Lecture Notes on

Quantum Computing and Cryptography

Part I – Linear Algebra and Quantum Mechanics

Academic Year 2025–2026

Dr Marco Paviotti

Department of Computing
University of Kent

Last updated: April 6, 2026

Contents

1 Introduction **3**
1.1 Further Reading . . . 3

2 Linear Algebra **4**
2.1 Vector Spaces . . . 4
2.2 Linear Transformations . . . 5
2.3 Linear combinations, Spans and Linear Independence . . . 7
2.4 The norm of a vector . . . 8
2.5 Rank and Nullity . . . 9
2.5.1 Solutions of Linear Systems . . . 10
2.6 Products . . . 11
2.6.1 Cross Product . . . 11
2.6.2 Tensor Product . . . 12
2.6.3 The dot product . . . 12
2.7 Outer product . . . 13
2.8 Inner product . . . 13
2.9 Diagonalisation and Changes of Basis . . . 15
2.10 Eigen decomposition . . . 16
2.11 Hermitian Matrices and the Spectral Decomposition Theorem . . . 18
2.12 Unitary Matrices . . . 20
2.13 Normal Matrices . . . 21
2.14 Trace of a Matrix . . . 22
2.15 Dirac Notation . . . 23
2.16 Hilbert Spaces (*) . . . 23
2.16.1 Cauchy complete spaces . . . 23

3 Quantum Mechanics **24**
3.1 Quantum States and Quantum Superposition . . . 24
3.2 Quantum Evolution and Quantum Gates . . . 25
3.2.1 Basic Quantum Gates . . . 26
3.2.2 Phase Operations . . . 27
3.3 Measurements . . . 28
3.3.1 Measurement in the Computational Basis . . . 29
3.3.2 Projective Measurements . . . 30
3.3.3 Distinguishing Quantum States . . . 30
3.4 Multiple Quantum Systems . . . 31
3.4.1 The CNOT Gate . . . 31
3.5 Quantum Entanglement . . . 32
3.6 Coherence and Density Matrices . . . 33
3.7 The Schrödinger Equation (*) . . . 35
3.8 The Bloch Sphere (*) . . . 36

A Elements of Set Theory **36**
A.1 Basic Trigonometry . . . 37
A.2 Complex Numbers . . . 39
A.3 The Fundamental Theorem of Algebra . . . 41
A.4 From Complex to Real Linear Combinations . . . 41

1 Introduction

Quantum computing is a paradigm which uses the principles of quantum mechanics to process information. Unlike classical computers, which operate on bits that represent either 0 or 1, quantum computers use quantum bits, or *qubits*. Qubits can exist in a superposition of states, represented mathematically as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where α and β are complex numbers that determine the probabilities of the qubit being measured as 0 or 1. This property allows quantum computers to explore multiple solutions simultaneously, offering a significant advantage for certain computational tasks.

Another key feature of quantum systems is *entanglement*, where qubits become correlated in such a way that the state of one qubit depends on the state of another, regardless of the physical distance between them. This enables highly coordinated operations across qubits, a phenomenon that classical systems cannot replicate. To perform computations, quantum computers use *quantum gates*, such as the Hadamard gate, which creates superposition, and the CNOT gate, which introduces entanglement. These gates manipulate qubits to execute algorithms in fundamentally different ways from classical logic gates.

Quantum computing holds promise for solving problems that are intractable for classical computers. For instance, Shor's algorithm can factorize large numbers exponentially faster than classical methods, posing a challenge to current cryptographic systems. Grover's algorithm, on the other hand, provides a quadratic speedup for searching unsorted databases. Quantum computers are also uniquely suited for simulating quantum systems, making them invaluable for advancing material science, drug discovery, and fundamental physics.

Despite its potential, quantum computing faces significant challenges. Qubits are highly sensitive to their environment, leading to *decoherence*, where quantum states lose their coherence over time. Additionally, quantum systems are prone to errors, necessitating robust error correction methods. Scaling quantum computers to include thousands or millions of qubits also remains a formidable engineering hurdle.

As research and technology progress, quantum computing promises to revolutionize fields such as cryptography, artificial intelligence, and healthcare, offering transformative solutions to some of humanity's most complex problems.

In these notes we will give the reader a comprehensive introduction to how quantum computing works. First we will introduce the reader to basic mathematical concepts like set theory and linear algebra, which are essential to undersand how quantum mechanics work.

1.1 Further Reading

Most linear algebra books cover all the materials from Section 2. I have drawn most material from

Elementary Linear Algebra: Applications Version, Howard Anton and Chris Rorres, Tenth Edition [1].

Here's the chapters that are relevant for this course:

- Chapter 1. Sections: 1.3, 1.4
- Chapter 3. Sections: 3.1, 3.2, 3.3, 3.5
- Chapter 4. Sections: 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.9, 4.10
- Chapter 5. Sections: 5.1, 5.2, 5.3
- Chapter 6. Sections: 6.1, 6.2, 6.3
- Chapter 7. Sections: 7.1, 7.2, 7.5

For Section 3 the source is

Quantum Computation and Quantum Information, M. Nielsen, and I. Chuang. Cambridge University Press, (2000) [3].

Here's the chapters that are relevant for this course:

- Chapter 2. Section 2.1. Linear Algebra.
  - Sections: 2.1.1, 2.1.2,2.1.3,2.1.4,2.1.5,2.1.6,2.1.7, 2.1.8, 2.1.10
- Chapter 2. Section 2.2. Postulates of Quantum Mechanics.
  - Sections: 2.2.1, 2.2.2, 2.2.3,2.2.4,2.2.5, 2.2.6,2.2.7,2.2.8
- Chapter 2. Section 2.4

If you need a refresher on basic logic, set theory and trigonometry read Appendix A and B.

2 Linear Algebra

Linear algebra is a branch of mathematics that deals with vector spaces, linear equations, and transformations. It is foundational to many areas of mathematics and has applications in physics, engineering, economics, but in particular, computer science, machine learning, quantum computing and numerical analysis.

The purpose of linear algebra is to study the structure and behavior of vectors, matrices, and systems of linear equations. This is useful in quantum because qbits will be represented by probability vectors and quantum algorithms will be represented essentially by compositions of linear transformations or matrices.

2.1 Vector Spaces

Definition 2.1 (Vector space)**.** *A* vector space *(or linear space) over a set of scalars $\mathbb{F}$ is a set with a binary operation associated with a neutral element $(V, +, \vec{0})$ together with a scalar multiplication $\cdot : \mathbb{F} \times V \to V$. These operations must satisfy the following axioms for all $\vec{u}, \vec{v}, \vec{w} \in V$ and all $a, b \in \mathbb{F}$:*

1. $v + 0 = v$
2. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
3. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.
4. *For each $\vec{v} \in V$, there exists a vector (called the inverse) $-\vec{v} \in V$ such that $\vec{v} + (-\vec{v}) = \vec{0}$.*
5. *There exist $1 \in \mathbb{F}$, such that $1\vec{v} = \vec{v}$ for all $\vec{v} \in V$*
6. $a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$
7. $(a + b)\vec{v} = a\vec{v} + b\vec{v}$
8. $a(b\vec{v}) = (ab)\vec{v}$

Most examples of vector spaces in this introduction will use the real numbers $\mathbb{R}$. However, in the formal definitions we will use the complex numbers, simply because vector spaces over the complexes have all the properties we need to model quantum computations and all the examples that work for the reals, work also for the complexes since $\mathbb{R} \subseteq \mathbb{C}$, by sending a real number r into a complex numbers $r + i0$ with the null imaginary part.

A two-dimensional Euclidean space $\mathbb{R}^2$ is a vector space over $\mathbb{R}$ of ordered pairs (x, y) where x and y are real numbers. The addition operation is given by point-wise addition and scalar multiplication

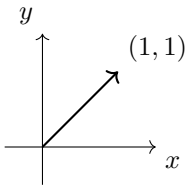


Figure 1: A vector of coordinates $(1, 1)$

is obtained by multiplying all the elements of a vector by a scalar in $\mathbb{R}$. For example, adding $(1, 2)$ and $(3, 4)$ gives $(4, 6)$, while multiplying $(1, 2)$ by a scalar 3 results in $(3, 6)$.

Similarly, in three-dimensional Euclidean space $\mathbb{R}^3$, vectors are represented as triples (x, y, z). Adding $(1, 0, -1)$ and $(2, 3, 4)$ yields $(3, 3, 3)$, and multiplying $(2, -1, 0)$ by a scalar -2 gives $(-4, 2, 0)$. The zero vector, which acts as the additive identity, is $(0, 0)$ in $\mathbb{R}^2$ and $(0, 0, 0)$ in $\mathbb{R}^3$.

Here's the formal definition of a vector space over the real numbers.

2.2 Linear Transformations

A *linear transformation* (or linear map) is a function between two vector spaces that preserves the operations of vector addition and scalar multiplication.

Definition 2.2 (Linear map)**.** *A* linear map *is a function $f : V \to W$ between vector spaces, with the following properties, for all $v, w \in V$ and $s \in \mathbb{C}$:*

$$f(v + w) = f(v) + f(w) \tag{1}$$
$$f(s \cdot v) = s \cdot f(v) \tag{2}$$

An anti-linear map *is a function that satisfies (1) but instead of (2), satisfies*

$$f(s \cdot v) = s^* \cdot f(v)$$

where s^* is the complex conjugate.

These two properties ensure that a linear transformation preserves the structure of the vector space. A simple example of a linear transformation is a scaling transformation. If f is a scaling transformation by a factor of k, then $f(\vec{v}) = k\vec{v}$, where k is a scalar. This transformation scales all vectors by the same factor k, preserving their direction but altering their magnitude.

Obviously, when the vector space is over the reals, the linear map is the same as the anti-linear map as the complex conjugate of a real number is the identity.

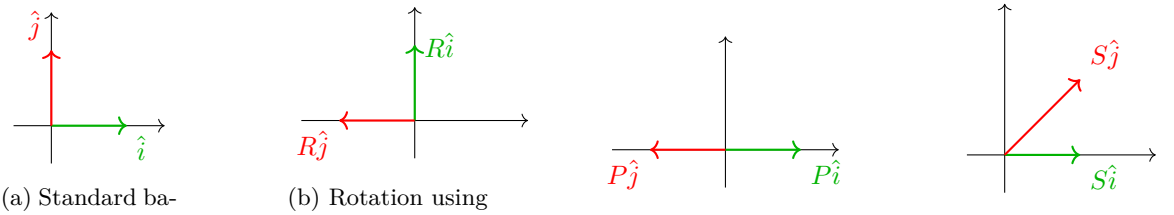
We now list a few examples of linear transformations. The easiest way to understand what a linear transformation is doing is to look at how it modifies the *basis vectors* (See Definition 2.6). For the time being the standard basis for $\mathbb{R}^2$ is simply

Rotation. A *rotation* in $\mathbb{R}^2$ by an angle θ around the origin is a linear transformation. The matrix representing this transformation is defined as follows:

$$f(\vec{v}) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

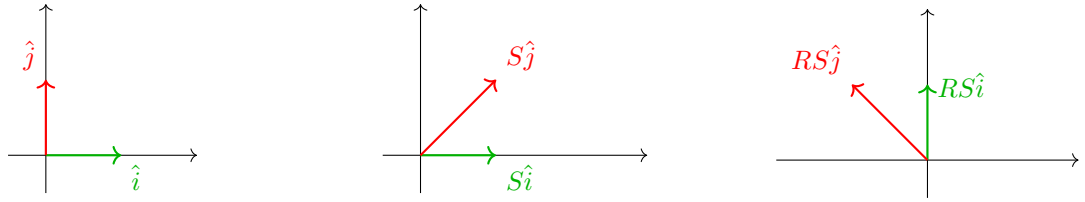
For example, for $\theta = 90°$ we have

$$f(\vec{v}) = R \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$



(a) Standard basis (b) Rotation using R (c) After applying P (d) After applying S

Figure 2: Linear transformations on the standard basis



(a) Standard basis (b) After shear A (c) After applying R

Figure 3: Illustration of the composition $R \circ A$ on the standard basis

where $R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $\vec{v} = (v_1, v_2)$ is a vector in $\mathbb{R}^2$, and $f(\vec{v})$ is the rotated vector. In Figure 2b the left-hand side is the basis of the vector space $\mathbb{R}^2$ and on the right-hand is the rotated basis vectors.

Projection The matrix $P = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$ represents a projection onto the x-axis, intuitively, all vectors are "flattened" onto the x-axis as in Figure 2c.

Shear The matrix $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ represents a horizontal shear of the plane. Geometrically, it leaves the x-axis invariant since $S\hat{i} = \hat{i}$, while the vector $\hat{j}$ is mapped to $A\hat{j} = (1, 1)$. More generally, a vector (x, y) is transformed to

$$S(x, y) = (x + y,\ y)$$

which means that each point is shifted horizontally by an amount proportional to its height y, while its vertical coordinate remains unchanged. Thus horizontal lines stay horizontal, the origin is fixed, areas are preserved ($\det S = 1$), and the transformation produces a slanted, or "sheared," version of the plane. The effect of the shear applied to the basis vectors of $\mathbb{R}^2$ is depicted in Figure 2d.

Composition The *composition* of two linear transformations A and B, written $B \circ A$, is the transformation that applies A first and then B to the result is represented by the matrix product BA, because

$$(B \circ A)(v) = B(A(v)) = B(Av) = (BA)v.$$

For example, consider a horizontal shear S followed by a rotation by 90° counterclockwise R. The composition *rotation after shear* is RS, which generally produces a different result than *shear after rotation* SR. This is illustrated in Figure 3. Notice that the composition of linear transformations is *not commutative*.

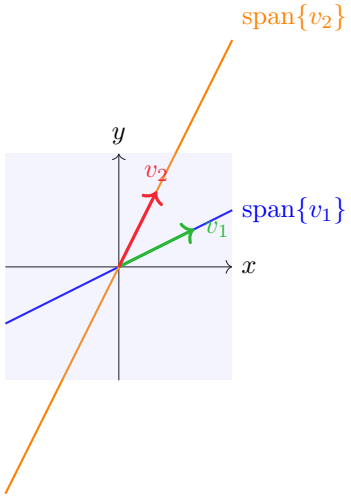


Figure 4: Span of $\{v_1, v_2\} = \mathbb{R}^2$

Definition 2.3 (Determinant)**.** *Let $A = (a_{ij}) \in \mathbb{F}^{n\times n}$. The* determinant *of A is defined by*

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}$$

where S_n is the symmetric group on $\{1, \ldots, n\}$ and $\operatorname{sgn}(\sigma)$ denotes the sign of the permutation σ.

2.3 Linear combinations, Spans and Linear Independence

Two vectors can be added together as depicted in Figure 5a, using the addition operation as in Definition 2.1. Moreover, using the scalar multiplication we can *rescale* a vector (Figure 5b). In general, give a family of vectors v_i and scalars s_i, we can generate a new vector by scaling each single v_i by s_i and then adding the results together. This is called a *linear combination*:

$$v = \sum_i s_i v_i$$

For example, in Figure 5a the vector w can be written as a linear combination of the vectors $u, v \in V$. In this case, we say w is *linearly dependent* from v and u. In other terms, we can also say that $v + u - w = 0$.

Now we want to ask, given a vector space V and some vectors in V, what is the set of vectors generated by all possible linear combinations of these vectors. First we define the notion of subspace.

Definition 2.4 (Vector Subspace)**.** *Let V be a vector space over a field $\mathbb{F}$. A subset $W \subseteq V$ is a* subspace *of V if W is a vector space under the operations inherited from V.*

Now, given a family of vectors $W = w_i$, the *span* of W, written $\operatorname{span}(W)$, is the subspace of vectors that can be written as a linear combination of vectors in W, which can be defined as

$$\operatorname{span}(W) = \{a_1 w_1 + a_2 w_w + \cdots + a_n w_n \mid w_1, \ldots, w_n \in W, a_1, \ldots, a_n \in S\}$$

For example, in $\mathbb{R}^2$ the span of the two vectors v_1, v_2 in Figure 4 is the entire space $\mathbb{R}^2$. Moreover, the span for each single vector is the entire line along that vector.

In $\mathbb{R}^2$ if two vectors do not share the same direction they are called *linearly independent* and their span is the whole space. This is an example of a more general fact which we are going to now illustrate.

A family of vectors $\{v_i\}$ in V is *linearly independent* when it is not linearly dependent, meaning that every vector v_i cannot be written as a linear combination of the others.

Definition 2.5 (Linear Independence)**.** *A sequence of vectors $v_1, \ldots, v_n$ from a vector space V is* linearly independent*, such that if*

$$a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = 0$$

then $a_1 = a_2 = \cdots = a_n = 0$.

In other words, the family of vectors $\{v_i\}$ is linearly independent if for every vector e_i the only way to write v_i as a linear combination of the other vectors is to set the scalars at 0.

Since we learned that vectors can be written as linear combinations of other vectors, we are looking at a sequence of vectors that can generate all vectors in a vector space. Moreover, we want this sequence to be minimal, i.e. the smallest sequence of vectors generating the whole space. This leads to the definition of basis of a vector space.

Definition 2.6 (Basis)**.** *A* basis *for a vector space V is a sequence e_i of linearly independent vectors such that every $v \in V$ can be written as a linear combination of vectors e_i.*

In other words, the base E is the smallest subset of linearly independent vectors such that $\text{span}(E) = V$. The *span* of a set of vectors is the set of all possible linear combinations of those vectors.

The standard basis for the Euclidean space $\mathbb{R}^2$ is $\hat{\imath} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\hat{\jmath} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. In general, the standard basis of the space $\mathbb{R}^n$ is the set of vectors $\{\hat{e}_i\}$ defined as

$$\hat{e}_{i,j} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{otherwise} \end{cases}$$

For example, the standard basis of the 4-dimensional real vector space $\mathbb{R}^4$ is

$$\mathcal{B} = \{e_1, e_2, e_3, e_4\},$$

where

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad e_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

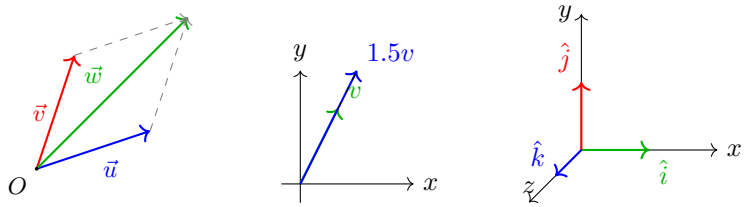
The standard basis is not the only basis for $\mathbb{R}^n$ as every set of n linearly independent vectors will generate the space $\mathbb{R}^n$ as in, for example, those in Figure 5c where $v_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$, $v_2 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$ and $v_3 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ has the one of the components at zero.

Example 2.1 (Canonical Basis for $\mathbb{R}^n$)**.**

2.4 The norm of a vector

The *norm* (or length) of a vector $\vec{v}$ in a vector space V is simply given by Pythagoras' theorem:

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



(a) The sum of v and u (b) Rescaling (c) Standard basis

where $v_1, \ldots, v_n$ are the components of the vector, or in other terms, the projections of the vector onto the respective axis. An example for $\mathbb{R}^2$ is depicted in Figure 7a.

The norm gives rise to the *distance* between two vectors $\vec{u}$ and $\vec{v}$ which can be calculuated as the norm of the difference of the two vectors:

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

To see this consider how to compute the distance of two vectors u and v in the space $\mathbb{R}^2$ as in Figure 7b. The difference between the two vectors is a new vector calculated by substracting component-wise, then the norm is the length of the difference of the two vectors hence it is the distance between the two points.

2.5 Rank and Nullity

The *rank* of a matrix is the maximum number of linearly independent rows or columns in the matrix. In other words, it measures how many rows or columns contain genuinely new information. To better understand this idea, consider that each row (or column) of a matrix can be viewed as a vector. If one row can be written as a linear combination of other rows, then it does not contribute any new information. For example, consider the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

The second row is equal to twice the first row, so there is only one linearly independent row. Therefore, the rank of this matrix is 1. This means that the $2x2$ matrix transforms vectors in $\mathbb{R}^2$ into vectors in $\mathbb{R}$.

From the perspective of linear transformations, suppose a matrix represents a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. The rank of the matrix is equal to the dimension of the image (or range) of T which is the subspace of $\mathbb{R}^m$ of vectors mapped by T. In geometric terms, it tells us how much of the output space is actually reached. A rank of 3 means the image is three-dimensional; a rank of 1 means that all outputs lie along a single line; and rank 0 means everything is mapped to the zero vector. Full rank indicates that the transformation does not collapse any dimensions beyond what is forced by the size of the matrix.

The formal definition is as follows:

Definition 2.7 (Rank)**.** *Let $A \in \mathbb{R}^{m \times n}$. The* rank *of A, denoted* $\operatorname{rank}(A)$*, is the dimension of the column space of A:*

$$\operatorname{rank}(A) = \dim\big(\operatorname{Col}(A)\big).$$

where the column space *of A, denoted* $\operatorname{Col}(A)$*, is the subspace of $\mathbb{R}^m$ spanned by the columns of A.*

Equivalently, the rank of A is the dimension of the row space of A, that is,

$$\operatorname{rank}(A) = \dim\big(\operatorname{Row}(A)\big).$$

where the row space *of A, denoted* $\operatorname{Row}(A)$*, is the subspace of $\mathbb{R}^m$ spanned by the rows of A.*

The column space of A is the set of all possible outputs of this transformation — every vector you can reach by taking a linear combination of the columns. The rank of A counts the number of linearly independent columns. Each independent column adds a "new direction" that the transformation can reach. Therefore, the rank is exactly the dimension of the column space, because it tells you how many independent directions you can reach in the output space.

On the other hand, the *nullity* counts the number of linearly independent solutions of the homogeneous system $Ax = 0$. Equivalently, it equals the number of free variables in the solution.

Definition 2.8 (Null Space and Nullity)**.** *Let $A \in \mathbb{R}^{m\times n}$. The* null space *(or kernel) of A is*

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}.$$

The nullity *of A is the dimension of its null space:*

$$\text{nullity}(A) = \dim \mathcal{N}(A).$$

This leads to the following theorem:

Theorem 2.1 (Rank–Nullity Theorem)**.** *If $A \in \mathbb{R}^{m\times n}$, then*

$$\text{rank}(A) + \text{nullity}(A) = n.$$

Here, the nullity is the dimension of the null space, which consists of all vectors mapped to zero. This theorem shows that the number of independent directions preserved by a transformation (the rank) plus the number of directions collapsed to zero (the nullity) equals the total number of input dimensions.

Beyond pure mathematics, rank has important practical applications. In data science and machine learning, the rank of a data matrix reveals how many features are truly independent. A low-rank matrix indicates redundancy among variables and forms the basis of dimensionality reduction techniques such as principal component analysis. Rank also appears in signal processing, economics, and control theory, where it measures the effective degrees of freedom in systems and models.

In summary, the rank of a matrix measures the amount of independent information it contains. It determines whether systems of equations have solutions, whether matrices are invertible, and how linear transformations affect dimension. For these reasons, rank is one of the most fundamental and powerful concepts in linear algebra.

2.5.1 Solutions of Linear Systems

A *system of linear equations* is a collection of linear equations involving the same set of variables. For example, consider the following system of equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

This system can be equivalently written using the matrix-vector multiplication between a matrix of coefficients A, the vector of unknown variables x and the vector of knowns b :

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{x} \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}}_{v}.$$

Thus, the system of linear equations is represented by the matrix equation

$$Ax = b$$

It is useful to compute now the rank of a matrix because it determines both the *existence of solutions* to $Ax = b$ and whether the associated linear transformation can be *reversed*. As explained earlier the rank of the matrix indicates the range of the linear transformation in terms of the number of dimensions of the space the transformation maps to.

The rank is useful when we want to ask the question

"what is the vector x which is mapped to b by A?"

Geometrically, the equation $Ax = b$ asks whether the vector b lies in the column space of A, or, in other words, is there one or more vectors x that are mapped into b by A.

The system is *consistent* (has at least one solution) if and only if b can be expressed as a linear combination of the columns of A, that is, if

$$\text{rank}(A) = \text{rank}([A \mid b]).$$

If adding b increases the rank, i.e., $\text{rank}([A \mid b]) > \text{rank}(A)$, then b lies outside the column space and the system has *no solution*.

The fact that a system as a solution does not mean necessarily that the matrix is invertible, in fact there may be more than one x for a vector b and in this case the matrix is not invertible.

Definition 2.9 (Singular Matrix)**.** *A square matrix $A \in \mathbb{R}^{n\times n}$ is called* singular *if it is not invertible.*

Equivalently, A is singular if $\det(A) = 0$ $\text{rank}(A) < n$. If $\det(A) \neq 0$ (full rank), the matrix is called *nonsingular* or invertible. A singular matrix "collapse" space along one or more directions. Its column space does not fill $\mathbb{R}^n$, so the linear transformation A cannot be reversed. Geometrically, it might squash a plane into a line, or a volume into a plane, losing information along some dimensions.

For a square $n \times n$ matrix A, the rank also determines invertibility:

- If $\text{rank}(A) = n$, all columns are independent, the column space spans $\mathbb{R}^n$, and A is *invertible*.
- If $\text{rank}(A) < n$, some columns are dependent, the column space does not fill $\mathbb{R}^n$, and A is *singular* (non-invertible).

2.6 Products

2.6.1 Cross Product

The *cross product* (or vector product) is a binary operation on two vectors in three-dimensional space $\mathbb{R}^3$ that produces another vector. For two vectors $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$, the cross product is defined as:

$$u \times v = \begin{pmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix}$$

where $\hat{i}, \hat{j}, \hat{k}$ are the standard basis vectors. Expanding the determinant:

$$u \times v = (u_2v_3 - u_3v_2)\hat{i} - (u_1v_3 - u_3v_1)\hat{j} + (u_1v_2 - u_2v_1)\hat{k}$$

The magnitude of the cross product is given by:

$$\|\vec{u} \times \vec{v}\| = \|\vec{u}\|\|\vec{v}\| \sin(\theta)$$

where θ is the angle between the two vectors.

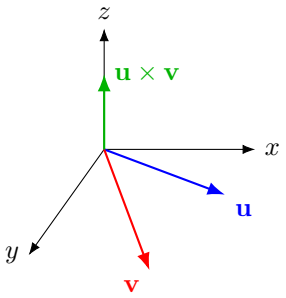


Figure 6: Cross product

2.6.2 Tensor Product

The *tensor product* is a generalization of the outer product (or dyadic product) and is used to construct higher-dimensional objects from two vectors or two matrices. For two vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, the tensor product $u \otimes v$ is an $m \times n$ matrix defined by:

$$\vec{u} \otimes \vec{v} = \begin{pmatrix} u_1v_1 & u_1v_2 & \cdots & u_1v_n \\ u_2v_1 & u_2v_2 & \cdots & u_2v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_mv_1 & u_mv_2 & \cdots & u_mv_n \end{pmatrix}$$

Definition 2.10 (Kronecker Product)**.** *Let $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{p\times q}$ two matrices. The* Kronecker product *of A and B is the block matrix*

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix} \in \mathbb{C}^{mp\times nq}.$$

2.6.3 The dot product

The *dot product* is a way of multiplying two vectors to get a scalar (a real or complex number in our case). For two vectors $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ in $\mathbb{R}^n$, the dot product is defined as:

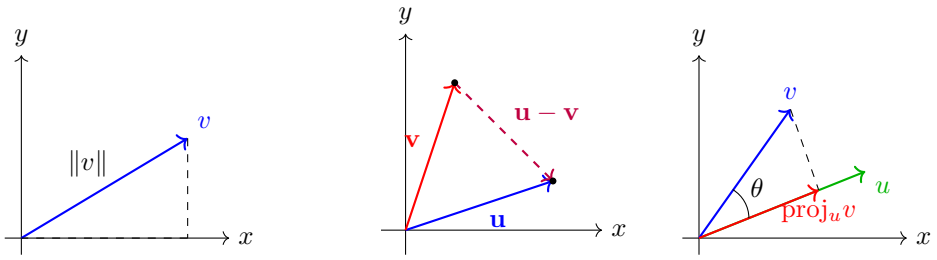
$$u \cdot v \triangleq \sum_{i=1}^{n} u_i v_i = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

The dot product is related to the cosine of the angle θ between the two vectors. Specifically, for two vectors $u, v \in V$ as in Figure 7c, the dot product can be seen as the influence of the vector v onto the vector u. This depends of course on what is the angle θ between the two. Thus, the dot product $\langle \cdot, \cdot \rangle : V \times V \to V$ can be expressed by the following formula:

$$u \cdot v = \|u\| \cdot \text{proj}_u(v) = \|u\|\|v\| \cos(\theta)$$

where $\|u\|$ and $\|v\|$ are the magnitudes (norms) of the vectors, and θ is the angle between them. Now it is easy to see that the dot product gives a notion of distance between two vectors. First we define the norm of a vector in terms of dot product:

$$\|u\| = \sqrt{u \cdot u} = \sqrt{u_1^2 + \cdots + u_n^2}$$



(a) Norm of a vector (b) Distance (c) Dot Product

Figure 7: Norm and Dot Product

As explained previously, the norm can be used to compute the distance between two vectors, hence once we have the dot product we can also define a notion of distance between two vectors as well:

$$d(\vec{u}, \vec{v}) = \|u - v\| = \sqrt{(u - v) \cdot (u - v)}$$

For example, for the vectors $\vec{u} = (2, 3, 4)$ and $\vec{v} = (1, 0, -1)$, to calculate their distance we first compute their difference component-wise:

$$\vec{u} - \vec{v} = (2 - 1, 3 - 0, 4 - (-1)) = (1, 3, 5)$$

then, we compute the norm of $\vec{u} - \vec{v}$:

$$\|\vec{u} - \vec{v}\| = \sqrt{1^2 + 3^2 + 5^2} = \sqrt{1 + 9 + 25} = \sqrt{35}$$

So, the distance between $\vec{u}$ and $\vec{v}$ is $\sqrt{35}$.

In general, the dot product has the following abstract definition.

Definition 2.11 (Dot product)**.** *For a vector space V the dot product $(\cdot) : V \times V \to V$ is a binary operation such that the dot product has several important properties:*

1. $u \cdot v = v \cdot u$
2. $u \cdot (v + w) = u \cdot v + u \cdot w$
3. $(au) \cdot v = a(u \cdot v)$

2.7 Outer product

Definition 2.12. *Let $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ be column vectors. The* outer product *of u and v is the $m \times n$ matrix*

$$u \otimes v = uv^T = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 & \cdots & u_1 v_n \\ \vdots & \ddots & \vdots \\ u_m v_1 & \cdots & u_m v_n \end{pmatrix}.$$

Each column of uv^T is a scaled copy of u, and each row is a scaled copy of v^T.

2.8 Inner product

The inner product is a generalisation of the dot product. In particular, in the Euclidean space $\mathbb{R}^n$ the inner product *is* the dot product. Formally, the inner product $\langle \cdot, \cdot \rangle : V \times V \to V$ between two vectors $u, v \in \mathbb{R}^n$ is defined as

$$\langle u, v \rangle \triangleq u \cdot v$$

In the complex space $\mathbb{C}^n$ the dot product is the Hermitian inner product:

$$\langle u, v\rangle = \sum_{k=1}^{n} u_k^* \, v_k$$

where u_k^* is the complex conjugate of the complex number u_k. This can be written more concisely by defining the conjugate transpose of a matrix.

$$\overline{A} = (A^T)^*$$

Then the complex inner product can be defined simply as

$$\langle v, w\rangle = \overline{v} w$$

Notice that, the conjugate of a real number is just the real number itself. So this definition of inner product coincides precisely with the one for the real vector spaces.

For example, for $u = (1+i, 2), v = (2, 1-i)$ the inner product is calculated as

$$\langle u, v\rangle = (1+i)^* \cdot 2 + 2^* \cdot (1-i) = (1-i)\cdot 2 + 2 \cdot (1-i) = 4 - 4i.$$

So, the inner product of u and v is -2.

The idea of the inner product is to generalise the notion of dot product to an arbitrary space where we want to talk about angles and distances.

Definition 2.13 (Inner Product)**.** *Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$. An* inner product *on V is a function*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$$

(sometimes called a scalar product) satisfying the following properties for all $u, v, w \in V$ and all scalars $\alpha \in \mathbb{F}$:

1. $\langle v + w, z\rangle = \langle v, z\rangle + \langle w, z\rangle$
2. $\langle v, w + z\rangle = \langle v, w\rangle + \langle v, z\rangle$
3. $\langle v, \alpha w\rangle = \alpha \langle v, w\rangle$
4. $\langle \alpha v, w\rangle = \alpha^* \langle v, w\rangle$
5. $\langle v, w\rangle = \langle w, v\rangle^*$
6. $\langle v, v\rangle \geq 0$, *with* $\langle v, v\rangle = 0$ *if and only if* $v = 0$

Exercise 2.1. *Prove the dot product is an inner product.*

As it was the case for the dot product, the norm of a vector v can be computed using the inner product:

$$\|\vec{v}\| = \sqrt{\langle v, v\rangle} \tag{3}$$

Therefore, we can define the distance of two vectors similarly to how we defined it for the dot product:

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

A property of the inner product that is going to be useful in quantum is the *Cauchy–Schwarz inequality* which states that in any inner product space,

$$|\langle u, v\rangle| \leq \|u\| \, \|v\|$$

for all vectors u and v. It says that the absolute value of the inner product can never exceed the product of the norms of the vectors. Geometrically, since

$$\langle u, v\rangle = \|u\|\,\|v\|\cos\theta,$$

the inequality simply reflects that $|\cos\theta| \leq 1$. Equality holds if and only if u and v are linearly dependent. Finally, the inner product is also useful to compute whether a basis is made of orthogonal or orthonormal vectors. Intuitively, two orthogonal vectors in $\mathbb{R}^2$ are two vectors that are at $90°$ to each other, these are also orthonormal if they are *normal* that is their length is 1. However, for higher dimensions, or simply when we do not want to rely on the specifics of one space or another we can simply rely on the inner product.

Definition 2.14 (Orthogonal and Orthonormal Basis)**.** *For a Hilbert space H, an orthogonal basis is a family of elements $\{e_i\}$ with the following properties:*

- *they are pairwise orthogonal, i.e. $\langle e_i, e_j\rangle = 0$ for all $i \neq j$;*
- *every element $a \in H$ can be written as an linear combination of e_i*

Additionally, an orthogonal basis is orthonormal if all its vectors are of length 1, i.e. $\langle e_i, e_i\rangle = 1$, for all i.

This definition becomes even easier when we introduce the Kronecker δ function, pronounced the "Kronecker delta".

Definition 2.15 (Kronecker δ)**.** *The* Kronecker delta *is the function*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The Kronecker delta acts like an identity selector:

$$\sum_j \delta_{ij} a_j = a_i$$

It is the discrete analogue of the identity matrix entries. This is convenient to define orthornormal basis very concisely.

Proposition 2.1. *Let $\mathbb{B} = \{e_i\}$ be a set of basis vectors. Then $\mathbb{B}$ is orthonormal iff*

$$\langle e_i, e_j\rangle = \delta_{ij}$$

for all i, j.

2.9 Diagonalisation and Changes of Basis

In linear algebra, a *change of basis* means describing the same vector using a different coordinate system. The vector itself does not change; only the way we represent it numerically changes. In $\mathbb{R}^2$, the standard basis is

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and every vector v can be written uniquely as a linear combination of these two vectors, for example $v = ae_1 + be_2$, where a, b are the coefficients of the vector v.

Now consider the matrix

$$P = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}$$

whose columns are $v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Since these two vectors are linearly independent, they form a new basis of $\mathbb{R}^2$ — albeit not an either orthogonal or normal one — hence, every vector v can be also be written as a linear combination of the basis vectors in P

$$v = xp_1 + yp_2$$

However, for the same vector v the coefficients x, y will not be the same as a, b which means that despite being the same vector, from the point of view of a different basis the vector has different coordinates. The equation above can be rewrite as

$$v = P \begin{pmatrix} x \\ y \end{pmatrix} \tag{4}$$

Then to find the coefficients of the vector v we have to compute the inverse P^{-1} and compute

$$\begin{pmatrix} x \\ y \end{pmatrix} = P^{-1}v$$

The vector obtained, namely $\begin{pmatrix} x \\ y \end{pmatrix}$, is the vector v as if v was given in the new basis P, and $\begin{pmatrix} x \\ y \end{pmatrix}$ is how the vector looks like in the standard basis. Similarly, to convert a vector $\begin{pmatrix} x \\ y \end{pmatrix}$ from the standard basis to the new basis we have to use the formula (4).

In this sense, the matrix P can be viewed as a *change of basis matrix*.

Now consider the rotation matrix R as in Figure 2b. This transformation operates in the standard basis interpretation of $\mathbb{R}^2$. We now want interpret a rotation as if the space was seen as from the lenses of a new basis whose vectors yield the linear transformation P. In other words, we want a 90° rotation matrix in the new basis.

Given a vector v in the new basis given by P, we have to transform it into a vector in the standard basis by multiplying by P. At this point we can rotate it using the rotation R in the standard basis and then we can convert the rotated vector into a vector in the new basis given by P. This translates to the following:

$$P^{-1}RPv$$

In particular, if P is defined as in the previous example we obtain the follow expression:

$$\overbrace{\begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}^{-1}}^{\text{Change of basis}} \underbrace{\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}}_{\text{Rotation}} \overbrace{\begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}}^{\text{Change of basis}} v$$

Note, that the composition of matrices $P^{-1}RP$ is the rotation matrix in the new basis we were looking for.

2.10 Eigen decomposition

Consider the standard basis vectors for $\mathbb{R}^2$ (this is just a subset of the standard basis in $\mathbb{R}^3$ as in Figure 2a) and consider again the shear matrix as in Figure 2d. After applying the shear to each vector $\hat{\imath}$ and $\hat{\jmath}$ we can notice that certain vectors get knocked off their trajectory. Formally, for a vector v, the applying a transformation S to v means computing Sv which may or may not not belong to the span of the original vector v, which is formed by all the possible rescalings of the vector $\hat{\imath}$.

In particular, for the standard basis $S\hat{\imath}$ belongs to the span of $\hat{\imath}$ and $S\hat{\jmath}$ does not belong to the span of $\hat{\jmath}$ as in Figure 8. This leads to the observation that there are certain vectors of the space which retain their trajectory and others that do not.

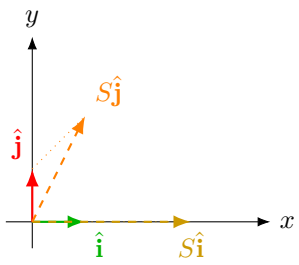


Figure 8: Eigenvectors and Eigenvalues

For a linear transformation, the vectors which retain their trajectory are called *Eigenvectors* and for each *Eigenvector* its rescaling factor is called the Eigenvalue of the Eigenvector.

Formally, let A be an $n \times n$ matrix. A nonzero vector $\mathbf{v}$ is called an *eigenvector* of A if there exists a scalar λ such that

$$A\mathbf{v} = \lambda\mathbf{v}.$$

The scalar λ is called the *eigenvalue* corresponding to the eigenvector $\mathbf{v}$. This equation means that when the matrix A acts on the vector $\mathbf{v}$, the result is simply a scalar multiple of $\mathbf{v}$. In other words, the direction of $\mathbf{v}$ does not change under the transformation; it is only stretched or compressed (and possibly reversed if λ is negative). Most vectors change direction when multiplied by a matrix, but eigenvectors do not.

To find eigenvalues, we rewrite the equation as

$$A\mathbf{v} = \lambda\mathbf{v} \quad \Longrightarrow \quad (A - \lambda I)\mathbf{v} = \mathbf{0}.$$

For this equation to have a nontrivial solution $\mathbf{v} \neq \mathbf{0}$, the matrix $A - \lambda I$ must be singular. Therefore, we require

$$\det(A - \lambda I) = 0.$$

This equation is called the *characteristic equation* of A, and its solutions are the eigenvalues. Once the eigenvalues are found, the corresponding eigenvectors are non-zero vectors obtained by solving

$$(A - \lambda I)\mathbf{v} = \mathbf{0}$$

for each eigenvalue λ.

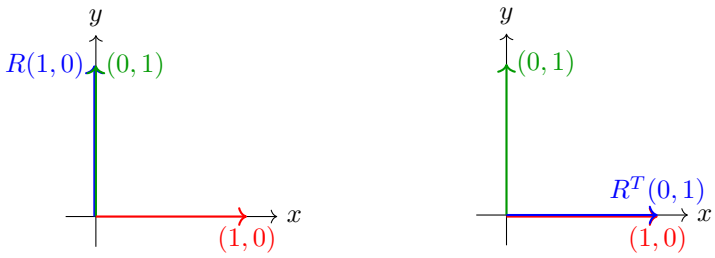
Geometrically, eigenvectors represent directions in which the linear transformation acts in a simple way—by pure scaling. Eigenvalues measure how much scaling occurs along those directions. These eigenvectors are the nonzero vectors in the null space of the matrix $\lambda I - A$. This null space is called the *eigenspace* of A corresponding to one particular λ.

Theorem 2.2 (Eigendecomposition)**.** *if $A \in V^{n\times n}$ then the following statements are equivalent:*

- *A is diagonalisable*
- *A has n linearly independent eigenvectors*

Decomposition Procedure

1. Find out if the matrix is diagonalisable by finding n linearly independent Eigenvectors. One way to do this, is to find a basis for each Eigenspace and merge these basis vectors into a set S. If this set has fewer than n vectors, the matrix is not diagonalisable.
2. Form the matrix $P = \begin{pmatrix} p_1 p_2 \dots p_n \end{pmatrix}$ where the vectors (p_i) are in S.



(a) $\langle R\hat{i}, \hat{j}\rangle = 1$ (b) $\langle \hat{i}, R^T\hat{j}\rangle = 1$

Figure 9: Hermitian adjoint

3. The matrix $P^{-1}AP$ will be diagonal and have Eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ corresponding to the eigenvectors $p_1, p_2, \ldots, p_n$ as its diagonal entries.

Exercise 2.2. *Find (if they exists) the eigenvectors and eigenvalues of the rotation matrix, the shear and the projection matrix.*

2.11 Hermitian Matrices and the Spectral Decomposition Theorem

In this section we are interested in linear transformation that preserve the inner product of two vectors, but in a "complementary way". For example, consdier the standard basis of the Euclidean space $\mathbb{R}^2$ and consider the two basis vectors $\hat{i}$ and $\hat{j}$ as in Figure 9. The rotation matrix applied only to $\hat{i}$, as in Figure 9a moves the vector in the same position as $\hat{j}$. Hence the inner product $\langle R\hat{i}, \hat{j}, =\rangle 1$. Their angle is $0°$ and both their lengths are 1 hence the contribution of $R\hat{i}$ onto $\hat{j}$ is 1. The matrix associated with R, written $R^\dagger$ and pronounced the Hermitian adjoint, is the unique matrix which transforms $\hat{j}$ such that it does not change the inner product with $\hat{i}$. In particular, in Figure 9b the Hermitian adjoint $R^\dagger$ $\hat{j}$ in the same position as $\hat{i}$, thus maintaining the inner product.

We now define the Hermitian adjoint formally.

Definition 2.16 (Conjugate transpose / Hermitian transpose / Hermitian adjoint)**.** *Let $A = (A_{ij}) \in \mathbb{C}^{n\times n}$ be a complex square matrix. The* conjugate transpose *of A, denoted $A^\dagger$, is the unique matrix which satisfies the equation*

$$\langle Av, w\rangle = \langle v, A^\dagger w\rangle \tag{5}$$

We are going to make extensive use of this fact in Section 2.12.

In the case of the complex vector space $\mathbb{C}^n$ the Hermitian adjoint is just the complex conjugate.

Proposition 2.2. *The Hermitian adjoint of a complex matrix is the conjugate transpose*

$$A^\dagger = (A^T)^*$$

Proof. Assume two vectors v, w, such that $\langle Av, w\rangle = \langle v, A^\dagger w\rangle$. By unfolding the definitions of inner product we get that $\overline{Av}w = \overline{v}A^\dagger w$, for all v, w. By the properties of the conjugate transpose $\overline{Av}w = \overline{v}\overline{A}w$. Since this is true for all w, v we can cancel them, thus obtaining $\overline{A} = A^\dagger$. □

Corollary 1. *Let $A \in \mathbb{C}^{n\times m}$. The Hermitian adjoint is the matrix obtained by taking the transpose and then the complex conjugate of each entry:*

$$(A^\dagger)_{ij} = A^*_{ji}, \quad i = 1, \ldots, n,\ j = 1, \ldots, m,$$

Exercise 2.3. *Prove that the Hermitian adjoint satisfies the following properties:*

1. $A^{\dagger^\dagger} = A$
2. $(A+B)^\dagger = (A^\dagger + B^\dagger)$
3. $(AB)^\dagger = B^\dagger A^\dagger$

Notice however, that even thought $R^\dagger$ happens to coincide with R^{-1}, it is not in general true that the conjugate transpose is the inverse of the matrix. For example, the Hermitian adjoint of $A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ is A itself since its entries are reals and the matrix is diagonal, but its inverse A^{-1} is $\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}$. This matrix however is particular since its Hermitian adjoint is itself, and it's called an Hermitian matrix.

Definition 2.17 (Hermitian Matrix)**.** *A square matrix $A \in \mathbb{C}^{n\times n}$ is called* Hermitian *if it is equal to its conjugate transpose:*

$$A = A^\dagger$$

that is,

$$A_{ij} = A^*_{ji} \quad \text{for all } i, j = 1, \ldots, n,$$

where A^*_{ji} denotes the complex conjugate of A_{ji}.

Hermitian matrices are real symmetric matrices. That is for a matrix A, all the entries A_{ij} are real and $A^T = A$. This is a very useful property to have as it admits a diagonal decomposition which we described in Section 2.9.

Theorem 2.3 (Spectral Decomposition Theorem)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix (or $A \in \mathbb{C}^{n\times n}$ be Hermitian, i.e., $A = A^\dagger$). Then there exist orthonormal eigenvectors $v_1, \ldots, v_n$ and real eigenvalues $\lambda_1, \ldots, \lambda_n$ such that*

$$A = \sum_{i=1}^{n} \lambda_i \, v_i v_i^T.$$

Equivalently, A can be written as

$$A = Q\Lambda Q^T,$$

where Q is an orthogonal (unitary) matrix whose columns are the eigenvectors of A, and Λ is a diagonal matrix whose diagonal entries are the eigenvalues of A.

Proof. To see this we explain how to construct Q and Λ. Since A is symmetric, its characteristic polynomial has at least one real root. Hence, there exists $\lambda_1 \in \mathbb{R}$ and a nonzero vector v_1 such that

$$Av_1 = \lambda_1 v_1.$$

Normalize v_1 so that $\|v_1\| = 1$. Let

$$V_1 = \{x \in \mathbb{R}^n : v_1^T x = 0\}$$

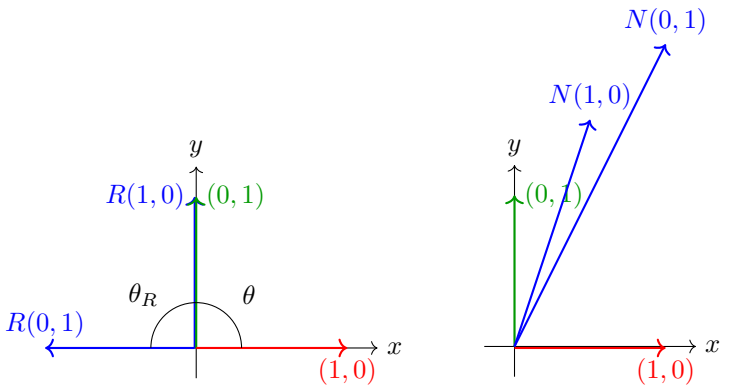
be the orthogonal complement of spanv_1. We claim V_1 is closed under application of A. Indeed, for any $x \in V_1$,

$$v_1^T(Ax) = (Av_1)^T x = \lambda_1 v_1^T x = 0,$$

so $Ax \in V_1$.

Restrict A to V_1. This restriction is again symmetric, so we can repeat the argument: there exists a unit eigenvector $v_2 \in V_1$ with eigenvalue λ_2.

Continuing inductively, we obtain an orthonormal set $\{v_1, \ldots, v_n\}$ of eigenvectors with corresponding real eigenvalues $\lambda_1, \ldots, \lambda_n$.



(a) A Unitary Matrix (b) A Non-Unitary Matrix

Figure 10: Unitary and Non-Unitary Matrices

We now construct Q as the matrix formed by the vectors v_i

$$Q = [v_1 \ \cdots \ v_n], \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n).$$

Then Q is orthogonal ($Q^T Q = I$), and

$$AQ = Q\Lambda.$$

Multiplying on the right by Q^T gives

$$A = Q\Lambda Q^T.$$

□

2.12 Unitary Matrices

A *unitary* matrix represents a transformation that preserves the geometry of the space. In particular, it does not change lengths or angles between vectors. Following this the matrix U is unitary and $x, y \in \mathbb{C}^n$, then

$$\|Ux\| = \|x\| \quad \text{and} \quad \langle Ux, Uy\rangle = \langle x, y\rangle.$$

The first equation states that the length of the vector does not change after applying the unitary transformation, while the second states that apply U to two vectors preserves their inner product. Thus, a unitary matrix can be thought of as a *rigid motion* in complex vector space: it may rotate vectors, reflect them, or change their complex phase, but it never stretches or distorts them.

For example, we consider the rotation matrix R and the matrix $N = \begin{pmatrix} 1 & 2 \\ 1.5 & 2 \end{pmatrix}$. The rotation matrix is a classic example of a unitary matrix whilst N is non-unitary. To see this it suffices to see what the respective matrices do on the standard basis vectors $\hat{i}$ and $\hat{j}$. The rotation sends the standard basis vectors to $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$, respectively, as in Figure 10a. On the other hand, the matrix N maps the standard basis vectors to two vectors with a smaller angle between them and different lengths as in Figure 10b.

From these intuition we can derive the formal definition of a unitary matrix. First recall that the inner product is a complex space is defined as $\langle x, y\rangle = x^\dagger y$. Then we compute as follows:

$$\begin{aligned}
&\langle Ux, Uy\rangle = \langle x, y\rangle \\
&\iff (Ux)^\dagger Uy = x^\dagger y \\
&\{ \text{ By Exercise 2.3 } \} \\
&\iff x^\dagger U^\dagger Uy = x^\dagger y \\
&\{ \text{ Bring everything to one side } \} \\
&\iff x^\dagger U^\dagger (Uy) - x^\dagger y = 0 \\
&\{ \text{ factor out } x^\dagger \} \\
&\iff x^\dagger (U^\dagger (Uy) - y) = 0 \\
&\{ \text{ factor out } y \} \\
&\iff x^\dagger (U^\dagger U - I) y = 0
\end{aligned}$$

Now if we let $A := U^\dagger U - I$, this amounts to solve the equation $x^\dagger A y = 0$, for all vectors x, y. Now, since this has to hold for all vectors x, y it has to hold also for the standard basis vectors $\{e_i\}$. Since $e_j^\dagger A e_i = A_{i,j}$ then, for all i, j, $A_{i,j} = 0$ which implies that $A = 0$. Hence we get

$$U^\dagger U - I = 0$$

This leads to the formal definition of the a unitary matrix.

Definition 2.18 (Unitary Matrix)**.** *A matrix $U \in \mathbb{C}^{n\times n}$ is called* unitary *if*

$$U^\dagger U = U U^\dagger = I,$$

where $U^\dagger$ denotes the conjugate transpose of U, and I is the identity matrix.

It follows that the determinant of a unitary matrix is always 1 and its eigenvalues are also always 1.

Furthermore, a matrix U is unitary if and only if its columns (and rows) form an orthonormal basis of $\mathbb{C}^n$, i.e.

$$\langle u_i, u_j\rangle = \delta_{ij}.$$

where δ_{ij} is the Kronecker δ.

2.13 Normal Matrices

Normal matrices generalise several kinds of well-behaved matrices including all those in Table 1, that is Hermitian, unitary and diagonal matrices. The formal definition is as follows:

Definition 2.19 (Normal Matrix)**.** *A matrix is* normal *if it commutes with the conjugate transpose, that is*

$$\overline{A} A = A \overline{A}$$

Exercise 2.4. *Prove that Hermitian, unitary and diagonal matrices are normal.*

Property	Hermitian	Unitary	Diagonal
Definition	$A = A^\dagger$	$A^\dagger A = I$	Diagonal entries $\neq 0$
Geometric mean.	Scaling on orthogonal dir.	Rotation/reflection	Independ. scaling on axes
Eigenvalues	Real	$\lvert\lambda\rvert = 1$	Diagonal entries
Eigenvectors	Orthonormal basis	Orthonormal basis	Standard basis
Diagonalizable	Yes (unitarily)	Yes (unitarily)	Already diagonal
Preserves length	No	Yes	No (unless ± 1)
Orthogonal eigenvec.	Yes	Yes	Yes
Inverse	If no zero eigenvalues	$A^{-1} = A^\dagger$	Invert diagonal entries
Spectrum	Real axis	Unit circle	Arbitrary

Table 1: Comparison of Hermitian, Unitary and Diagonal matrices

2.14 Trace of a Matrix

In a matrix, the diagonal elements measure how much each basis vector is stretched along itself. More precisely, if we consider a matrix A and a standard basis e_i then the the entry A_{ii} specifies how much of the basis vector e_i is stretched along its direction. The trace of a matrix, essentially, sums all these contributions.

The *trace* of a square matrix $A \in \mathbb{C}^{n\times n}$ is defined as the sum of its diagonal elements:

$$\operatorname{trace}(A) = \sum_{i=1}^{n} A_{ii}$$

Intuitively, the trace measures the total amount by which a linear transformation stretches space along its principal directions. Recall from Section 2.10 that each eigenvalue represents scaling the associated eigenvector along one direction. Essentially, the trace adds up all these scalings.

For example, the trace of the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

can be computed as $\operatorname{trace}(A) = 1 + 5 + 9 = 15$. We list the properties of the trace operation:

- *Linearity:*
$$\operatorname{trace}(A + B) = \operatorname{trace}(A) + \operatorname{trace}(B) \qquad \operatorname{trace}(cA) = c\operatorname{trace}(A)$$
- *Cyclic property:*
$$\operatorname{trace}(AB) = \operatorname{trace}(BA)$$
- *Invariance under similarity:*
$$\operatorname{trace}(P^{-1}AP) = \operatorname{trace}(A)$$

More in general, the cyclic property implies

$$\operatorname{trace}(ABC) = \operatorname{trace}(BCA) = \operatorname{trace}(CAB)$$

Another useful result is that the trace equals the sum of the eigenvalues of A:

$$\operatorname{trace}(A) = \sum_i \lambda_i$$

which holds even if A is not diagonal.

Notation	Description
z^*	Complex conjugate of the complex number z. Example: $(1+i)^* = 1-i$
$\|\psi\rangle$	Vector, also known as a ket
$\langle\psi\|$	Vector dual to $\|\psi\rangle$, also known as a bra
$\langle\phi\|\psi\rangle$	Inner product between the vectors $\|\phi\rangle$ and $\|\psi\rangle$
$\|\phi\rangle \otimes \|\psi\rangle$	Tensor product of $\|\phi\rangle$ and $\|\psi\rangle$
$\|\phi\rangle\|\psi\rangle$	Abbreviated notation for the tensor product
A^*	Complex conjugate of the matrix A
A^T	Transpose of the matrix A
$A^\dagger$	Hermitian conjugate (adjoint) of A
$\langle\phi\|A\|\psi\rangle$	Inner product between $\|\phi\rangle$ and $A\|\psi\rangle$; equivalently between $A^\dagger\|\phi\rangle$ and $\|\psi\rangle$
$\overline{A}$	The conjugate transpose $(A^T)^*$

Table 2: Dirac Notation from Nielsen and Chuang [3]

2.15 Dirac Notation

We conclude with introducing a notation for vectors and inner products which is summarized in Table 2. This is usually introduced with quantum mechanics albeit it is not strictly connected to it, in the sense, that we could have introduced at the beginning of these notes but we chose not to .

In Dirac notation, vectors in a complex Hilbert space are written as *kets* and *bras*. For a vector $\psi = \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_n \end{pmatrix}$, a ket is simply written $|\psi\rangle$. Its conjugate transpose $\psi^\dagger$ is written as

$$\langle\psi| = \begin{pmatrix} v_0^* & v_1^* & \cdots & v_n^* \end{pmatrix}$$

and pronounced "bra". The reason why this notation is convenient is because the inner product arises as the product of a ket and bra, in words, a bra-ket (bracket):

$$\langle\phi|\psi\rangle = \langle\phi, \psi\rangle.$$

This is simply the matrix-vector multiplication of the row vector $\langle\psi|$ with the column vector $|\psi\rangle$. A linear transformation (or operator) A acting on a ket is written $A|\psi\rangle$. Moreover, applying a linear transformation on a $|\psi\rangle$ and then performing the inner product between $\langle\psi|$ and $A|\psi\rangle$ is written as follows:

$$\langle\phi|\, A\,|\psi\rangle = \langle\phi, A\psi\rangle.$$

Finally, the outer product of $|\phi\rangle$ and $\langle\psi|$ is

$$|\phi\rangle\langle\psi|$$

which is simply the matrix-matrix multiplication of a 1-column matrix $|\phi\rangle$ with a 1-row matrix $\langle\psi|$.

2.16 Hilbert Spaces (*)

2.16.1 Cauchy complete spaces

A sequence (a_n) is called a *Cauchy sequence* if its terms become arbitrarily close to each other as the sequence progresses.

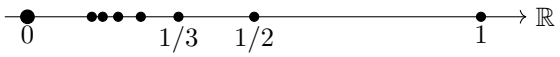


Figure 11: The Cauchy sequence $a_n = 1/n$

Definition 2.20 (Cauchy sequence)**.** *A sequence of points a_i in $\mathbb{R}^n$ is called a* Cauchy sequence*, iff for every $\varepsilon > 0$, there exists an index N such that for all $m, n \geq N$, the distance between the terms satisfies*

$$|a_m - a_n| < \varepsilon$$

This definition does not mention a limit; instead, it focuses only on how close the terms are to one another.

For example, the sequence $a_n = \frac{1}{n}$ depicted in Figure 11 is Cauchy in $\mathbb{R}$. As n becomes large, the values $1/n$ become very small, and the difference between any two large terms becomes arbitrarily small.

In the real numbers every Cauchy sequence converges, which is why $\mathbb{R}$ is called a complete space.

Definition 2.21 (Limit)**.** *Let (a_n) be a sequence in $\mathbb{R}$. We say that (a_n)* converges *to a real number L, and write*

$$\lim_{n\to\infty} a_n = L \quad \textit{or} \quad a_n \to L,$$

if for every $\varepsilon > 0$ there exists a natural number N such that for all $n \geq N$,

$$|a_n - L| < \varepsilon.$$

A space is complete *iff every Cauchy sequence has a limit.*

This definition means that the terms of the sequence eventually get arbitrarily close to L. No matter how small a distance ε we choose, there is some index N after which all terms of the sequence lie within the ε-neighborhood of L. In other words, beyond a certain point, the sequence stays as close to L as we wish.

So far we have defined the notion of Cauchy sequence and limit using the space $\mathbb{R}^n$ in which the notion of distance between two points (vectors) is given by the norm of the difference between the vectors. The norm is in turn defined using the inner product, so in order to generalise these definitions we just need to postulate a space with an inner product in which all Cauchy sequences are complete. This idea leads to the definition of a Hilbert space.

Definition 2.22 (Hilbert Space)**.** *A* Hilbert space *is a vector space H equipped with an inner product such that the metric induced by the norm is complete, i.e. every Cauchy sequence converges.*

3 Quantum Mechanics

The first postulate states that Hilbert spaces are correct mathematical objects in which to interpret quantum state spaces.

Postulate 1. *Associated to any isolated physical system is a Hilbert space known as* state space *of the system. The system is completely described by its state vector which is a unit vector in the system's state space.*

3.1 Quantum States and Quantum Superposition

A quantum state of a system is represented by a column vector whose indices are placed in correspondence with the classical states of that system. The entries are complex numbers such that the sum of

the absolute values squared of the entries must equal 1. In other words, a quantum state is simply a column vector over the complexes such that its norm is 1:

$$\|v\| = \sqrt{\sum_{k=1}^{n} v_k^2} = 1$$

The qubit $|0\rangle$ is the column vector representing the *ground state* which is 0 with probability 1 and 1 with probability 0. The qubit $|1\rangle$ is the column vector representing the *excited state* which is 1 with probability 1 and 0 with probability 0.

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

These two states form the *basis* of the complex Euclidean space $\mathbb{C}^2$ which is called the *computational basis*. This means that every other state can be written as a linear combination of these two states.

The linear combinations and computational basis are a fundamental concepts that allows us to define a *superposition* of states. Superposition describes the ability of a quantum system to exist simultaneously in multiple states. Unlike classical bits, which can only be in one state at a time (0 or 1), a qubit can be in a combination of both states at once. A quantum state in superposition is modelled by a linear combination as follows:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \tag{6}$$

where α and β are complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$. This is equivalent to saying that $|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. In other words, the qubit does not have a definite state but rather a probabilistic combination of states. Intuitively, when measured, the qubit collapses into one of the basis states, with probabilities determined by $|\alpha|^2$ and $|\beta|^2$.

Two examples of quantum states in superposition $|+\rangle$ and $|-\rangle$. These two states are interesting because they are both *equal superpositions* of $|0\rangle$ and $|1\rangle$, in the sense that the linear combinations have different coefficients, but they have a different *phase*:

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

This means that the minus sign makes them behave differently, but one measures them they will look identical. Intuitively, $|0\rangle$ has half probability of being measured at $|0\rangle$ and half to be $|1\rangle$ and so is $|1\rangle$.

Exercise 3.1. *Show that $|+\rangle$ and $|-\rangle$ form an orthonormal basis.*

There are two things we can do with a quantum state: measure it or let it evolve unitarily without measuring it. We will explain evolution first.

3.2 Quantum Evolution and Quantum Gates

We now turn our attention to how quantum states can evolve. This first and perhaps most straightforward requirement is that, whatever the evolution of a particular quantum state may be, this has to preserve the sum of the probabilities for each outcome. In other words, if U is a linear operator acting on quantum system $|\psi\rangle$ then, if $\| |\psi\rangle \| = 1$ then also $\|U|\psi\rangle\| = 1$. By unfolding the definition of norm of a vector we obtain that the inner product $\langle U|\psi\rangle, U|\psi\rangle\rangle$ must be equal to $\langle\psi,\psi\rangle$, that is 1 (see Section 2.4). By Definition 2.16 the inner product $\langle U|\psi\rangle, U|\psi\rangle\rangle$ is equal to $\langle|\psi\rangle, U^\dagger U|\psi\rangle\rangle$ using its hermitian adjoint. We obtain the follow equation

$$\langle\psi|U^\dagger U|\psi\rangle = 1$$

for all $|\psi\rangle$. Since this has to hold for all $|\psi\rangle$ it has to hold in particular for the computational basis $\{|e_i\rangle\}$ which leads to the fact that $U^\dagger U$ must be the identity matrix.

In simple words, since quantum gates (or quantum evolutions) must be preserve probabilities, the operator U must be unitary as described in Section 2.12.

Postulate 2. *The evolution of a closed quantum system is described by a* unitary transformation*. That is the state $|\psi\rangle$ of the system at time t_1 is related to the state $|\psi'\rangle$ of the system t_2 by a unitary operator U which depends only on the times t_1 and t_2.*

Given a sequence of unitary operators indexed by a timestamp t, that is $U(t)$, and an initial state $|\psi(0)\rangle$, the time evolution of the system is given by:

$$|\psi(t)\rangle = U(t)\ |\psi(0)\rangle$$

which is a sequence of vectors indexed by the timestamp.

Definition 3.1 (Quantum Gate)**.** *A* quantum gate *is a unitary operator acting on one or more qubits.*

We list formally the properties arising from this definition:

- *Unitary:* $U^\dagger U = I$
- *Linear:* $U(a|\psi\rangle + b|\phi\rangle) = aU|\psi\rangle + bU|\phi\rangle$
- *Reversible:* $U^{-1} = U^\dagger$
- *Probability Conservation:* $\|U|\psi\rangle\| = \||\psi\rangle\|$

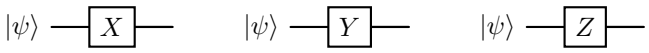
3.2.1 Basic Quantum Gates

The *Pauli matrices* are a set of three fundamental 2×2 complex matrices that act on a single qubit. They are denoted by σ_x, σ_y, σ_z.

The first quantum gate we are going to see is the σ_x or bit-flip operator. The task is to come up with a matrix which maps $|0\rangle \mapsto |1\rangle$ and $|1\rangle \mapsto |0\rangle$. As we have seen in Section 2.2, we can just construct the matrix by placing $|1\rangle$ in the first column and $|0\rangle$ in the second column, that is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Similarly, we can construct the bit-flip with a phase and the phase-flip operators:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In circuit notation, these operators are represented as single-qubit gates:



The Pauli matrices essentially correspond to rotations on the Bloch sphere, i.e. they generate rotations around the x, y, and z axes.

Exercise 3.2. *Check that all these operators are Hermitian, unitary, their traces is* 0 *and their determinant is* -1*. That is, for all $i \in \{x, y, z\}$,*

- *Hermitian:* $\sigma_i^\dagger = \sigma_i$
- *Unitary:* $\sigma_i^\dagger \sigma_i = I$
- *Self-inverse:* $\sigma_i^2 = I$

- *Trace and determinant:* $\text{trace}(\sigma_i) = 0, \quad \det(\sigma_i) = -1$

The *Hadamard gate*, denoted H, is a fundamental single-qubit quantum gate that creates superposition:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

In circuit diagrams, it is represented as:

$$|\psi\rangle \text{ —} \boxed{H} \text{—}$$

To see why the Hadamard gate creates superposition, it is sufficient to apply it to the computational basis and observe that $H|0\rangle$ yields the superposition state $|+\rangle$ and, conversely, $H|1\rangle$ yields $|-\rangle$:

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = |+\rangle, \qquad H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) = |-\rangle.$$

As all the other gates, the Hadamard also acts as a rotation on the Bloch sphere. Conversely, because the Hadamard is reversible, we also have

$$H|+\rangle = |0\rangle, \qquad H|-\rangle = |1\rangle.$$

Exercise 3.3. *Check that the Hadamard gate also satisfies the properties in Exercise 3.2.*

Exercise 3.4. *What are the eigenvalues and eigenvectors of H?*

3.2.2 Phase Operations

Phase operations act by modifying the complex phases of the amplitudes without changing their magnitudes. Consider a general qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. A *global phase* transformation multiplies the entire state by a common complex factor:

$$|\psi\rangle \longrightarrow e^{i\phi}|\psi\rangle. \tag{7}$$

Such a transformation has no physical or observable consequence, since all measurement probabilities depend only on $|\alpha|^2$ and $|\beta|^2$. For example,

$$H|0\rangle \longrightarrow e^{i\pi/3}H|0\rangle \tag{8}$$

represents the same physical state. In contrast, a *relative phase* transformation alters the phase between components of the superposition:

$$|\psi\rangle \longrightarrow \alpha|0\rangle + e^{i\theta}\beta|1\rangle. \tag{9}$$

This relative phase is physically meaningful and can affect interference phenomena, making it observable in subsequent measurements.

All standard single-qubit phase gates can be understood as special cases of a general phase rotation acting only on the $|1\rangle$ component:

$$R_\phi = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}, \qquad R_\phi|\psi\rangle = \alpha|0\rangle + e^{i\phi}\beta|1\rangle. \tag{10}$$

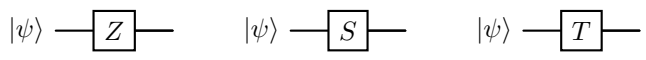
In circuit form, this general phase rotation is represented as a single-qubit gate:

$$|\psi\rangle \text{ —} \boxed{R_\phi} \text{—}$$

All the other phase operations are instances of this general transformation. In particular, the Z *gate* corresponds to a phase flip with $\phi = \pi$, the *Phase (S) gate* corresponds to $\phi = \frac{\pi}{2}$, and the T *gate* corresponds to $\phi = \frac{\pi}{4}$. Their matrix representations are:

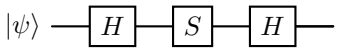
$$Z = R_\pi = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad S = R_{\pi/2} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \qquad T = R_{\pi/4} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}.$$

These gates are represented in circuit diagrams as follows:



Therefore, all these gates implement the same fundamental operation: they leave the $|0\rangle$ component unchanged while applying a controlled phase shift to the $|1\rangle$ component. This highlights that phase gates do not change measurement probabilities directly, but instead modify interference patterns, which is essential for many quantum algorithms.

Quantum gates can be composed to construct new quantum gates, or derive old ones. In particular, the composition of quantum gates HSH produces a linear transformation whose square equals the bit-flip operation. This composition can be visualized as the sequential application of three gates:



Exercise 3.5. *Prove that* $(HSH)^2 = \sigma_x$.

3.3 Measurements

A general measurement is described by a set of operators $\{M_m\}$, where each operator corresponds to a possible outcome m. The central requirement is that if we could apply all of these operators obtaining the probability of each outcome the sum of the probabilities for each outcome needs to be equal to 1.

Assume we have a collection of operators $\{M_m\}$ and we apply one of these M_m to a state $|\psi\rangle$:

$$|\phi_m\rangle = M_m|\psi\rangle. \tag{11}$$

The probability of $|\phi_m\rangle$ is given by the norm $\|M_m\,|\psi\rangle\,\|$. As we did in Section 3.2, the norm is computed using the inner product $\langle\psi|M_m^\dagger M_m|\psi\rangle$. The requirement is that the sum of all of these inner products must be 1, formally:

$$\sum_m \langle\psi|M_m^\dagger M_m|\psi\rangle = 1,$$

for all $|\psi\rangle$. Since the inner product preserves addition (2) and by factoring out ψ the former is equivalent to $\langle\psi|\left(\sum_m M_m^\dagger M_m\right)|\psi\rangle = 1$. Now since this has to be true for all $|\psi\rangle$, it has in particular, to be true for the computational basis $\{|i\rangle\}$, hence:

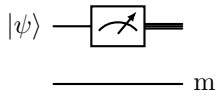
$$\sum_m M_m^\dagger M_m = I \tag{12}$$

ensuring that the total probability over all outcomes is 1.

If outcome m occurs, the measurement acts linearly on the state: $M_m\,|\psi\rangle$. In general, the resulting state is not normalized. The norm of the state is: $\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}$. To obtain a valid quantum state, we must normalize the vector. Therefore, the post-measurement state becomes:

$$|\psi_m\rangle = \frac{M_m\,|\psi\rangle}{\|M_m|\psi\rangle\|} = \frac{M_m\,|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \tag{13}$$

In circuit terms, one can think of the measurement as producing a classical outcome while collapsing the quantum state:



Intuitively, the operator M_m extracts the component of the state corresponding to outcome m. The norm of this component gives the probability of that outcome. After the measurement, the state is renormalized to reflect the fact that outcome m has occurred.

These observations lead to the third postulate of quantum mechanics:

Postulate 3. *Quantum measurements are described by a collection $\{M_m\}$ of measurement operators. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is $|\psi\rangle$ immediately before the measurement then the probability that result m occurs is given by*

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle.$$

The state after the measurement is

$$\frac{M_m\,|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}.$$

3.3.1 Measurement in the Computational Basis

The simplest kind of measurement is called measurement in the computational basis and it is obtained by using the measurement operators $M_i = |i\rangle\langle i|$. For example, for a state $|\psi\rangle = a\,|0\rangle + b\,|1\rangle$ we have the measurement operators $M_0 = |0\rangle\langle 0|$ and $M_1 = |1\rangle\langle 1|$. By using Postulate 3, the probability of the outcome 0 is obtained by computing $p(0)$:

$$p(0) = \langle\psi|M_0^\dagger M_0|\psi\rangle = \langle\psi| \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} |\psi\rangle = \langle\psi| \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} |\psi\rangle = a^* a = a^2$$

and the probability of the outcome 1 is obtained by computing $p(1)$ which is b^2.

After the measurement the state collapses to one of the two computational basis states.

Definition 3.2 (Measurement in the Computational Basis)**.** *Given a quantum state $|\psi\rangle$ in the computational basis $|e_i\rangle$, for a sequence $i \in \{1, \ldots, n\}$ defined as*

$$|\psi\rangle = \alpha_1\,|e_1\rangle + \cdots + \alpha_n\,|e_n\rangle$$

The measurement in the computational basis states that upon measurement the state $|\psi\rangle$ collapses to the state $|e_i\rangle$ with probability $\|\alpha_i\|^2$.

It follows directly that the sum of the probabilities for each outcome is $\sum_{j=0}^{n-1} \|\alpha_j\|^2 = 1$.

Measurement in quantum mechanics is the process of observing a quantum system, causing it to collapse into one of its basis states. For a qubit in the state $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$, measurement in the computational basis collapses the state $|\psi\rangle$ to either $|0\rangle$ with probability $|\alpha|^2$ or $|1\rangle$ with probability $|\beta|^2$.

For example, measuring the state $|\psi\rangle = \frac{1}{\sqrt{2}}\,|0\rangle + \frac{1}{\sqrt{2}}\,|1\rangle$ has equal probabilities of yielding $|0\rangle$ or $|1\rangle$, that is 50%. Similarly, measuring the state $|\psi\rangle = |0\rangle$ always yields $|0\rangle$ with 100% probability.

For an n-qubit system, the computational basis states are $|b_1 b_2 \ldots b_n\rangle$, where $b_i \in \{0, 1\}$. A measurement collapses the system to one of these 2^n states. Consider the two-qubit state:

$$|\psi\rangle = \frac{1}{\sqrt{2}}\,|00\rangle + \frac{1}{\sqrt{2}}\,|11\rangle\,.$$

If measured in the computational basis the system collapses to either $|00\rangle$ with probability of $|\frac{1}{\sqrt{2}}|^2 = 0.5$ or $|11\rangle$ with probability $|\frac{1}{\sqrt{2}}|^2 = 0.5$.

3.3.2 Projective Measurements

Definition 3.3 (Projective Measurement)**.** *A* projective *measurement with m possible outcomes, is a collection of projectors $P_1, \ldots, P_m$, pairwise orthogonal, that is $P_i P_j = 0$ if $i \neq j$, and, moreover, that sum identity $\sum_{j=1}^{m} P_j = I$.*

3.3.3 Distinguishing Quantum States

Distinguishing quantum states is a fundamental (and surprisingly subtle) problem in quantum mechanics. Suppose a quantum system is prepared in one of several known states:

$$\{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_n\rangle\} \tag{14}$$

The goal is to determine which state was prepared by performing a measurement. A quantum system is prepared in one of several known states $\{|\psi_1\rangle, |\psi_2\rangle, \ldots\}$, but the observer does not know which one. The goal is to determine the prepared state by performing a measurement.

The task is to design the measurement so that the outcome m correctly identifies the state $|\psi_i\rangle$ with high probability.

Perfect discrimination is possible if and only if the states are orthogonal. If the states are non-orthogonal, errors are unavoidable, and one must either minimize the probability of error, or allow for inconclusive outcomes. Thus, distinguishing quantum states is a problem of statistical inference: measurement outcomes provide partial information that is used to infer the unknown prepared state.

Assume we have two orthogonal states:

$$\langle\psi_1|\psi_2\rangle = 0 \tag{15}$$

and define measurement operators:

$$M_i = |\psi_i\rangle \langle\psi_i| \tag{16}$$

for $i \in \{1, 2\}$ which satisfying the properties

$$p(i|\psi_i) = 1 \tag{17}$$

$$p(j|\psi_i) = 0 \quad (j \neq i) \tag{18}$$

Assume we have two states $|\psi_1\rangle$ and $|\psi_2\rangle$ such that are not orthogonal, that is

$$\langle\psi_1|\psi_2\rangle \neq 0, \tag{19}$$

then perfect discrimination is impossible. The proof is given by contradiction. Suppose we attempt to distinguish $|\psi_1\rangle$ and $|\psi_2\rangle$ perfectly. Then there must exist measurement operators M_1, M_2 such that:

$$\langle\psi_1|M_2^\dagger M_2|\psi_1\rangle = 0 \tag{20}$$

$$\langle\psi_2|M_1^\dagger M_1|\psi_2\rangle = 0 \tag{21}$$

which implies:

$$M_2|\psi_1\rangle = 0, \quad M_1|\psi_2\rangle = 0 \tag{22}$$

By linearity, this leads to a contradiction unless the states are orthogonal.

If perfect discrimination is not possible, one can minimize the probability of error. For two states with prior probabilities p_1, p_2, the optimal error is given by the Helstrom bound:

$$P_{\text{error}} = \frac{1}{2}\left(1 - \sqrt{1 - 4p_1 p_2 |\langle\psi_1|\psi_2\rangle|^2}\right) \tag{23}$$

3.4 Multiple Quantum Systems

Postulate 4. *The state space of a composite physical system is the tensor product of the state spaces of the component physical system. Moreover, if we have systems numbered* $1, 2, \ldots n$*, and a system* I *is prepared in the state* $|\psi_i\rangle$ *then the joint state of the total system is*

$$|\psi_1\rangle \otimes |\psi_2\rangle \otimes \ldots \otimes |\psi_n\rangle$$

In quantum mechanics, when two quantum systems are combined, their states are described in a *tensor product* space. Suppose you have two quantum systems, A and B, with state spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. The state of the combined system is represented in the tensor product space:

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$$

The tensor product is a way of combining the state spaces of two systems into a larger state space. If $|\psi_A\rangle$ is a state in $\mathcal{H}_A$ and $|\psi_B\rangle$ is a state in $\mathcal{H}_B$, then the state of the combined system is written as:

$$|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$$

The tensor product allows for the creation of entangled states, which are not separable into individual states of A and B. These entangled states are crucial in quantum computation and quantum information theory.

3.4.1 The CNOT Gate

The Controlled-NOT (CNOT) gate is a two-qubit quantum gate acting on a *control* qubit c and a *target* qubit t. It flips the target qubit if and only if the control qubit is in state $|1\rangle$, and does nothing otherwise.

In the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, its matrix representation is:

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Exercise 3.6. *Convince yourself that the matrix CNOT changes the second qubit if the first one is set at* 1*.*

The construction of CNOT is equivalent to the expression

$$\text{CNOT} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \sigma_X,$$

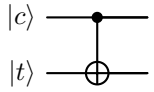
where σ_X is the Pauli-X (NOT) operator.

In an n-qubit system, the CNOT acting on qubits c and t is extended by tensoring identity operators on all other qubits:

$$U = I \otimes \cdots \otimes \text{CNOT} \otimes \cdots \otimes I.$$

However, if the control and target qubits are not adjacent, the operator must be embedded carefully, e.g., via qubit reordering or SWAP operations.

The circuit below represents a controlled-NOT (CNOT) operation acting on two qubits, arranged as horizontal lines evolving from left to right in time.



The upper wire corresponds to the *control qubit*, denoted $|c\rangle$, while the lower wire represents the *target qubit*, denoted $|t\rangle$. A filled black dot is placed on the control wire, indicating that the operation is conditioned on the state of this qubit. Specifically, the gate is activated only when the control qubit is in the state $|1\rangle$.

On the target wire, a circle containing a plus sign ($\oplus$) denotes a NOT (Pauli-X) operation, which flips the state of the qubit:

$$|0\rangle \leftrightarrow |1\rangle .$$

A vertical line connects the control dot to the target symbol, indicating that the NOT operation is applied conditionally.

The overall action of the CNOT gate is therefore:

$$|c, t\rangle \longrightarrow |c,\, t \oplus c\rangle ,$$

where $\oplus$ denotes addition modulo 2. In words, the target qubit is flipped if and only if the control qubit is in the state $|1\rangle$; otherwise, the system remains unchanged.

The CNOT gate is fundamental for entanglement generation and, together with single-qubit gates, forms a universal set for quantum computation.

3.5 Quantum Entanglement

Quantum entanglement is a phenomenon where two or more qubits become linked, so that the state of one qubit instantly affects the state of the other, no matter how far apart they are. Entangled qubits cannot be described independently. The combined state contains more information than the individual qubits.

The classic example is the Bell state is:

$$\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \tag{24}$$

whose measurement outcomes are perfectly correlated. It should be easy for the reader to see that if the first qubit of $|\Phi^+\rangle$ is measured as 0, then the second qubit is 0. Similarly, if the first 1 then the second is 1. The first qubit has 50% probability of being 0, but if it is so is the second bit, as the probability of the first qubit of being 0 and the second one 1 is 0% and similarly for the case when the first qubit is 1. Hence the two states are entangled: If we measure one qubit we automatically now what the second one is going to be.

We now show how to create entangled states. We start with two qubits in $|00\rangle = |0\rangle \otimes |0\rangle$ and apply the Hadamard gate to the first qubit, that is

$$H\,|0\rangle \otimes |0\rangle = |+\rangle \otimes |0\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle) \otimes |0\rangle = 1/\sqrt{2}(|00\rangle + |10\rangle).$$

Then we apply a CNOT gate with the first qubit control and the second qubit as target

$$\mathrm{CNOT}\left(\frac{|00\rangle + |10\rangle}{\sqrt{2}}\right) = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

In summary, the construction is given by

$$\mathrm{CNOT}(H\,|0\rangle \otimes |0\rangle)$$

Notice, we are secretly starting to do categorical quantum mechanics here [2].

We now formalise precisely the notion of entanglement by showing how the Bell state is measured. To measure the first qubit, we apply projectors on the first subsystem:

$$P_0 = |0\rangle\langle 0| \otimes I, \quad P_1 = |1\rangle\langle 1| \otimes I.$$

Both of these measurements are hermitian and unitary, hence $P_0^\dagger P_0 = P_0 P_0 = P_0^2 = P_0$ and similarly for P_1. The probability that the first qubit is 0 is given by

$$p(0) = \langle \Phi^+ | P_0^\dagger P_0 | \Phi^+ \rangle = \frac{1}{2}.$$

wherease the resulting state in case 0 is measured is given by the stated

$$P_0 |\Phi^+\rangle = \frac{1}{\sqrt{2}} |00\rangle.$$

which, after the post-measurement normalisation becomes $|00\rangle$. Hence measuring 0 for the first qubit results in the second qubit to be 0 as well.

Exercise 3.7. *Convince yourself that this is the same when measuring* 1

3.6 Coherence and Density Matrices

In quantum computing, *coherence* refers to the ability of a quantum system to maintain well-defined phase relationships between the components of a superposition. Coherence is responsible for many uniquely quantum phenomena, including interference and entanglement. Consider a single qubit in the state

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle ,$$

where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$.

This state is a superposition of the computational basis states $|0\rangle$ and $|1\rangle$. The relative phase between α and β carries physical significance. If this phase relationship is preserved, the system is said to be *coherent*.

Coherence allows quantum states to produce interference effects. For example, many quantum algorithms rely on constructive and destructive interference between amplitudes in order to amplify correct computational paths.

While pure states can be described using state vectors, many physical systems interact with their environment and cannot be described by a single wavefunction. In such cases we use the *density matrix* formalism.

For a pure state $|\psi\rangle$, the density matrix is defined as

$$\rho = |\psi\rangle \langle\psi| .$$

For the qubit state

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) ,$$

the density matrix becomes

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

In particular, the *diagonal elements* represent the probabilities of measuring the system in each basis state and the *off-diagonal elements* encode the phase relationships between components of the superposition. In the previous example, the nonzero off-diagonal terms

$$\rho_{01} = \rho_{10} = \frac{1}{2}$$

indicate that the system possesses coherence between the states $|0\rangle$ and $|1\rangle$. Hence the density matrix provides a convenient way to identify the presence of quantum coherence.

Now consider a situation where the system is not in a coherent superposition, but instead in a classical mixture: with probability $1/2$ the system is in $|0\rangle$ and with probability $1/2$ the system is in $|1\rangle$. The density matrix describing this mixture is

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

Here, the off-diagonal elements vanish. This indicates that the phase relationship between the states has been lost, and therefore the system no longer exhibits quantum coherence.

In practical quantum systems, interaction with the environment gradually destroys coherence. This process is known as *decoherence*. Mathematically, decoherence suppresses the off-diagonal elements of the density matrix over time:

$$\rho = \begin{pmatrix} p_0 & c(t) \\ c^*(t) & p_1 \end{pmatrix} \quad \longrightarrow \quad \begin{pmatrix} p_0 & 0 \\ 0 & p_1 \end{pmatrix}.$$

As the coherence terms $c(t)$ approach zero, the system behaves more like a classical probabilistic system.

It is important to note that coherence is *basis dependent*. A density matrix that is diagonal in one basis may contain off-diagonal terms in another basis. Therefore, statements about coherence must always specify the basis with respect to which the density matrix is written.

Maintaining coherence is one of the central challenges in the physical realization of quantum computers. In quantum mechanics, a *state* can either be *pure* or *mixed*. A *pure state* is described by a single vector in a Hilbert space. It represents a quantum system that is in a definite state. In the case of a composite system, a pure state can be written as a *product state or tensor state* or an *entangled state*.

The density matrix for a pure state has the following properties:

- It is *idempotent*: $\rho^2 = \rho$.
- It has *trace 1*: $\text{trace}(\rho) = 1$, ensuring that the total probability is normalized.
- It is a *rank-1* matrix if the state is pure, meaning it can be written as a single outer product.

A *mixed state* represents a quantum system that is in a statistical mixture of pure states, which means the system's state is not known exactly. Mixed states arise when there is some uncertainty or lack of information about the system, often due to decoherence or interaction with an environment.

Mathematically, a mixed state is described by a *density matrix* ρ, which is a weighted sum of the outer products of pure states:

$$\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|$$

where p_i are probabilities and $|\psi_i\rangle$ are pure states. The density matrix for a mixed state satisfies $\text{trace}(\rho) = 1$ and ρ is Hermitian, that is $\rho^\dagger = \rho$.

Hence, the *density operator* (or *density matrix*) is a more general representation of a quantum state, used to describe both *pure* and *mixed* states. It is especially useful when dealing with situations where the system is in a statistical mixture of states or when the system is entangled with an environment, which is common in quantum computation.

For example, the expectation value of an observable $\hat{O}$ for a quantum state ρ is given by:

$$\left|\hat{O}\right\rangle = \text{trace}(\rho\hat{O})$$

This is a generalization of the expectation value for pure states, where $\langle\hat{O}\rangle = \langle\psi|\hat{O}|\psi\rangle$.

A *mixed state* represents a probabilistic mixture of pure states. If the system is in one of several pure states $|\psi_i\rangle$ with probability p_i, the density operator for the mixed state is:

$$\rho = \sum_i p_i \, |\psi_i\rangle \, \langle\psi_i|$$

where p_i are the probabilities that the system is in the pure state $|\psi_i\rangle$. Note that, the density matrix for a mixed state is *not idempotent* (i.e., $\rho^2 \neq \rho$), it has *more than one non-zero eigenvalue* and may have a rank greater than 1.

The density matrix for a mixed state has the following properties:

- It is *positive semi-definite*: $\langle\phi|\rho|\phi\rangle \geq 0$ for any state $|\phi\rangle$.
- It has *trace 1*: $\mathrm{Tr}(\rho) = 1$

3.7 The Schrödinger Equation (*)

Postulate 2' *The time evolution of the state of a closed quantum system is described by the Schrödinger equation*

$$i\hbar \frac{d\,|\psi(t)\rangle}{dt} = H\,|\psi(t)\rangle \tag{25}$$

where $\hbar$ is the Planck's constant and H is an Hermitian operator called the Hamiltonian of the closed system.

The Schrödinger Equation essentially postulates that the evolution in time of a quantum system is given by an Hermitian operator. To see why this is an equivalent version of Postulate 2 we have to solve the differential equation given by Schrödinger. The solution is give by

$$|\psi(t)\rangle = e^{-\frac{itH}{\hbar}}\,|\psi(0)\rangle\,. \tag{26}$$

We are going to prove that this is indeed a solution to Equation 25, but first notice that if we set $U(t) = e^{-ihT/\hbar}$ then $|\psi(t)\rangle = U(t)\,|\psi(0)\rangle$ which is a more direct way to calculate the evolution of a quantum state at the time t starting from an initial quantum state $|\psi(0)\rangle$. If H is Hermitian, namely $H^\dagger = H$ by definition of Hermitian matrices, then $U(t)$ is unitary. To see this it suffices to compute

$$U^\dagger(t)U(t) = I$$

This is easy by noticing that $U^\dagger(t) = (e^{-iHt/\hbar})^\dagger = e^{iHt/\hbar}$, then $e^{iHt/\hbar}e^{-iHt/\hbar} = I$.

We now prove that (26) is a solution to Equation 25.

Proof of 25. The first step is to calculate the derivative of $|\psi(t)\rangle$:

$$\frac{d\,|\psi(t)\rangle}{dt} = \frac{-iH}{\hbar} \cdot e^{-\frac{itH}{\hbar}}\,|\psi(0)\rangle\,. \tag{27}$$

By substituting this derivative into Equation 25 we get

$$i\hbar \frac{-iH}{\hbar} \cdot e^{-\frac{itH}{\hbar}}\,|\psi(0)\rangle = H\,|\psi\rangle\,.$$

This can be simplified to

$$H \cdot \underbrace{e^{-\frac{itH}{\hbar}}\,|\psi(0)\rangle}_{|\psi(t)\rangle} = H\,|\psi\rangle\,.$$

Now notice that $e^{-\frac{itH}{\hbar}}\,|\psi(0)\rangle$ is $|\psi(t)\rangle$, hence the equation above can be rewritten as

$$H \cdot |\psi(t)\rangle = H\,|\psi\rangle$$

as wanted. □

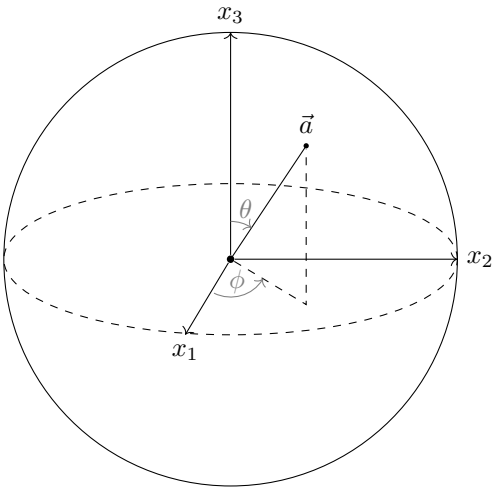


Figure 12: Bloch Sphere

3.8 The Bloch Sphere (*)

Consider qubit written as a linear combination of the computational basis

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \quad |\alpha|^2 + |\beta|^2 = 1.$$

Although α and β are complex numbers, the state can be expressed using two real parameters:

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

The reader who is not familiar with this conversion is encouraged to take a look at the appendix, Section A.4, for a more detailed explanation of how this works.

The Bloch sphere is a way of representing qubits graphically. In Figure 12 a vector corresponds to a point on a unit sphere. The north pole is $|0\rangle$, the south pole is $|1\rangle$ and any other point is a vector in superposition. Thus, every pure qubit state is a point on the surface of the sphere whereas mixed states lie inside the sphere.

References

[1] Howard Anton and Chris Rorres. *Elementary Linear Algebra: Applications Version.* Wiley, eleventh edition, 2014.

[2] Chris Heunen and Jamie Vicary. Introduction to categorical quantum mechanics, February 2013. Lecture notes.

[3] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

A Elements of Set Theory

A *set* (or class) is an unordered collection of objects called *elements* of the set. We write $a \in X$ when a is an element of the set X and we read it as "a belongs to X". Here are some important sets:

- $\emptyset$ the empty set with no elements. The reader should ponder about the difference between the empty set $\emptyset$ and the singleton set $\{\emptyset\}$ containing the empty set.
- $\mathbb{N}$ the set of *natural numbers* $\{1, 2, 3, \dots\}$
- $\mathbb{N}_0$ the set of *natural numbers with zero* $\{0,1,2,3,\dots\}$
- $A \times B$ the *cartesian product* of sets A and B is the set containing the pairs (a, b) such that $a \in A$ and $b \in B$.
- $A \cup B$ the *union* of sets A and B is the set containing all elements of A and of B. If there are equal elements in A and B, these become squashed together.
- $A \uplus B$ the *disjoint union* of sets A and B is the set containing all elements $(1, a)$ for $a \in A$ and $(2, b)$ for $b \in B$.

A *relation* $R \subseteq A \times B$ is a subset of the cartesian product $A \times B$ relating some elements in A with some elements in B. A *function* is a relation $f \subseteq A \times B$ such that for an $a \in A$, there exists only one $b \in B$. We write the set of functions between A and B as $A \to B$.

There is only one function from the empty set $\emptyset$ into any other set A, that is the empty relation denoted by $! \subseteq \emptyset \times A$. Dually, there is only one function from any set A to the singleton set $\{*\}$ for an element $*$. That is the function sending every $a \in A$ into $*$. We denote this map $!$ as well, although it should be clear from the context which one we mean.

An important notion in set theory is the one of *size*, which indicates the *cardinality* of a set. Two sets are said to be *isomorphic* when they have the same cardinality. An isomorphism is given by a function $f : A \to B$ and its *inverse* $f^{-1} : B \to A$ such that $f(f^{-1}(x)) = x$ and $f^{-1}(f(x)) = x$.

A set A is *finite* if and only if it is isomorphic to the set $\{m \in \mathbb{N} \mid m \leq n\}$ for some $n \in \mathbb{N}$. If this is the case, it means we can enumerate the elements of A and write A as $\{a_1, a_2, a_3, \dots, a_n\}$. This is because intuitively we could write down its elements on a piece of paper in a finite amount of time. We say a set is *infinite* if and only if it is not finite. A set is *countable* if it is isomorphic to the natural numbers.

For a set I, we denote the family of sets indexed by I as $\{A_i\}_{i \in I}$. In the case that I is finite, the family $\{A_i\}_{i \in I}$ is finite, and we can write both the finite union and product of these sets as follows:

$$A_1 \cup A_2 \cup \cdots \cup A_n$$

and similarly for the product. In the case that I is infinite, the union of the infinite family of sets can be written simply as:

$$\bigcup_{i \in I} A_i = \{a \in A_i \mid i \in I\}$$

Dually, the *infinite product* or *dependent product* is defined by the set of functions that, given an index $i \in I$, return an element in A_i:

$$\Pi_{i \in I} A_i = \{f : I \to \bigcup_{i \in I} A_i \mid f(i) \in A_i\}$$

A.1 Basic Trigonometry

The sine (sin) and cosine (cos) functions are fundamental trigonometric functions. The unit circle provides a geometric interpretation of sin and cos. Figure 13 illustrates the unit circle with an angle θ, where:

$$\cos(\theta) = x \text{ (horizontal projection)}, \quad \sin(\theta) = y \text{ (vertical projection)}.$$

The *domain* of $\sin(x)$ and $\cos(x)$ is all real numbers: $x \in \mathbb{R}$ while the range of both functions is $[-1, 1]$. Moreover, both functions are periodic with a period of 2π:

$$\sin(x + 2\pi) = \sin(x), \quad \cos(x + 2\pi) = \cos(x).$$

this means that a periodic point in time (the multiples of 2π the functions go back to their starting value.

The function $\sin(x)$ is an odd function while the $\cos(x)$ function is even, meaning that :

$$\sin(-x) = -\sin(x) \qquad \cos(-x) = \cos(x).$$

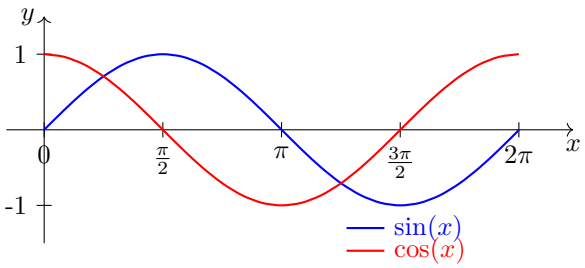
Proposition A.1 (Pythagora's identity)**.** *The sine and cosine functions satisfy the fundamental Pythagorean identity:*

$$\sin^2(\theta) + \cos^2(\theta) = 1.$$

where θ is an angle.

We list below the values for particular key angles:

$$\sin(0) = 0, \quad \sin\left(\frac{\pi}{2}\right) = 1, \quad \sin(\pi) = 0, \quad \sin\left(\frac{3\pi}{2}\right) = -1, \quad \sin(2\pi) = 0,$$
$$\cos(0) = 1, \quad \cos\left(\frac{\pi}{2}\right) = 0, \quad \cos(\pi) = -1, \quad \cos\left(\frac{3\pi}{2}\right) = 0, \quad \cos(2\pi) = 1.$$



Finally, the sine and cosine functions satisfy the following addition formulas:

$$\sin(a + b) = \sin(a)\cos(b) + \cos(a)\sin(b),$$

$$\cos(a + b) = \cos(a)\cos(b) - \sin(a)\sin(b).$$

Step 3: Parametrize the Magnitudes with θ

After factoring out the global phase, the qubit can be written as

$$|\psi\rangle = |\alpha|\,|0\rangle + |\beta|\,e^{i\phi}|1\rangle,$$

with $|\alpha|, |\beta| \geq 0$ and $|\alpha|^2 + |\beta|^2 = 1$.

Introducing the Bloch Sphere Angle θ: To map this qubit onto the Bloch sphere, we parametrize the magnitudes using a single angle $\theta \in [0, \pi]$:

$$|\alpha| = \cos\frac{\theta}{2}, \quad |\beta| = \sin\frac{\theta}{2}.$$

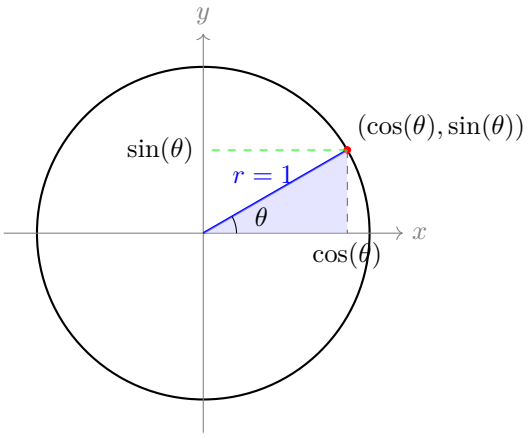


Figure 13: The Unit Circle

Why $\theta/2$?

- Using $\theta/2$ ensures the state remains normalized automatically:

$$|\alpha|^2 + |\beta|^2 = \cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2} = 1.$$

- It also aligns the polar angle θ with the standard Bloch sphere representation:
  - $\theta = 0 \implies |\psi\rangle = |0\rangle$ (north pole)
  - $\theta = \pi \implies |\psi\rangle = |1\rangle$ (south pole)
  - $0 < \theta < \pi \implies$ superposition between $|0\rangle$ and $|1\rangle$

Intuition: This is analogous to a point on the unit circle in 2D, where

$$x^2 + y^2 = 1 \implies x = \cos\theta,\ y = \sin\theta.$$

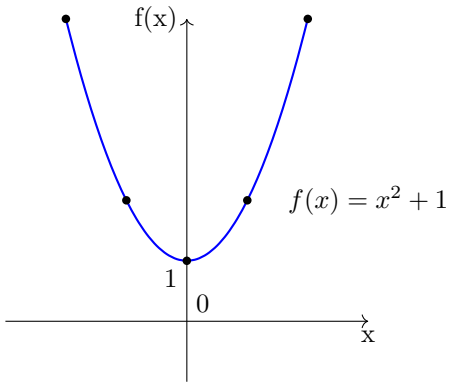
Here, $|\alpha|$ and $|\beta|$ are like x and y, and the factor of $1/2$ is a conventional choice to match the Bloch sphere representation.

A.2 Complex Numbers

Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as follows:

$$f(x) = x^2 + 1$$

Graphically, this function can be viewed as in the following picture



Obviously, if we ask the question, "Is there a value for x such that $f(x) = 0$" the answer would be negative: there is no x at which point the function f returns 0 because we would need to solve the equation $x^2 = -1$ and we know that there is no real number which multiplied by itself would yield a negative number. However, for some kinds of mathematical problems, such as the ones that we are going to treat in these notes, it would be useful to know that every function has a solution of this kind.

Complex numbers were invented to deal with this type of scenario by just assuming the square root of -1 exists. This imaginary number is called i. Now, if we take a look at the equation above again, we can compute $x^2 + 1 = 0$, which implies $x^2 = -1$, and thus $x = \sqrt{-1} = i$. Hence, i is our solution to the equation.

Algebraic Form Any complex number is formed by a real part and an imaginary part and is written as

$$z = a + bi \tag{28}$$

with $a, b \in \mathbb{R}$. Operations include addition, subtraction, multiplication, and division.

Polar Form A complex number $z = a + bi$ can also be represented in *polar form*, which expresses the number in terms of its *modulus* and *argument*.

In the complex plane, the complex number $z = a + bi$ is represented as a point (a, b). The modulus r is the distance from the origin to this point and the argument θ is the angle formed with the positive real axis. The polar form provides a more convenient way to perform operations such as multiplication and division of complex numbers.

The diagram in Figure 14 illustrates a complex number $z = a + bi$ in the complex plane, where a is the real part ($\mathrm{Re}(z)$), b is the imaginary part ($\mathrm{Im}(z)$), $|z| = \sqrt{a^2 + b^2}$ is the magnitude (or modulus) of z and $\theta = \arg(z)$ is the argument (angle) of z. Clearly $a = \cos\theta$ and $b = \sin\theta$. So we can rewrite (28) as

$$z = r\left(\cos(\theta) + i\sin(\theta)\right) \tag{29}$$

Exponential Form The exponential form is based on the following theorem.

Theorem A.1 (Euler's Formula)**.** *For any real number θ, the following equation holds:*

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

where:

- *e is Euler's number (the base of the natural logarithm),*
- *i is the imaginary unit ($i^2 = -1$),*

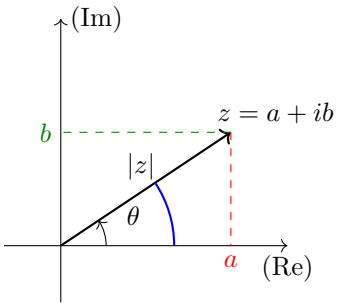


Figure 14: A geometric representation of complex numbers.

- *θ is a real number representing the angle (in radians),*

Thus, the polar form

$$z = r(\cos(\theta) + i\sin(\theta))$$

can be conveniently written using Euler's formula as:

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

where $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ by Euler's formula.

A.3 The Fundamental Theorem of Algebra

As we mentioned above, when we work with complexes every polynomial function has a solution. This is made formal by the fundamental theorem of algebra and it is going to be a crucial result for describing models of quantum computation:

Theorem A.2 (Fundamental Theorem of Algebra)**.** *Every non-constant polynomial with complex coefficients has at least one complex root. In other words, for any polynomial $p(z)$ of degree $n \geq 1$, where*

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0,$$

with $a_n \neq 0$, there exists at least one complex number z_0 such that

$$p(z_0) = 0.$$

A.4 From Complex to Real Linear Combinations

Often it is useful to consider a vector with complex coefficients and convert into a vector with real coefficients. The key idea behind this transformation is that a complex number corresponds to a 2D real vector:

$$a + ib \;\leftrightarrow\; (a, b)$$

thus there is an isomorphism of vector spaces

$$\mathbb{C}^n \cong \mathbb{R}^{2n}$$

In words, a vector of n complex entries and be represented as a vector of $2n$ entries where there first n are the real parts of the entries and the second half are the imaginary parts.

More precisely, let $|\psi\rangle$ be a vector written as a linear combination:

$$|\psi\rangle = \sum_k c_k |k\rangle, \quad c_k \in \mathbb{C}$$

Each coefficient can be decomposed into real and imaginary parts:

$$c_k = a_k + ib_k, \quad a_k, b_k \in \mathbb{R}$$

Substituting:

$$|\psi\rangle = \sum_k (a_k + ib_k)|k\rangle = \sum_k a_k|k\rangle + i\sum_k b_k|k\rangle$$

Thus the vector separates into a real and an imaginary component. We can represent the complex vector as a real vector of dimension $2n$:

$$|\psi\rangle \longleftrightarrow \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^{2n}$$

The first n entries are the real parts, and the last n entries are the imaginary parts. Each basis vector $|k\rangle$ corresponds to two real directions:

$$|k\rangle \mapsto (|k\rangle_{\mathrm{Re}}, |k\rangle_{\mathrm{Im}})$$

Example Consider the vector

$$|\psi\rangle = (1+2i)|0\rangle + (3-i)|1\rangle,$$

then,

$$c_0 = 1 + 2i \to (1, 2), \quad c_1 = 3 - i \to (3, -1)$$

This the real representation is:

$$\begin{pmatrix} 1 \\ 3 \\ 2 \\ -1 \end{pmatrix}$$

Note that although the vector now has only real entries the multiplication by i becomes a rotation between real and imaginary parts and linear operators become real matrices of size $2n \times 2n$.