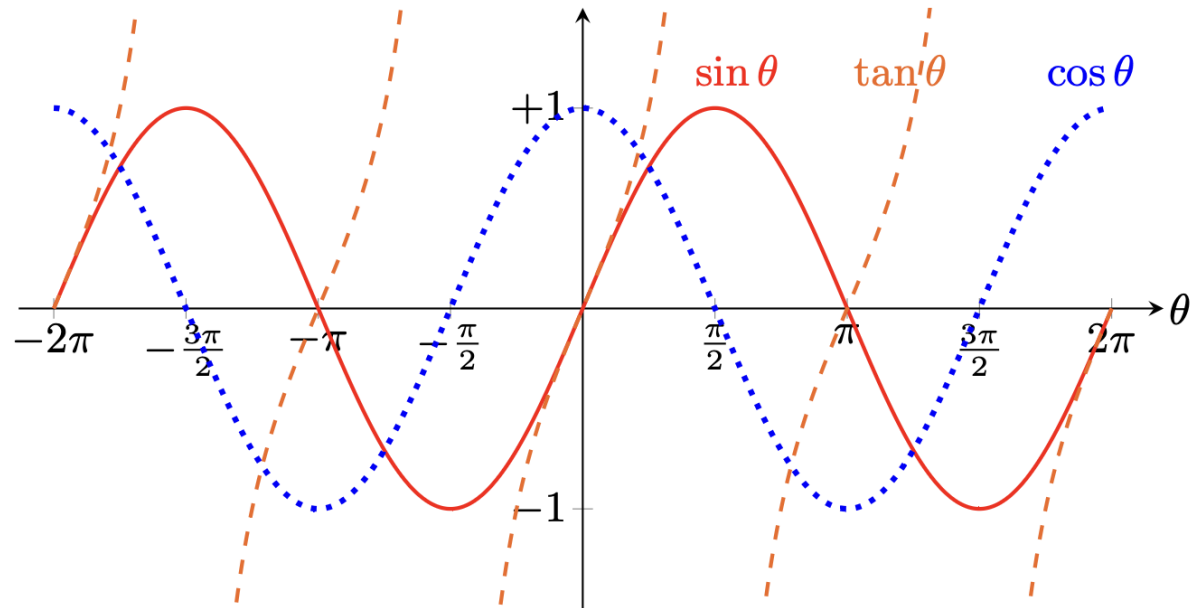
You can't escape this.

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 mathematical detail of how to prove these items. However, we still have to go over this. Don't ask me why.



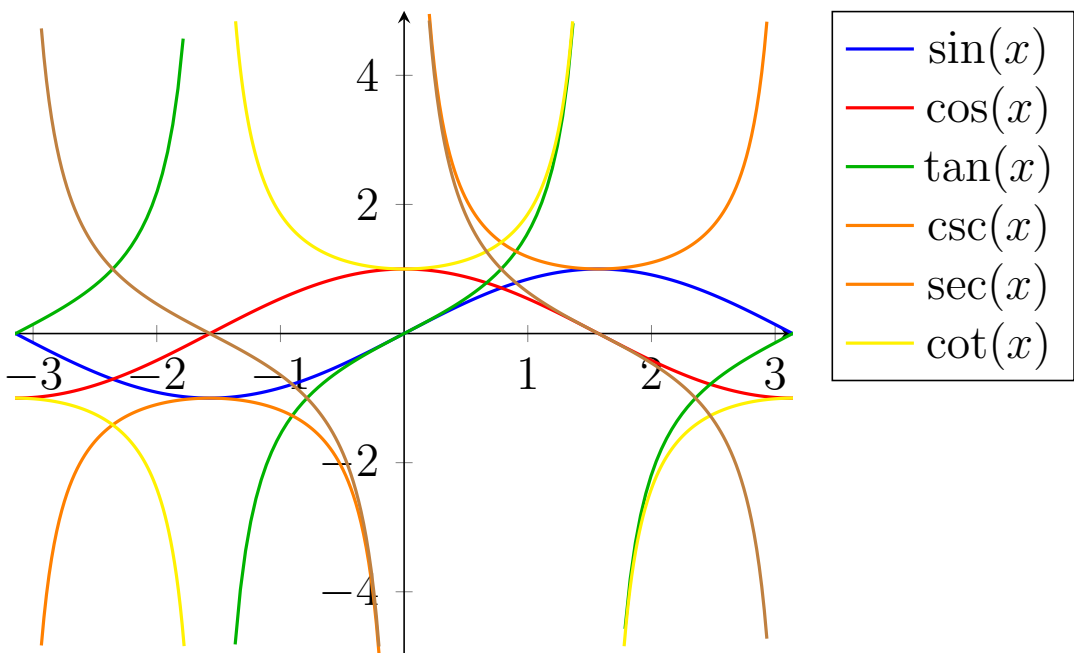
The functions of trigonometric functions can also be plotted out as follows:



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	$[-1, 1]$	$[-1, 1]$	$(-\infty, \infty)$	$(-\infty, -1] \cup [1, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
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

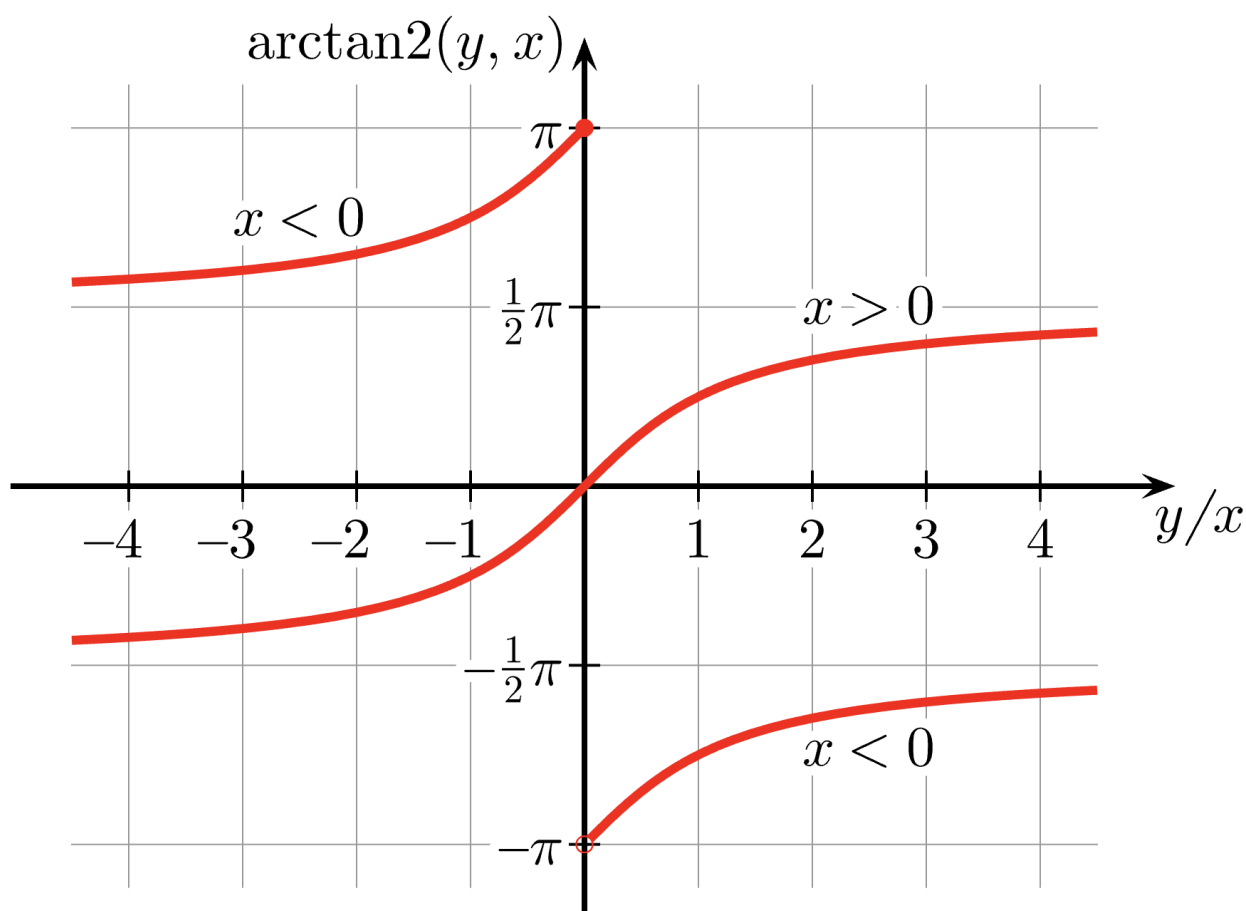
$$\begin{array}{lll}
 \sin(-\theta) = -\sin(\theta) & \sin(\pi - \theta) = \sin(\theta) & \sin(\pi + \theta) = -\sin(\theta) \\
 \cos(-\theta) = \cos(\theta) & \cos(\pi - \theta) = -\cos(\theta) & \cos(\pi + \theta) = -\cos(\theta) \\
 \tan(-\theta) = -\tan(\theta) & \tan(\pi - \theta) = -\tan(\theta)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

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

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

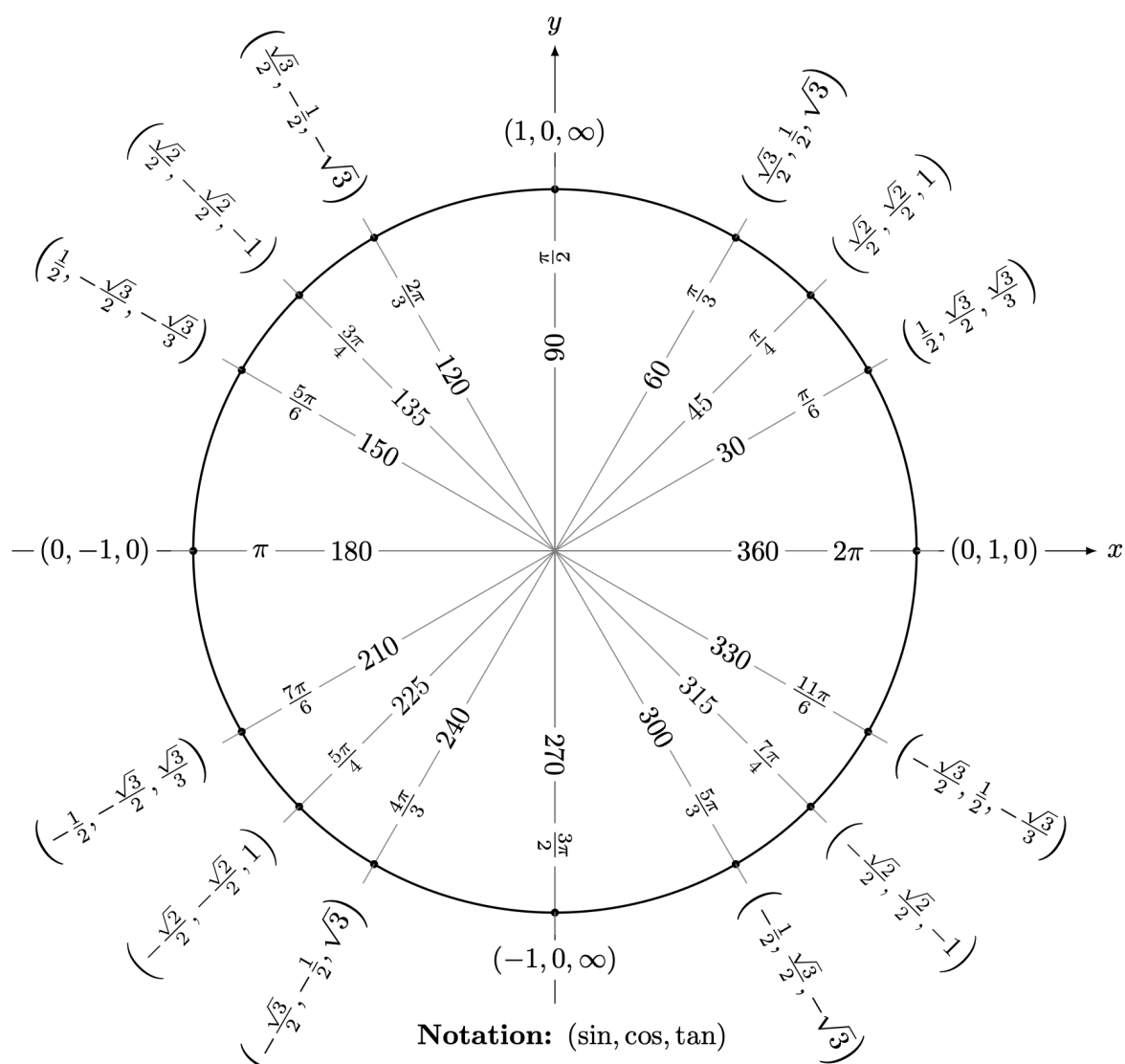
But also some interesting extensions of the commonly known inverse functions, with the example being arctan2, a function that effectively doubles the domain of the function while preserving its properties for the purpose of, say, converting values from cartesian to spherical coordinates for the azimuthal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

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



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

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

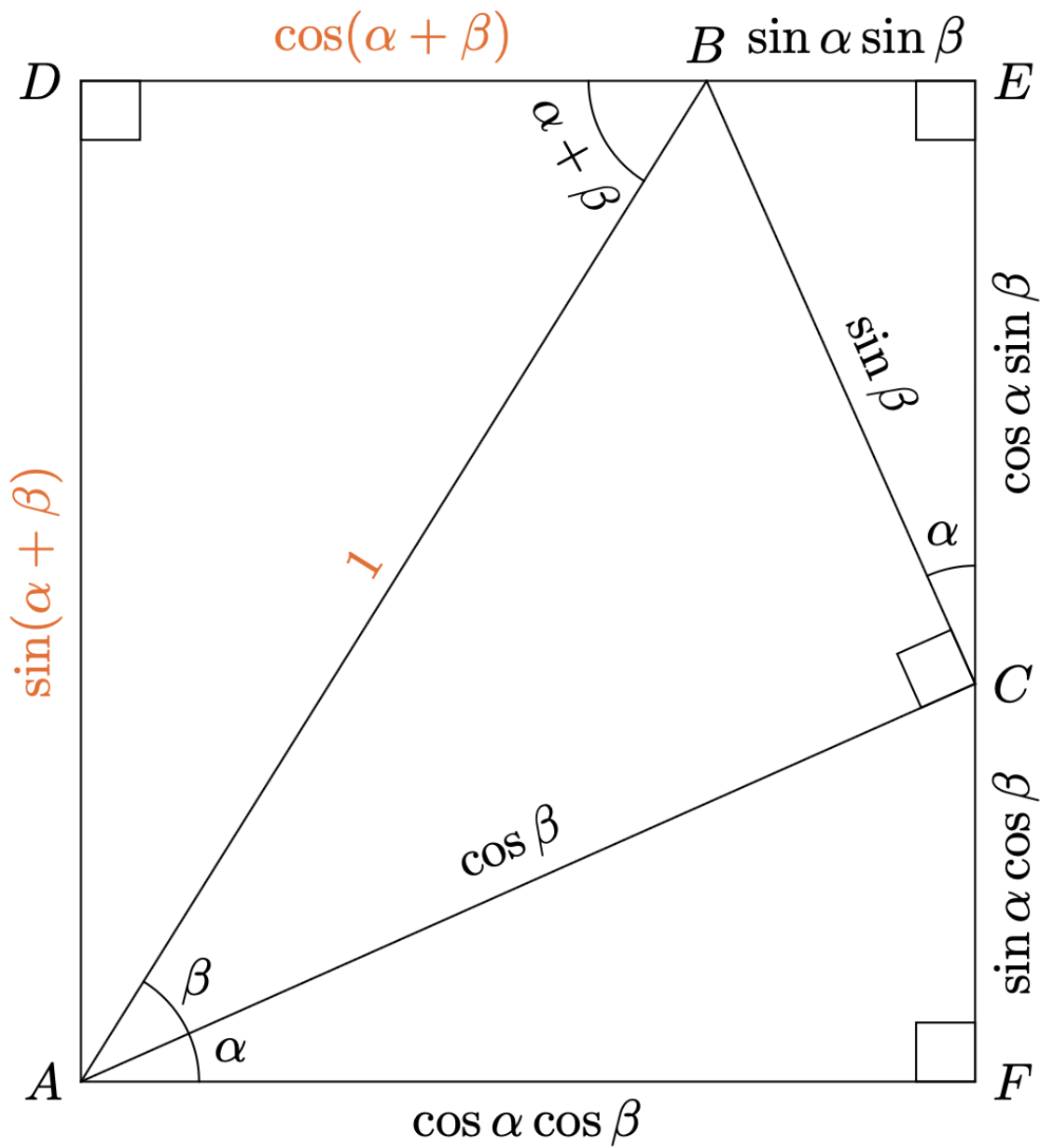
Product-to-Sum Identities

$$\begin{aligned}\sin \alpha \sin \beta &= \frac{1}{2} \left[\cos(\alpha - \beta) - \cos(\alpha + \beta) \right] \\ \cos \alpha \cos \beta &= \frac{1}{2} \left[\cos(\alpha - \beta) + \cos(\alpha + \beta) \right] \\ \sin \alpha \cos \beta &= \frac{1}{2} \left[\sin(\alpha + \beta) + \sin(\alpha - \beta) \right]\end{aligned}$$

Sum-to-Product Identities

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

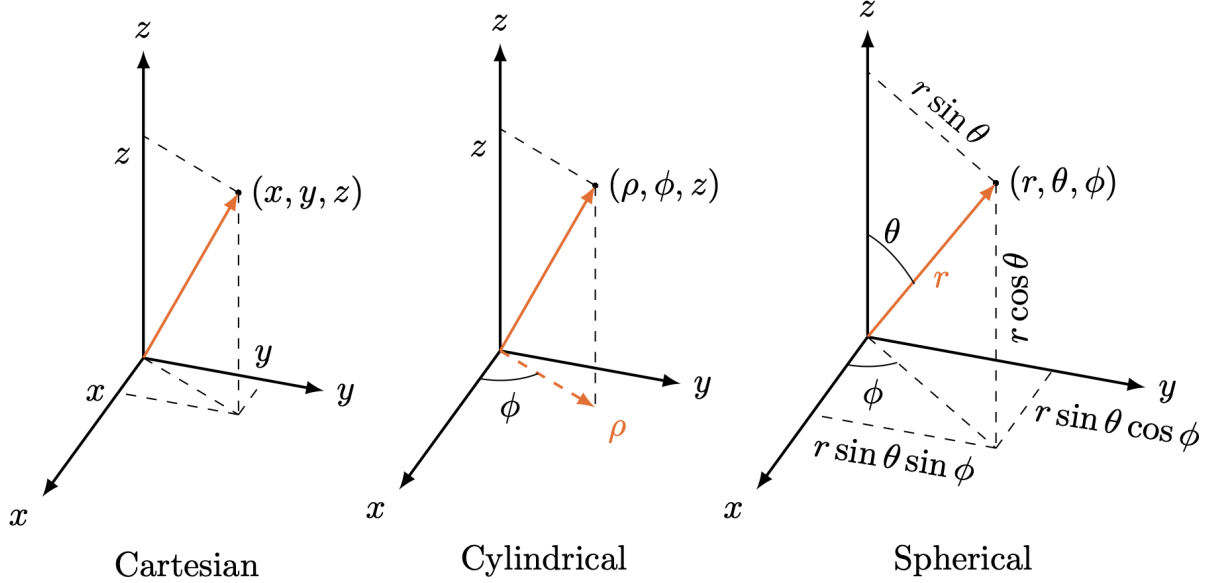
A rather nice photo to sum the seccion up is by relating the angles to each other using the following image



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

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

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

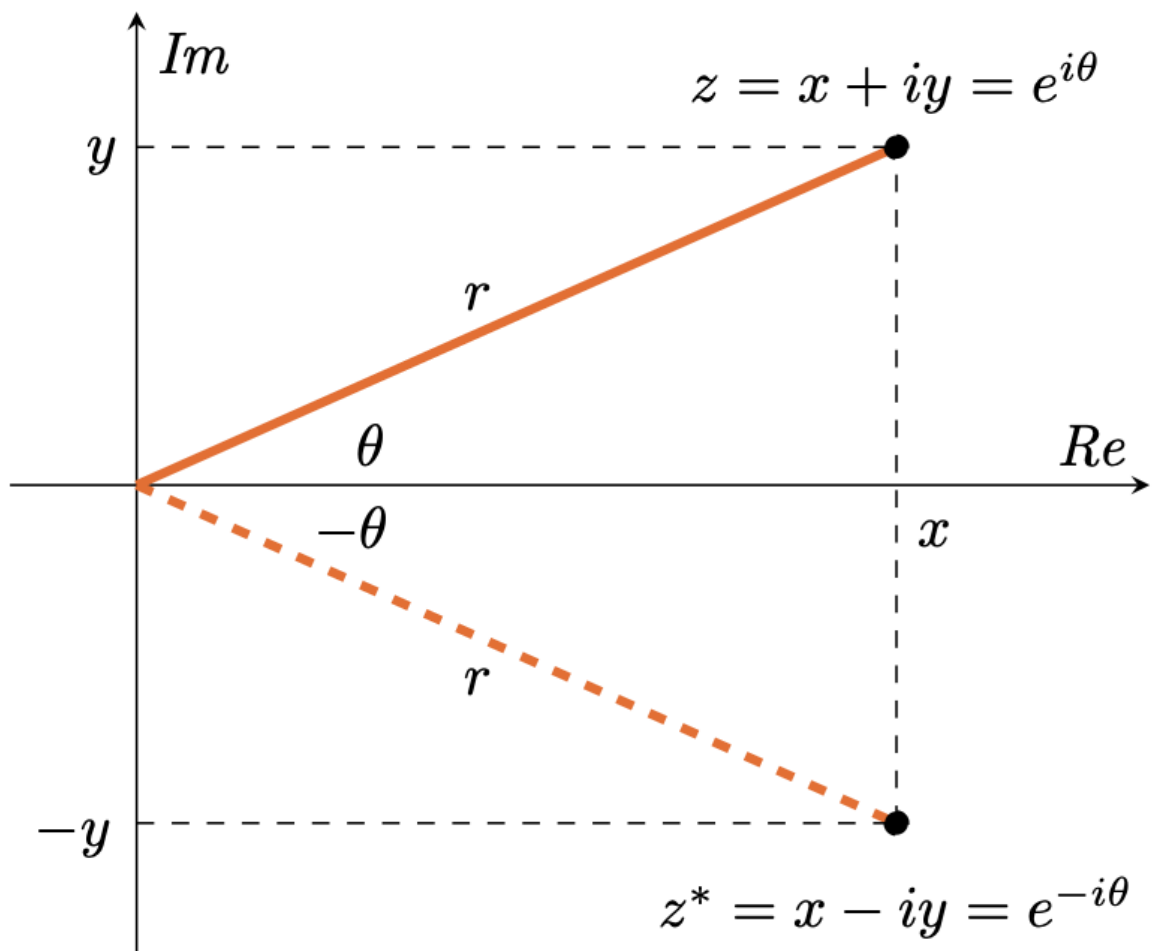
We consider numbers to be complex when they compose of a real and imaginary part, and they are not only fundamental to a complete understanding of algebra and mathematics as a whole, but also form the backbone of quantum mechanics, and, by extension, quantum computing. Mastering complex numbers is like 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

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

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 algebraic operations, which include addition, subtraction, 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

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

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 number, 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 given 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

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

Conversely, the conversion between cartesian and polar are

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

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:

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

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

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

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

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

which sums up essentially what mathematicians 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 L^AT_EX to 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	$ z = \sqrt{zz^*} = \sqrt{x^2 + y^2}$	$ z = 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 $z^s = r^s e^{is\theta}$.

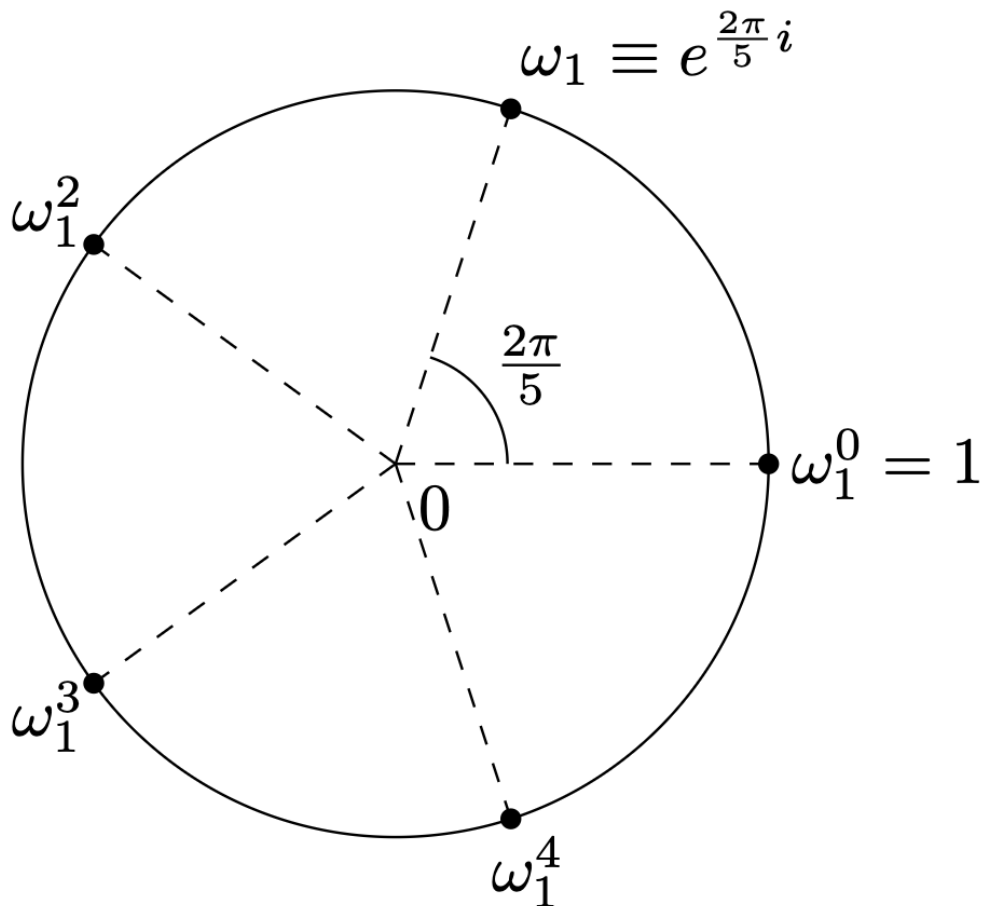
In particular, an application of this theorem is where we describe the roots of unity. Any root of unity can be described by a power of the first root of unity,

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

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

$$\omega_k = e^{\frac{2\pi i}{n}k} = \omega_1^k \quad k = 1, 2, \dots, n-1$$

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 conveniently proven by applying the formula for summing a geometric sequence to the summation. From this, we can conveniently derive a useful mathematical 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 $\delta_{k-l \pmod N}$ 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 N th root of unity (satisfying $\omega^N = 1$), we have the derivation

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

where we used the fact that $\omega^{nN} = (\omega^N)^n = 1^n = 1$ 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

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

Example. Evaluating

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

gives us

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

Example. Evaluating

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

gives us

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

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

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

from this, we can derive that

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

and that

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

Another way of expressing this is by saying

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

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 mathematical concepts that make up the backbone of many mathematical fields, including those relevant to quantum computing.

4.1 Sets

The concepts of sets are fundamental to many areas of mathematics. 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 $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ 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 \bar{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 $\bar{A} = \{1, 3, 5, 7, 9\}$, 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 $\{A_1, A_2, \dots, A_n\}$ is a partition of set A , and $B \subseteq A$, then $\{A_1 \cap B, A_2 \cap B, \dots, A_n \cap B\}$ 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:

$$A \cup \emptyset = A$$

$$A \cap U = A.$$

Domination Laws:

$$A \cup U = U$$

$$A \cap \emptyset = \emptyset.$$

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:

$$A \cup \bar{A} = U$$

$$A \cap \bar{A} = \emptyset$$

$$\overline{\bar{A}} = A.$$

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:

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

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

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 mathematics that build upon sets with additional operations in algebra. These structures are ubiquitous in mathematics 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 mathematical contexts, it is common to omit the symbol \cdot or $*$ for the group operation and simply write $x \cdot y$ as xy . Some common examples include the group 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 far are all abelian groups, there are some of groups that are not abelian. For example, 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. **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 preserve distance and the orientation of objects in question. SO(2) corresponds to a circle, and SO(3) to a sphere.
2. **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. **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.

$$|G| \equiv 0 \pmod{|H|}$$

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 (Partition 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 implies 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 $0 + 2\mathbb{Z} = \{\dots, -4, -2, 0, 2, 4, \dots\}$ represents the even integers.
- The coset $1 + 2\mathbb{Z} = \{\dots, -3, -1, 1, 3, 5, \dots\}$ 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 $f : A \rightarrow B$, 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 .

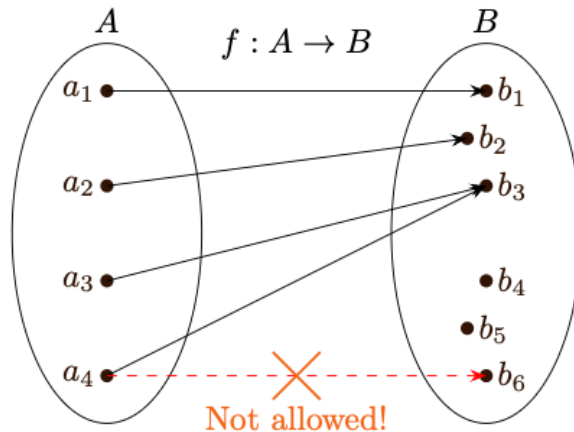
Example. $f(x) = x^2$ is a function $f : \mathbb{R} \rightarrow \mathbb{R}$. $r = \sqrt{x^2 + y^2}$ is a function $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$.

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

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



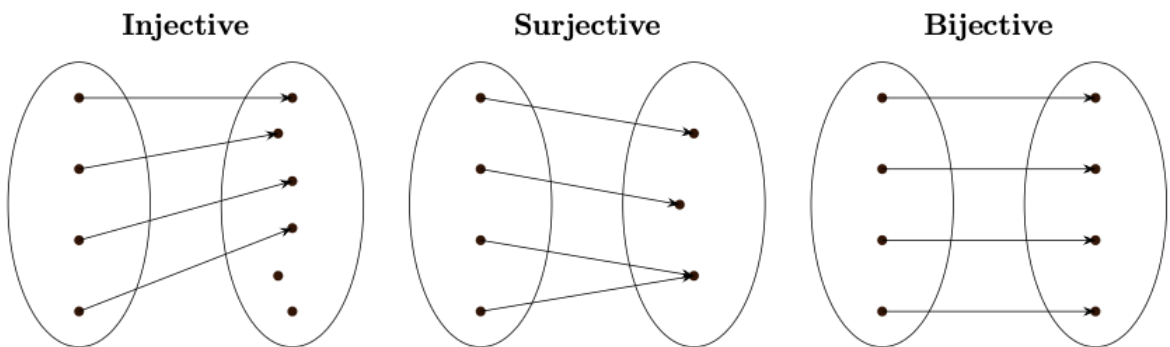
There are also a few common function types that we will define:

Definition 4.32 (Injective). A function $f : A \rightarrow B$ 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 $f : A \rightarrow B$ is surjective if the whole codomain is covered, meaning that $\forall b \in B, \exists a \in A$ 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 $f : A \rightarrow B$, its domain A and codomain B , we can define an inverse function $f^{-1} : B \rightarrow A$ such that $\forall b \in B, f^{-1}(b) = a$ if and only if $f(a) = b$.

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

4.4 Common Functions and Asymptotic Behavior

This section will be a brief review of real functions ($f : \mathbb{R} \rightarrow \mathbb{R}$) 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:

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

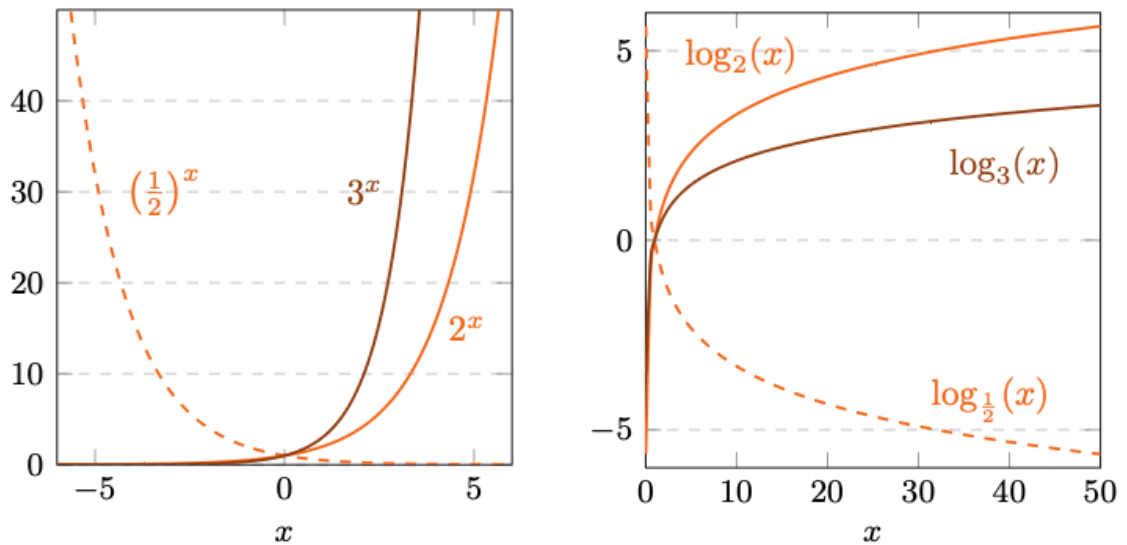
and the product is:

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

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: $\log_b(x) = \log_a(x) / \log_a(b)$

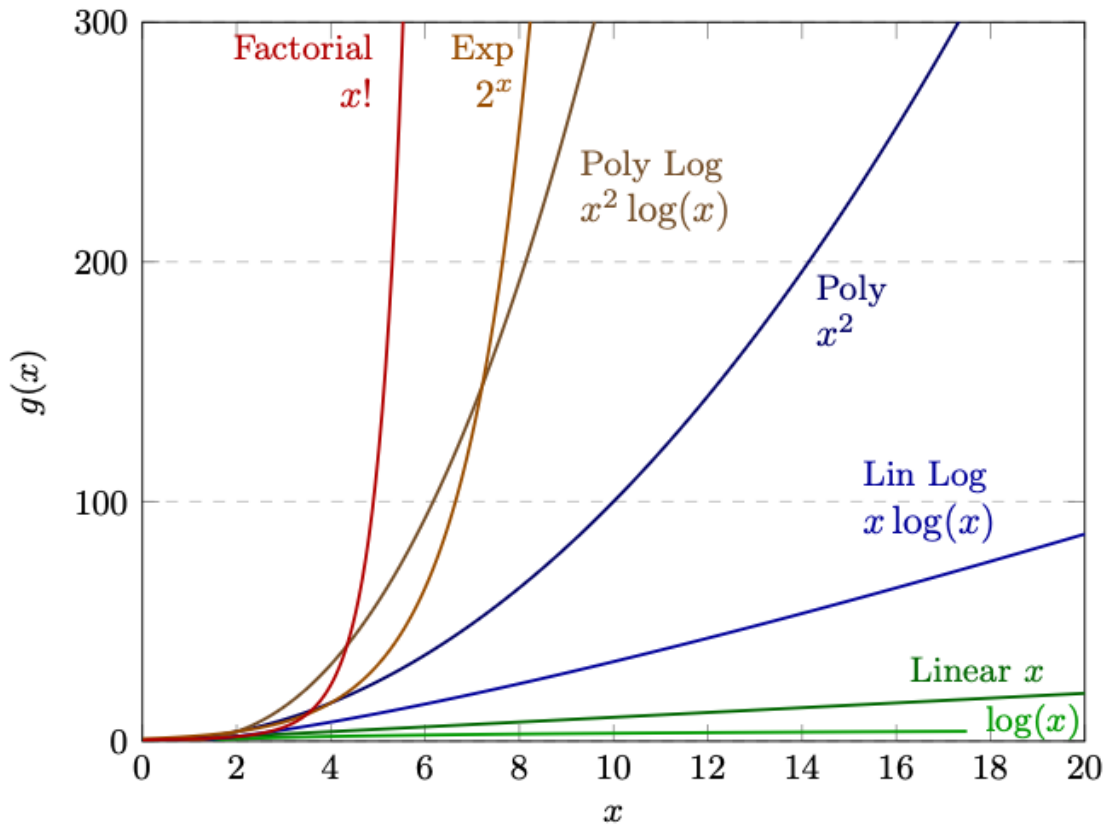
Scaling behavior, especially in the context of data structures and algorithm efficiency, is a significant topic for any computing related fields. It concerns itself with the evolution of function curves as the response becomes significantly large. We use the *Big O notation* to describe the behavior of a function $f(x)$, $x \rightarrow \infty$. 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 $x \rightarrow \infty$, 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 mathematical 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 concepts 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, \dots, 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.

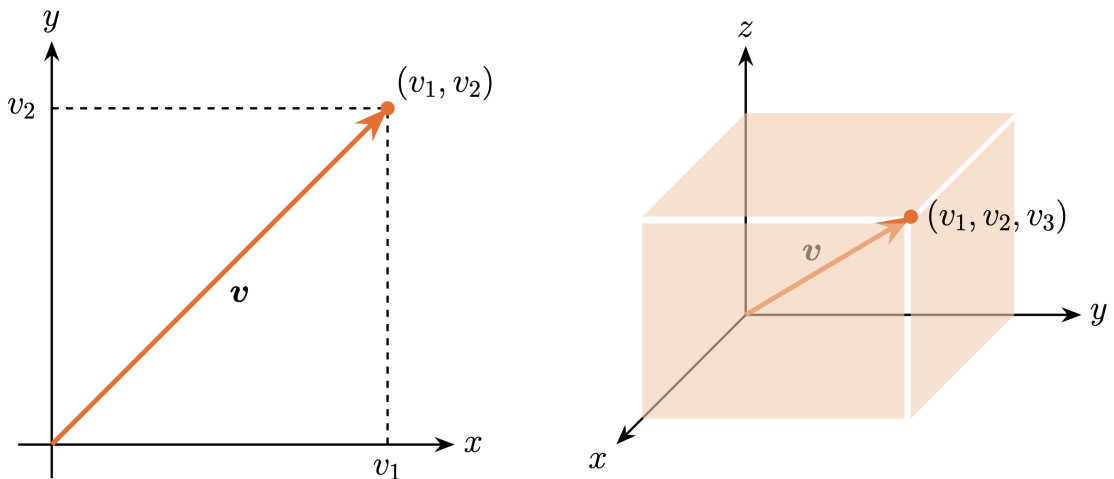
Now in the context of quantum computing, n is typically finite as we are usually considering finite quantum systems. A real vector is in \mathbb{R}^n , while a complex vector is in \mathbb{C}^n . Much like $\{\mathbf{0}\}$, we will begin with the definition of the zero vector.

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.

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

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 possess magnitude and direction, conveniently 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 , $n > 3$ 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 $\mathbf{v} = (v_1, v_2, \dots, v_n)$ is denoted as \mathbf{v}^* and is given by:

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

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:

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

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

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

$$\begin{aligned}\mathbf{u} + \mathbf{v} &= \mathbf{v} + \mathbf{u} \\ \mathbf{u} + (\mathbf{v} + \mathbf{w}) &= (\mathbf{u} + \mathbf{v}) + \mathbf{w} \\ \mathbf{v} + \mathbf{0} &= \mathbf{v}\end{aligned}$$

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:

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

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:

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

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 $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ in \mathbb{R}^n or \mathbb{C}^n , we say that \mathbf{w} is a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ if \mathbf{w} can be expressed as:

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \dots + k_r\mathbf{v}_r = \sum_{i=1}^r k_i\mathbf{v}_i$$

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 objects 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. \square

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

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

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

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 corollary 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. Mathematically:

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

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

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 $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ 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 . \square

The formal definition also is as follows:

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

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

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 coordinate 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 mathematically obvious why we must choose these two 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 $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ 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, \dots, c_n not all zero such that

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

If not such scalars exist, 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 $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ 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. \square

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, \dots, 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

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

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 possess 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 mathematically 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 mathematical 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 $a_{11}, a_{12}, \dots, a_{mn}$ 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:

$$[v_1 \quad v_2 \quad \cdots \quad v_m].$$

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 $v = (v_1, v_2, \dots, v_n)$ 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 common 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 $j, k \in \{0, 1, 2, \dots, n-1\}$ to denote the computational basis vectors in \mathbb{C}^n .

There are also certain symbols reserved for specific vectors of significance, much like how we reserve π and e in common arithmetic. 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:

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

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:

$$v^\dagger \equiv (v^*)^T = (v^T)^* = [v_1^* \quad v_2^* \quad \dots \quad v_n^*].$$

We have a convenient 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 components. In three dimensional space, it means adding the squared value of the z component as well. This allows us to generalize to the n th 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:

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

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

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

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 defined as:

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \dots + u_nv_n$$

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

$$\begin{aligned} \|\mathbf{v}\|^2 &= \mathbf{v} \cdot \mathbf{v} \\ \|\mathbf{u} - \mathbf{v}\|^2 &= (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) \end{aligned}$$

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

At the same time, we have a few more intuitive and convenient 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:

$$\begin{array}{ll} \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} & \text{(Commutative Law)} \\ \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} & \text{(Left Distributive Law)} \\ (\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} & \text{(Right Distributive Law)} \\ (k\mathbf{u}) \cdot \mathbf{v} = k(\mathbf{u} \cdot \mathbf{v}) & \text{(Left Homogeneity Law)} \\ \mathbf{u} \cdot (k\mathbf{v}) = k(\mathbf{u} \cdot \mathbf{v}) & \text{(Right Homogeneity Law)} \end{array}$$

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:

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

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} , its norm can be expressed as:

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

Complex inner product spaces are complex vector spaces "equipped" with a complex inner product. 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: $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$. The inner product, for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $k \in \mathbb{C}$, 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. $\langle \mathbf{v} \cdot \mathbf{v} \rangle = 0 \iff \mathbf{v} = \mathbf{0}$

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.
$$\mathbf{v} \cdot \mathbf{v} = \sum_{i=1}^n v_i^* v_i = \sum_{i=1}^n |v_i|^2 \geq 0, \quad \forall \mathbf{v}$$
5.
$$\forall i \in [1, n], \quad |v_i| = 0 \implies v_i = 0 \implies \mathbf{v} = \mathbf{0}$$

□

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

Theorem 6.22. Given a complex inner product space V with an inner product $\langle \cdot, \cdot \rangle$, with $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $k \in \mathbb{C}$, 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) $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$. It must, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $k \in \mathbb{R}$, 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:

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

Which, conveniently, 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 :

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

their matrix product is defined as:

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

Now, noting that we can simplify the product between 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 convenience, 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:

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

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 $\langle\phi|\phi\rangle = \|\phi\|^2 = 1$ \square

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 mathematical 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}^∞

We first start with \mathbb{C}^∞ 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:

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

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

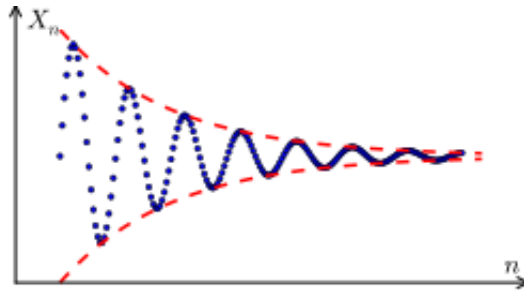
Building on the previous concepts of \mathbb{C}^n and functional spaces, we are ready to define Hilbert spaces. These are generalization of inner product spaces that is *complete*.

Definition 6.30 (Hilbert Space). A Hilbert 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 $\{s_n\} = (s_1, s_2, \dots)$ is called a **Cauchy Sequence** if:

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



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.

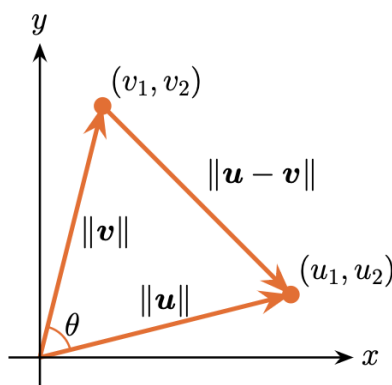
When we talk about the dot product, it is closely related to the two. Applying the law of cosines:

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

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 orthogonal, 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. Therefore, we have

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

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

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

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

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

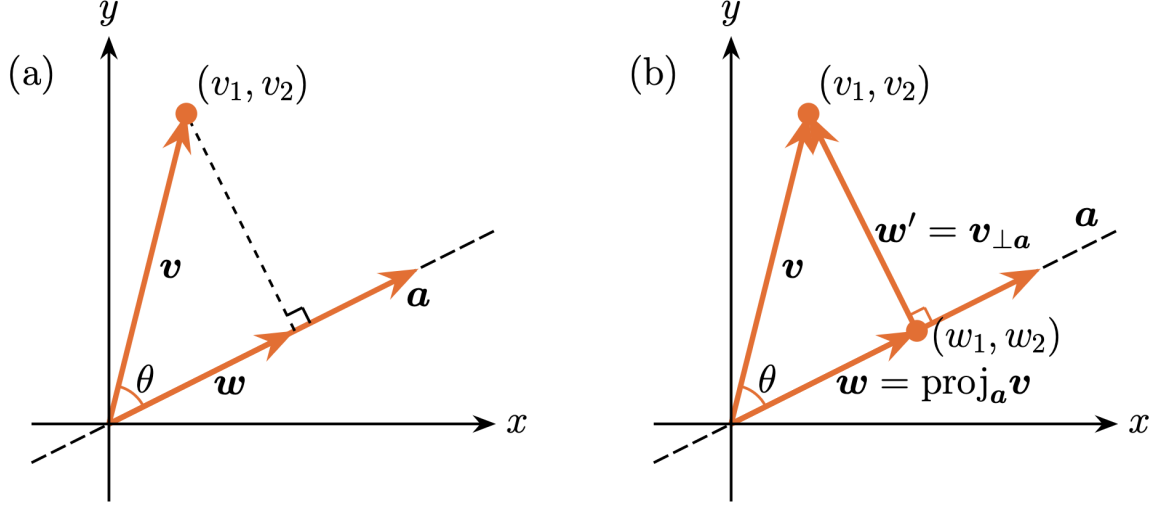
Implying that λ is of the form $e^{i\phi}$. □

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:

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

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

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

Proof. Since \mathbf{w} is in the span of the vector \mathbf{a} , it can be written 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:

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

□

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

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

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 observable 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 orthogonal and normalized, are an intuitive and fundamental concept for both quantum mechanics and quantum computing, simplifying the representation of quantum states and illustrating key concepts such as superposition, entanglement, state evolution, and measurement. Here, Dirac notation is the norm.

Definition 6.43 (Orthonormal Basis). A basis $S = \{|b_1\rangle, |b_2\rangle, \dots, |b_n\rangle\}$ 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:

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

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

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:

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

When expressing any unit vector in terms of orthonormal bases, 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:

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

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:

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

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

$$= \sum_{i=1}^n c_i^* c_i 1 + \sum_{i \neq j} c_i^* c_j 0 \quad (6)$$

$$= \sum_{i=1}^n |c_i|^2 \quad (7)$$

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

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

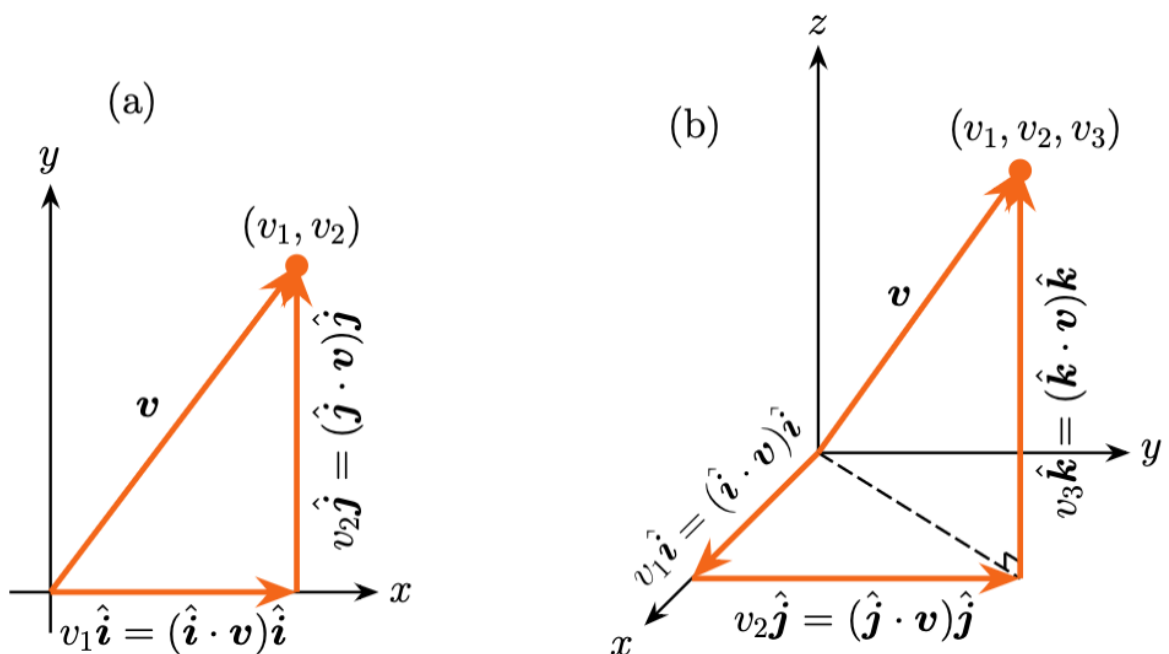
$$\begin{aligned} \langle b_i | \psi \rangle &= \langle b_i | (c_1 |b_1\rangle + c_2 |b_2\rangle + \cdots + c_n |b_n\rangle) \\ &= c_1 \langle b_i | b_1 \rangle + c_2 \langle b_i | b_2 \rangle + \cdots + c_n \langle b_i | b_n \rangle \\ &= c_1 \cdot 1 + \sum_{i \neq j} c_j \cdot 0 \\ &= c_i \end{aligned}$$

\square

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 this task, which is especially beneficial when working with degenerate eigenvalues.

In order to fully understand this process, 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 its 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. \square

Theorem 6.48 (Orthogonal Basis). Take $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ 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 constitutes 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:

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

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

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

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. \square

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 :

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

The Gram Schmidt process,

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

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

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

Proof. In this proof, we utilize the complex dot product as the inner product. Given that $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a basis for V , and hence $\dim(V) = n$, we aim to demonstrate that $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ 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. Mathematically:

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

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

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

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

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, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k+1}$ 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:

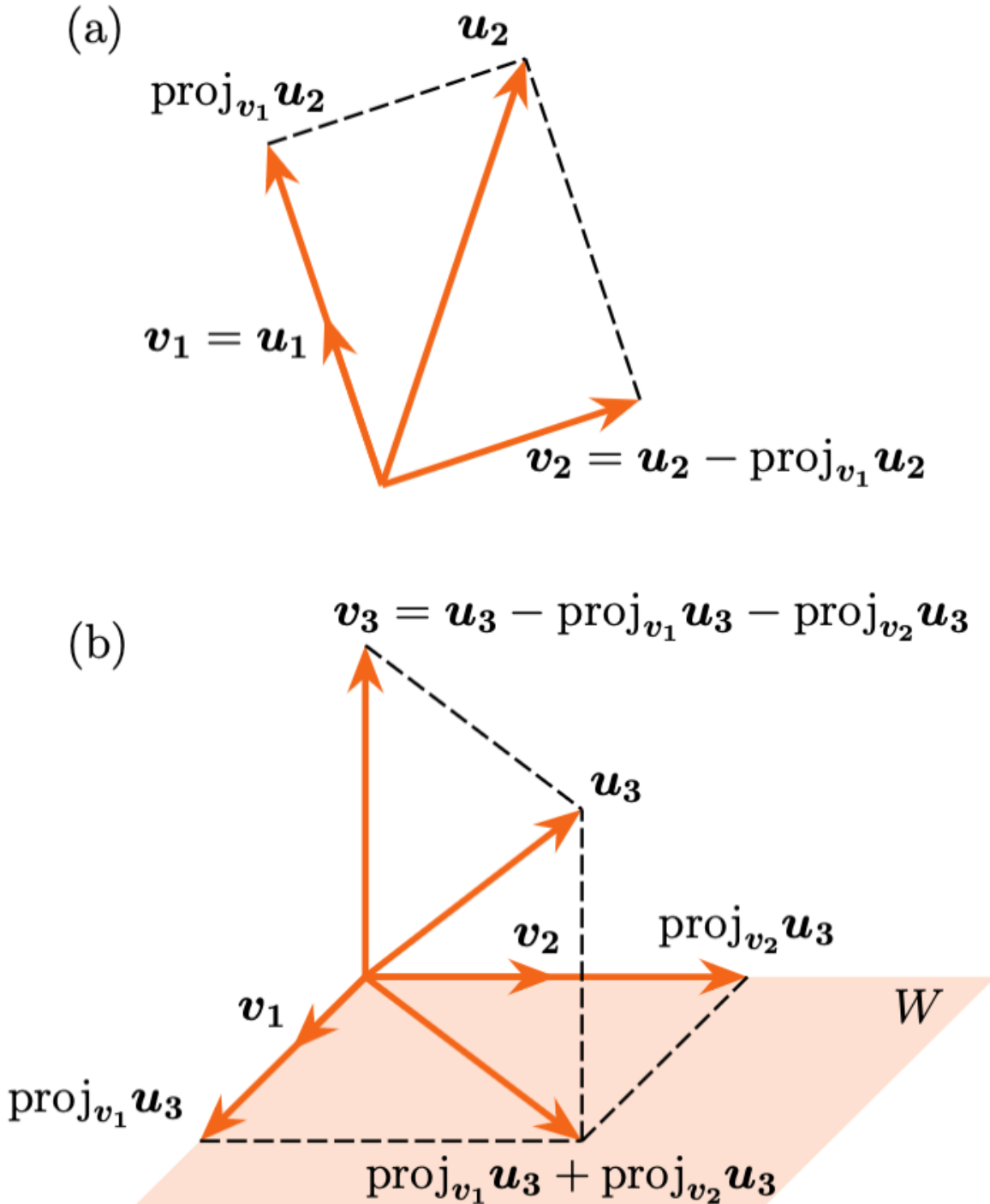
Definition 6.51 (Projection). Let V be an inner product space and $\mathbf{u} \in V$, and take $W \subseteq V$. If $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is an orthogonal basis for W , then we define $\text{proj}_W \mathbf{u}$ as:

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

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 $W_{n-1} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1}\}$,

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

Interestingly, there have been several proposed forms for the Gram-Schmidt process in Python, a few examples we will include below:

Classical Implementation

The original implementation by **iizukak** provides a straightforward translation of the Gram-Schmidt algorithm, directly computing projections and subtracting them iteratively.

However, this approach uses Python's `map` function and doesn't leverage NumPy's vectorization capabilities, making it slower for larger matrices:

```

1 import numpy
2
3 def gs_coefficient(v1, v2):
4     return numpy.dot(v2, v1) / numpy.dot(v1, v1)
5
6 def multiply(coefficient, v):
7     return map((lambda x : x * coefficient), v)
8
9 def proj(v1, v2):
10    return multiply(gs_coefficient(v1, v2), v1)
11
12 def gs(X):
13    Y = []
14    for i in range(len(X)):
15        temp_vec = X[i]
16        for inY in Y:
17            proj_vec = proj(inY, X[i])
18            temp_vec = map(lambda x, y : x - y, temp_vec, proj_vec)
19        Y.append(temp_vec)
20    return Y

```

Modified Gram-Schmidt (Clear Implementation)

`mreagan`'s implementation is notable for its clarity and adherence to the textbook algorithm. It processes each column sequentially, subtracting projections onto previously computed orthogonal vectors. This is the **modified Gram-Schmidt** algorithm, which offers better numerical stability than the classical version by reducing floating-point error accumulation:

```

1 import numpy as np
2
3 def gram_schmidt(A):
4     # Get the number of vectors.
5     n = A.shape[1]
6     for j in range(n):
7         # To orthogonalize the vector in column j with respect to the
8         # previous vectors, subtract from it its projection onto
9         # each of the previous vectors.
10        for k in range(j):
11            A[:, j] -= np.dot(A[:, k], A[:, j]) * A[:, k]
12        A[:, j] = A[:, j] / np.linalg.norm(A[:, j])
13    return A

```

Vectorized Implementation

`aditya0by0` improved upon the classical approach with a vectorized version that computes projections using matrix operations rather than loops. The key advantage is computing all projection coefficients simultaneously, which significantly improves performance for large matrices:

```

1 def gram_schmidt(A, norm=True, row_vect=False):
2
3     if row_vect:
4         A = A.T
5
6     no_of_vectors = A.shape[1]
7     G = A[:, 0:1].copy()
8
9     for i in range(1, no_of_vectors):
10        numerator = A[:, i].dot(G)
11        denominator = np.diag(np.dot(G.T, G))
12        weights = np.squeeze(numerator / denominator)
13
14        projected_vector = np.sum(weights * G, axis=1, keepdims=True)
15        orthogonalized_vector = A[:, i:i+1] - projected_vector
16        G = np.hstack((G, orthogonalized_vector))
17
18    if norm:
19        G = G / np.linalg.norm(G, axis=0)
20
21    if row_vect:

```

```

22         return G.T
23
24     return G

```

Handling Singular Matrices

JJGO addressed an important practical issue: handling linearly dependent vectors. This implementation checks if a vector is already in the span of previous vectors (using a small tolerance like $1e-10$) and skips it if so:

```

1 def gram_schmidt(vectors):
2     basis = []
3     for v in vectors:
4         w = v - np.sum(np.dot(v, b) * b for b in basis)
5         if (w > 1e-10).any():
6             basis.append(w / np.linalg.norm(w))
7     return np.array(basis)

```

NumPy QR Factorization

jakevdp pointed out that NumPy’s built-in `qr()` function provides the fastest and most numerically stable approach. The QR decomposition inherently performs Gram-Schmidt orthogonalization (using a modified version), but with optimized C implementations:

```

1 import numpy as np
2
3 def gram_schmidt_columns(X):
4     Q, R = np.linalg.qr(X)
5     return Q

```

Modified Gram-Schmidt for Complex Vectors

For complex inner product spaces (relevant to quantum computing), **scrapy** provided a modified implementation using `np.vdot()` for proper complex conjugation:

```

1 import numpy as np
2
3 def modifiedGramSchmidt(A):
4     dim = A.shape[0]
5     Q = np.zeros(A.shape, dtype=A.dtype)
6     for j in range(0, dim):
7         q = A[:, j]
8         for i in range(0, j):
9             rij = np.vdot(Q[:, i], q)
10            q = q - rij * Q[:, i]
11            rjj = np.linalg.norm(q, ord=2)
12            if np.isclose(rjj, 0.0):
13                raise ValueError
14            else:
15                Q[:, j] = q / rjj
16     return Q

```

Custom Inner Product Implementation

For non-Euclidean inner product spaces, the algorithm can be adapted by replacing the dot product with a custom inner product function:

```

1 import numpy as np
2
3 def inner_product(x, y):
4     # Example: weighted inner product
5     return x[0]*y[0] + 2*x[1]*y[1] + x[2]*y[2]
6
7 def gram_schmidt(V):
8     U = []
9     for i in range(len(V)):
10        u = V[i]
11        for j in range(i):
12            proj = (inner_product(V[i], U[j]) /
13                    inner_product(U[j], U[j])) * U[j]
14            u = u - proj
15        U.append(u / ((inner_product(u, u))**0.5))
16     return np.array(U)

```

Performance Comparison

According to **victorhcm**’s testing, the NumPy QR approach and manual Gram-Schmidt implementations produce identical results (up to sign differences in the basis vectors). How-

ever, the QR factorization is typically 2-3 times faster and more numerically stable, especially for ill-conditioned matrices.

7 Fundamentals of Matrix Algebra