SML (Mathematics of Quantum Information Theory), Summary Notes

Chris Bourne[1]

Spring Semester, 2025

The following are work-in-progress notes that summarise the lectures for the 'Special Mathematics Lecture (Mathematics of Quantum Information Theory)'. The content is mostly taken from the texts [3, 4]. These notes should not be considered as complete lecture notes as many details are still missing. Please also be wary of typos and other mistakes. Any corrections or suggestions are greatly appreciated ☺.

Quantum circuit diagrams have been made with the 'quantikz' LaTeX package, available for download at `arXiv:1809.03842`.

Contents

[1] cbourne@nagoya-u.jp

1 Preface

The aim of this Special Mathematics Lecture (SML) is to introduce the mathematical framework used to study quantum computing and quantum information. In order to do this, we will also need to introduce the basic postulates of quantum mechanics. The mathematics of quantum mechanics is quite complicated and often involves so-called infinite-dimensional Hilbert spaces. But if our interest is in quantum computing and information, we can restrict our attention to finite-dimensional spaces like $\mathbb{C}^n$, which makes many aspects much easier.

Let us briefly introduce a few important concepts that we will further study in the lectures.

Quantum mechanics

Very loosely speaking, quantum mechanics is a theory about the behaviour of very small objects (electrons, atoms, etc.). It is a theory developed by Heisenberg, Schrödinger, von Neumann, Dirac and many others in the 1920s and 1930s. The physics of quantum phenomenon necessitated the development of new *mathematical* constructions such as Hilbert spaces, (quantum) states and (quantum) observables. Many parts of quantum mechanics do not agree with our intuition, which is based on our observations of the macroscopic world. However, the experimental evidence in favour of quantum mechanics is substantial so, despite its counter-intuitive nature, its foundations remain almost universally accepted 100 years later.

A key part of quantum mechanics is the way in which measurement is understood, both physically and mathematically. Physically, measurement of a quantum system is not a neutral operation and will change the nature of the system itself. Mathematically, given a quantity we want to observe, we can understand its possible outcomes and the probability that each will occur. These outcomes and probabilities are closely related to the eigenvalues and eigenvectors of the mathematical object (a linear operator) that describes the observable.

A particularly non-intuitive but important aspect of quantum mechanics is the notion of quantum entanglement, where parts of a quantum system are intrinsically linked, even if they are very far from each other. Mathematically, we understand quantum entanglement via so-called tensor products of Hilbert spaces and what this implies for the states (vectors) in this composite system.

Quantum computing

One way we can understand quantum computing is as an application of quantum mechanics to information theory, a mathematical description of the transmission of data/information.

A classical computer performs operations (mathematical operations, algorithms, ...) on data that is represented by bits, strings of 0 or 1. A *quantum* computer is a device for computation that uses distinctly quantum mechanical phenomena (superposition, entanglement, ...) to perform operations on data represented by *quantum* bits (Qubits), strings of $|0\rangle$ or $|1\rangle$. Both methods of computation have similar underlying principles, using algorithms to transform data and communicate information. But the setting and implementation are very different. This means that one form of computation may be faster/more efficient at performing a specific task than another. As we will see, there are some tasks/algorithms that can be performed much more efficiently by a quantum computer.

What we will *not* discuss, is how to construct and implement a quantum computer in the real world, though remark that there are many different approaches such as adiabatic quantum computing and topological quantum computing with distinct theoretical foundations.

Quantum information

Roughly speaking, quantum computing is focused on the operations one can do on systems of Qubits in order to implement particular algorithms or solve particular problems. More generally, we can study what information/data is contained in a system of Qubits and what operations preserve this information. Put another way, if a Qubit system is described by a quantum state (or, more generally, a density matrix), what are the operations on such states/density matrices that preserve the information/data that is to be transmitted. A careful study of such operations is required to carefully analyse if errors have occurred in a quantum transmission and how much the information has been altered. With such information, one can try and implement error-correcting quantum codes to counteract these errors.

At the time of writing, it is unclear how much we will be able to say about the wider theory of quantum information within the context of this SML, but the interested reader can consult [3, Part III], which is the 'standard' reference.

A motivating example: (simplified) Deutsch's algorithm

As previously stated, a classical bit is an element $x \in \{0,1\}$. Suppose we have a function $f : \{0,1\} \to \{0,1\}$ and want to know if $f(0) = f(1)$ or $f(0) \neq f(1)$. Using classical bits, this takes two steps: compute $f(0)$, then compute $f(1)$ and compare the results.

If we use a system of two Qubits, then we can answer this question in one step. We will outline this process without proper definitions (which will come later in the text). A generic Qubit is an element

$$|x\rangle = a|0\rangle + b|1\rangle := a\begin{pmatrix}1\\0\end{pmatrix} + b\begin{pmatrix}0\\1\end{pmatrix} \in \mathbb{C}^2, \qquad a,b\in\mathbb{C}, \quad |a|^2+|b|^2 = 1.$$

Before we consider a 2-Qubit system, we briefly introduce the Hadamard transformation on Qubits, which is a linear and unitary operator $H : \mathbb{C}^2 \to \mathbb{C}^2$ such that

$$H|0\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big), \quad H|1\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle - |1\rangle\big), \qquad H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \text{ in matrix form.}$$

Note also that $H^2 = \mathbf{1}_2$ and so H will send the vectors $\frac{1}{\sqrt{2}}(|0\rangle + (-1)^y|1\rangle)$ to $|y\rangle$ for $y \in \{0,1\}$.

We now consider a composite system. A 2-Qubit vector is given by the product

$$\begin{aligned}|x,y\rangle = (|x\rangle)(|y\rangle) = |x\rangle\otimes|y\rangle &= \big(a|0\rangle + b|1\rangle\big)\otimes\big(c|0\rangle + d|1\rangle\big)\\ &= ac|0,0\rangle + ad|0,1\rangle + bc|1,0\rangle + bd|1,1\rangle \in \mathbb{C}^2\otimes\mathbb{C}^2, \quad a,b,c,d\in\mathbb{C}.\end{aligned}$$

We will precisely define the tensor product $\otimes$ later. Given the function $f : \{0,1\} \to \{0,1\}$, we can define the (unitary) operation on 2-Qubits,

$$U_f : \mathbb{C}^2\otimes\mathbb{C}^2 \to \mathbb{C}^2\otimes\mathbb{C}^2, \qquad U_f\big(|x,y\rangle\big) = |x, y\oplus f(x)\rangle,$$

where $y \oplus f(x) \in \{0,1\}$ denotes addition modulo 2. That is,

$$U_f(|0,0\rangle) = |0,f(0)\rangle, \quad U_f(|0,1\rangle) = |0,1\oplus f(0)\rangle, \quad U_f(|1,0\rangle = |1,f(1)\rangle, \quad U_f(|1,1\rangle) = |1,1\oplus f(1)\rangle.$$

We wish to apply U_f on a specific 2-Qubit vector, namely,

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

To help compute U_f of this vector, we first note that

$$\begin{aligned}U_f\big(|x\rangle(\tfrac{1}{\sqrt{2}}|0\rangle - \tfrac{1}{\sqrt{2}}|1\rangle)\big) &= \tfrac{1}{\sqrt{2}}\big(|x,f(x)\rangle - |x,1\oplus f(x)\rangle\big)\\ &= \begin{cases}\tfrac{1}{\sqrt{2}}(|x,0\rangle - |x,1\rangle), & f(x)=0,\\ \tfrac{1}{\sqrt{2}}(-|x,0\rangle + |x,1\rangle), & f(x)=1,\end{cases}\\ &= (-1)^{f(x)}|x\rangle(\tfrac{1}{\sqrt{2}}|0\rangle - \tfrac{1}{\sqrt{2}}|1\rangle).\end{aligned}$$

We now consider the operation U_f on the 2-Qubit $|x\rangle|y\rangle = \frac{1}{2}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle)$. A computation will give that

$$\begin{aligned} U_f\big(\tfrac{1}{2}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle)\big) &= \tfrac{1}{2}\big((-1)^{f(0)}|0\rangle(|0\rangle - |1\rangle) + (-1)^{f(1)}|1\rangle(|0\rangle - |1\rangle)\big) \\ &= \begin{cases} \frac{1}{2}(-1)^{f(0)}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle), & f(0) = f(1), \\ \frac{1}{2}(-1)^{f(0)}(|0\rangle - |1\rangle)(|0\rangle - |1\rangle), & f(0) \neq f(1). \end{cases} \end{aligned}$$

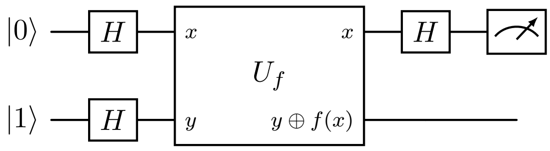
Finally, we apply the Hadmard transformation on the first Qubit only,

$$\begin{aligned} (H \otimes \mathbf{1})U_f\big(\tfrac{1}{2}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle)\big) &= \begin{cases} \frac{1}{2}(-1)^{f(0)}H(|0\rangle + |1\rangle)(|0\rangle - |1\rangle), & f(0) = f(1), \\ \frac{1}{2}(-1)^{f(0)}H(|0\rangle - |1\rangle)(|0\rangle - |1\rangle), & f(0) \neq f(1). \end{cases} \\ &= \begin{cases} \frac{1}{\sqrt{2}}(-1)^{f(0)}(|0\rangle)(|0\rangle - |1\rangle), & f(0) = f(1), \\ \frac{1}{\sqrt{2}}(-1)^{f(0)}(|1\rangle)(|0\rangle - |1\rangle), & f(0) \neq f(1). \end{cases} \\ &= \tfrac{1}{\sqrt{2}}(-1)^{f(0)}|f(0) \oplus f(1)\rangle(|0\rangle - |1\rangle) \end{aligned}$$

We can then measure the first Qubit to obtain the value $f(0) \oplus f(1) \in \{0, 1\}$ and determine if $f(0) = f(1)$ or not. We can combine all these process into one step via an application of the quantum *circuit*,

$$(H \otimes \mathbf{1})U_f(H \otimes H)|0, 1\rangle = \tfrac{1}{\sqrt{2}}(-1)^{f(0)}|f(0) \oplus f(1)\rangle(|0\rangle - |1\rangle)$$

We also represent the quantum circuit by the following diagram (which will be further explained later)



The difference between classical and quantum methods become more apparent in higher dimensions. Suppose $f : \{0, 1\}^n \to \{0, 1\}$ is a function that is either constant $f(\mathbf{x}) = c \in \{0, 1\}$ for all $\mathbf{x} \in \{0, 1\}^n$ or is balanced: takes value 0 for half of $\{0, 1\}^n$ and 1 for the other half. Classical methods require at least $2^{n-1} + 1$ queries to know with complete certainty if f is constant or balanced. On the other hand, the Deutsch–Joza algorithm solves this question by *one* operation of a quantum circuit built from H and U_f on a particular choice of n-Qubit. So there is an exponential difference in terms of steps taken.

The above example one case where 'quantum' methods can be used to obtain interesting and surprising results. There are many other surprising examples such as quantum teleportation and superdense coding that we will cover in the notes that follow.

2 Linear algebra review

We first review some points from linear algebra. Indeed, much of the mathematical content of quantum mechanics and quantum computing is directly related to complex vector spaces and linear operators on these spaces.

2.1 Hilbert spaces

The setting of quantum mechanics and quantum computing is a complex Hilbert space. A review of the complex numbers and its elementary properties are given in Appendix A.

Definition 2.1. A complex Hilbert space is a vector space $\mathcal{H}$ over the field $\mathbb{C}$ with a sesquilinear inner product $\langle \cdot \mid \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ with the properties

$$\begin{aligned} &\langle \psi \mid \phi \rangle = \overline{\langle \phi \mid \psi \rangle}, \\ &\langle \psi \mid a\phi_1 + b\phi_2 \rangle = a\langle \psi \mid \phi_1 \rangle + b\langle \psi \mid \phi_2 \rangle \\ &\langle \psi \mid \psi \rangle \geq 0 \quad \text{and} \quad \langle \psi \mid \psi \rangle = 0 \quad \iff \quad \psi = 0 \end{aligned}$$

for all $\psi, \phi, \phi_1, \phi_2 \in \mathcal{H}$ and $a, b \in \mathbb{C}$. Furthermore, the inner product induces a norm $\|\psi\| = \sqrt{\langle \psi \mid \psi \rangle}$ such that $\mathcal{H}$ is complete.

Remarks 2.2.

1. The conditions on the complex inner-product imply that $\langle a\psi \mid \phi \rangle = \overline{a}\langle \psi \mid \phi \rangle$ for any $a \in \mathbb{C}$.
2. For a norm to be complete, any sequence $\{\psi_n\}_{n\geq 0}$ in $\mathcal{H}$ such that $\|\psi_n - \psi_m\| < \epsilon$ for n, m large enough implies that there exists $\psi \in \mathcal{H}$ such that $\|\psi_n - \psi\| \to 0$.

To prove that $\|\psi\| = \sqrt{\langle \psi \mid \psi \rangle}$ is a norm, one uses the Cauchy–Schwarz inequality, whose proof is an exercise.

Lemma 2.3 (Cauchy–Schwarz inequality)**.** *For any $\psi, \phi \in \mathcal{H}$,*

$$\big|\langle \psi \mid \phi \rangle\big| \leq \|\psi\| \, \|\phi\|.$$

Exercise 2.1. Show that if $\psi_n \to \psi$ in $\mathcal{H}$, then for any $\phi \in \mathcal{H}$, $\langle \phi \mid \psi_n \rangle \to \langle \phi \mid \psi \rangle$ in $\mathbb{C}$.

Example 2.4. For any natural number $n \geq 1$, $\mathbb{C}^n$ is a complex Hilbert space with the obvious vector space structure,

$$a \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} + b \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} az_1 \\ az_2 \\ \vdots \\ az_n \end{pmatrix} + \begin{pmatrix} bw_1 \\ bw_2 \\ \vdots \\ bw_n \end{pmatrix}.$$

The inner-product is similar to that on $\mathbb{R}^n$, but with a complex conjugate on the left,

$$\left\langle \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \,\middle|\, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle = \sum_{j=1}^{n} \overline{z_j} w_j.$$

An inner-product lets us consider orthogonality and the angles between vectors. We say ψ and $\phi \in \mathcal{H}$ are orthogonal if $\langle \psi \mid \phi \rangle = 0$. Similarly, if $\mathcal{V} \subset \mathcal{H}$ is a vector subspace, we can define

$$\mathcal{V}^\perp = \big\{\phi \in \mathcal{H} \,:\, \langle \psi \mid \phi \rangle = 0 \text{ for all } \psi \in \mathcal{V}\big\}.$$

Exercise 2.2. Let $\mathcal{V} \subset \mathcal{H}$ be a closed vector subspace. Show that any vector $\phi \in \mathcal{H}$ can be decomposed as a sum $\phi = \phi_1 + \phi_2$ with $\phi_1 \in \mathcal{V}$, $\phi_2 \in \mathcal{V}^\perp$ and, hence, $\langle \phi_1 \mid \phi_2 \rangle = 0$. Put another way, there is a decomposition $\mathcal{H} = \mathcal{V} \oplus \mathcal{V}^\perp$.

Any vector space has a basis, a set of linearly independent vectors $\{\varphi_j\}_{j\in J} \subset \mathcal{H}$ such that any element $\psi \in \mathcal{H}$ can be decomposed as a linear span of basis elements,

$$\psi = \sum_{j\in J} a_j \varphi_j, \qquad a_j \in \mathbb{C} \text{ for all } j \in J.$$

Using the inner-product, we will generally work with an *orthonormal basis*, a basis $\{e_j\}_{j\in J}$ such that

$$\langle e_j \mid e_k \rangle = \delta_{j,k} = \begin{cases} 1, & j = k, \\ 0, & \text{otherwise.} \end{cases}$$

The Gram–Schmidt process can turn any basis into an orthogonal basis. We can also normalise any non-zero vector $\psi \mapsto \frac{1}{\|\psi\|}\psi$ to have unit length.

Example 2.5. The Hilbert space $\mathbb{C}^n$ has the canonical orthonormal basis

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \cdots, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

For quantum computing and quantum information theory, it generally suffices to only consider finite-dimensional Hilbert spaces, i.e. Hilbert spaces with a finite basis. But for many systems in quantum mechanics, infinite-dimensional spaces are needed. An important example is

$$L^2(\mathbb{R}^3) = \Big\{ f : \mathbb{R}^3 \to \mathbb{C} \,\Big|\, \int_{\mathbb{R}^3} |f(x)|^2 \, \mathrm{d}x < \infty \Big\}.$$

Exercise 2.3. Let $\{e_j\}_{j\in J}$ be an orthonormal basis of a Hilbert space $\mathcal{H}$. Show that the expansion of an element $\psi \in \mathcal{H}$ with respect to this basis is unique

$$\psi = \sum_j a_j e_j = \sum_j b_j e_j \quad \Longrightarrow \quad a_j = b_j \text{ for all } j.$$

Furthermore, show that the complex coefficients $a_j = \psi_j := \langle e_j \mid \psi \rangle$. If $\psi = \sum_j \psi_j e_j$ and $\phi = \sum_j \phi_j e_j$, show that

$$\langle \psi \mid \phi \rangle = \sum_{j\in J} \overline{\psi_j}\, \phi_j.$$

2.2 Dual spaces and Dirac bra-ket notation

Definition 2.6. Let $\mathcal{H}$ be a complex Hilbert space. The dual space

$$\mathcal{H}^* = \big\{ \varphi \mid \varphi : \mathcal{H} \to \mathbb{C} \;\text{ is a continuous and linear map} \big\}.$$

If $\mathcal{H}$ is a finite-dimensional space, then any linear map $\varphi : \mathcal{H} \to \mathbb{C}$ is continuous (exercise?).

For any element $\psi \in \mathcal{H}$, we can define an element $\langle\psi| \in \mathcal{H}^*$, where

$$\langle\psi| : \mathcal{H} \to \mathbb{C}, \qquad \langle\psi|\big(\phi\big) = \langle\psi \mid \phi\rangle \in \mathbb{C}.$$

It follows from properties of the inner-product that for any $a, b \in \mathbb{C}$ $\langle\psi|\big(a\phi_1 + b\phi_2\big) = a\langle\psi|\big(\phi_1\big) + b\langle\psi|\big(\phi_2\big)$, so $\langle\psi|$ is linear. Similarly, by a previous exercise, if $\phi_n \to \phi$, then

$$\langle\psi|(\phi_n) = \langle\psi \mid \phi_n\rangle \to \langle\psi \mid \phi\rangle = \langle\psi|(\phi).$$

Therefore $\langle\psi|$ is continuous and, hence, an element in $\mathcal{H}^*$.

Theorem 2.7 (Riesz Representation Theorem)**.** *Let $\mathcal{H}$ be a complex Hilbert space. There is a bijective correspondence $\mathfrak{R} : \mathcal{H} \to \mathcal{H}^*$ such that $\mathfrak{R}(\psi) = \langle\psi|$ for all $\psi \in \mathcal{H}$.*

Using the Riesz Representation Theorem as a guide, we use a more symmetric way to denote elements in $\mathcal{H}$ and the dual space $\mathcal{H}^*$.

- A vector in the Hilbert space $\mathcal{H}$ is denoted as $|\psi\rangle$, the **ket**.
- A dual vector in $\mathcal{H}^*$ is denoted as $\langle\psi|$, the **bra**.

We can pair a bra $\langle\psi|$ with a ket $|\phi\rangle$ to obtain a bra-ket $\langle\psi \mid \phi\rangle \in \mathbb{C}$. The Riesz map $\mathfrak{R}(|\psi\rangle) = \langle\psi|$. Note that $\mathfrak{R}$ is anti-linear,

$$\mathfrak{R}(a|\psi\rangle + b|\phi\rangle) = \mathfrak{R}(|a\psi + b\phi\rangle) = \langle a\psi + b\phi| = \overline{a}\langle\psi| + \overline{b}\langle\phi| = \overline{a}\,\mathfrak{R}(|\psi\rangle) + \overline{b}\,\mathfrak{R}(|\phi\rangle)$$

for any $a, b \in \mathbb{C}$.

We will generally use the ket notation for elements $|\psi\rangle \in \mathcal{H}$, though will sometimes pass back and forth $\psi \sim |\psi\rangle$. For example, we will write $\|\psi\|$ rather than $\||\psi\rangle\|$.

Given an orthonormal basis $\{e_j\}_{j\in J}$, we can therefore write

$$\psi = \sum_{j\in J}\langle e_j \mid \psi\rangle\,|e_j\rangle = \sum_{j\in J}|e_j\rangle\langle e_j \mid \psi\rangle.$$

Example 2.8. For the case $\mathcal{H} = \mathbb{C}^n$, we can naturally consider $\mathcal{H}^* \cong \mathbb{C}^n$, where the Riesz map

$$\mathfrak{R}\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} \overline{z_1} & \overline{z_2} & \cdots & \overline{z_n} \end{pmatrix}.$$

2.3 Linear operators

At this point, we now assume our Hilbert space $\mathcal{H}$ to be finite-dimensional and fix an orthonormal basis $\{|e_1\rangle, \ldots, |e_n\rangle\}$. Because $|\psi\rangle = \sum_j |e_j\rangle\langle e_j \mid \psi\rangle$, we have that $\sum_j |e_j\rangle\langle e_j| = \mathbf{1}_{\mathcal{H}}$ the identity operator.

Suppose that $A : \mathcal{H} \to \mathcal{H}$ is linear map. We write $A|\psi\rangle$ to denote the vector $|A\psi\rangle$. Using the linearity of A and the inner-product, we can decompose

$$\begin{aligned} A|\psi\rangle &= \sum_j |e_j\rangle\langle e_j \mid A\psi\rangle = \sum_j |e_j\rangle\langle e_j \mid A\Big(\sum_k |e_k\rangle\langle e_k \mid \psi\rangle\Big)\rangle \\ &= \sum_{j,k} |e_j\rangle\langle e_j \mid Ae_k\rangle\langle e_k \mid \psi\rangle. \end{aligned}$$

That is,

$$A = \sum_{j,k} |e_j\rangle\langle e_j \mid Ae_k\rangle\langle e_k| = \sum_{j,k} |e_j\rangle A_{jk}\langle e_k|,$$

where $\{A_{jk}\}_{j,k=1}^n = \{\langle e_j \mid Ae_k\rangle\}_{j,k=1}^n$ are complex-valued matrix coefficients of A *with respect to the orthonormal basis* $\{e_j\}_{j=1}^n$.

Definition 2.9. For a finite-dimensional space $\mathcal{H}$, we denote by $\mathcal{L}(\mathcal{H})$ the vector space of all linear operators $A : \mathcal{H} \to \mathcal{H}$.

Example 2.10. For any $|\psi\rangle$ and $|\phi\rangle \in \mathcal{H}$, we can define the linear operator

$$|\psi\rangle\langle\phi| : \mathcal{H} \to \mathcal{H}, \qquad |\psi\rangle\langle\phi|(|\eta\rangle) = |\psi\rangle\langle\phi \mid \eta\rangle.$$

Let's consider the matrix representation of this operator in the case $\mathcal{H} = \mathbb{C}^n$ with canonical orthonormal basis. Recalling Example 2.8, for $|\psi\rangle = (z_1, \ldots, z_n)$ and $|\phi\rangle = (w_1, \ldots, w_n)$, then

$$|\psi\rangle\langle\phi| = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \begin{pmatrix} \overline{w_1} & \cdots & \overline{w_n} \end{pmatrix} = \begin{pmatrix} z_1\overline{w_1} & \cdots & z_1\overline{w_n} \\ \vdots & \ddots & \vdots \\ z_n\overline{w_1} & \cdots & z_n\overline{w_n} \end{pmatrix}.$$

Definition 2.11.

1. Given a linear operator $A \in \mathcal{L}(\mathcal{H})$, the adjoint $A^* : \mathcal{H} \to \mathcal{H}$ is the linear operator such that
$$\langle \psi \mid A\phi\rangle = \langle A^*\psi \mid \phi\rangle \quad \text{for all } \phi, \psi \in \mathcal{H}.$$
If $A = A^*$, then A is self-adjoint (Hermitian).
2. A linear operator $U : \mathcal{H} \to \mathcal{H}$ is unitary if $U^*U = UU^* = \mathbf{1}_{\mathcal{H}}$. That is, U is invertible with its inverse given by the adjoint.
3. A linear operator $P : \mathcal{H} \to \mathcal{H}$ is an orthogonal projection if $P = P^* = P^2$.

Because we always work on Hilbert spaces, a projection P will always mean an *orthogonal* projection. Given a unit vector $|\psi\rangle$, the linear operator $P_\psi = |\psi\rangle\langle\psi|$ gives the projection on the subspace of $\mathcal{H}$ spanned by $|\psi\rangle$.

We list a few properties of the adjoint. The proof is an exercise.

Lemma 2.12.

1. *Let A and B be linear operators on $\mathcal{H}$. Then*
$$(A^*)_{jk} = \overline{A_{kj}}, \qquad (cA)^* = \bar{c}A^*, \qquad (AB)^* = B^*A^*,$$
where $\{A_{jk}\}$ denotes the matrix coefficients and $c \in \mathbb{C}$.
2. *If $\mathcal{H}$ is finite-dimensional, a linear operator $U : \mathcal{H} \to \mathcal{H}$ is unitary*
$$\langle U\psi \mid U\phi\rangle = \langle \psi \mid \phi\rangle \quad \textit{for all } \phi, \psi \in \mathcal{H}.$$

Remark 2.13. Part (2) of Lemma 2.12 fails when $\mathcal{H}$ is infinite-dimensional. Namely, there are operators $V : \mathcal{H} \to \mathcal{H}$ such that $\langle V\psi \mid V\phi\rangle = \langle \psi \mid \phi\rangle$ but where $VV^* \neq \mathbf{1}_{\mathcal{H}}$.

A concept that will play an essential role in our understanding of quantum mechanics is the eigenvalues (and eigenvectors) of a linear operator.

Definition 2.14. A number $\lambda \in \mathbb{C}$ is an eigenvalue of a linear operator $A : \mathcal{H} \to \mathcal{H}$ with eigenvector $|\psi\rangle \in \mathcal{H}$ if $|\psi\rangle \neq 0$ and $A|\psi\rangle = \lambda|\psi\rangle$. The set of eigenvalues

$$\sigma(A) = \big\{\lambda \in \mathbb{C} \,\big|\, (\lambda \mathbf{1}_{\mathcal{H}} - A) \text{ is not invertible}\big\}$$

is called the spectrum of the linear operator A.

One finds the eigenvalues of an operator by solving the characteristic polynomial equation,

$$(\lambda \mathbf{1}_{\mathcal{H}} - A) \text{ is not invertible} \quad \Longleftrightarrow \quad \det\big(\lambda \mathbf{1}_{\mathcal{H}} - A\big) = 0,$$

where $\det : M_n(\mathbb{C}) \to \mathbb{C}$ is the determinant of a (square) matrix. The determinant of a matrix $A \in M_n(\mathbb{C})$ is independent of the choice of orthonormal basis used to compute the coefficients $A_{jk} = \langle e_j \mid Ae_k\rangle$. Hence the solutions to the polynomial equation $\det\big(\lambda \mathbf{1}_{\mathcal{H}} - A\big) = 0$ is independent of the choice of orthonormal basis. Furthermore, if A is of size $n \times n$, then $p_n(\lambda) = \det\big(\lambda \mathbf{1}_{\mathcal{H}} - A\big)$ can be considered as a complex polynomial of order n in the variable λ. By the Fundamental Theorem of Algebra, the equation

$$p_n(\lambda) = \det\big(\lambda \mathbf{1}_{\mathcal{H}} - A\big) = 0 \quad \Longrightarrow \quad n \text{ solutions, } \lambda_1, \ldots, \lambda_n \in \mathbb{C},$$

where it may occur that $\lambda_j = \lambda_k$ for some $j \neq k$ (multiplicity).

A fixed eigenvalue $\lambda \in \mathbb{C}$ of A might have linearly independent eigenvalues. The span of eigenvectors of λ is called the eigenspace of A for the value λ. We say that λ is a non-degenerate eigenvalue if its eigenspace is one-dimensional. Otherwise, we say that λ is degenerate.

Exercise 2.4.

1. If A is self-adjoint, then $\sigma(A) \subset \mathbb{R}$ (eigenvalues are real).
2. If U is a unitary operator with eigenvalue λ, then $|\lambda| = 1$ (eigenvalues of unitary operators lie on the complex unit circle).
3. If $P = P^* = P^2$, then $\sigma(P) = \{0, 1\}$.

We say that a linear operator A is diagonalisable if there exists an orthonormal basis of eigenvalues of A. That is, there is an othorthonormal basis $\{|e_{j,\alpha}\rangle\}_{j,\alpha}$ such that

$$A|e_{j,\alpha}\rangle = \lambda_j|e_{j,\alpha}\rangle, \qquad \sigma(A) = \{\lambda_1, \ldots, \lambda_m\}.$$

In the above, the index α is used to denote the possible degeneracy of an eigenvalue.

Theorem 2.15 (Spectral Theorem)**.** *A linear operator $A : \mathcal{H} \to \mathcal{H}$ is diagonalisable if and only if it is normal, $A^*A = AA^*$.*

The above theorem shows that self-adjoint and unitary operators are diagonalisable.

Given a diagonalisable operator A, the operator

$$P_{\lambda_j} := \sum_{\alpha} |e_{j,\alpha}\rangle\langle e_{j,\alpha}|$$

gives the (orthogonal) projection onto the λ_j-eigenspace of A. We also note the important property that

$$P_{\lambda_j} P_{\lambda_k} = \delta_{j,k} P_{\lambda_j}.$$

We can also consider the matrix decomposition of A, where in the eigenvector orthonormal basis,

$$A = \sum_{j,\alpha} \sum_{k,\beta} |e_{j,\alpha}\rangle A_{j,\alpha,k,\beta} \langle e_{k,\beta}|.$$

Because $A|e_{j,\alpha}\rangle = \lambda_j|e_{j,\alpha}\rangle$, we have that $A_{j,\alpha,k,\beta} = \lambda_j\delta_{j,k}\delta_{\alpha,\beta}$. So we can simplify

$$\begin{aligned}A &= \sum_{j,\alpha}\sum_{k,\beta}|e_{j,\alpha}\rangle A_{j,\alpha,k,\beta}\langle e_{k,\beta}| = \sum_{j,\alpha}\sum_{k,\beta}|e_{j,\alpha}\rangle\lambda_j\delta_{j,k}\delta_{\alpha,\beta}\langle e_{k,\beta}| \\ &= \sum_{j,\alpha}\lambda_j|e_{j,\alpha}\rangle\langle e_{j,\alpha}| = \sum_j \lambda_j\Big(\sum_\alpha |e_{j,\alpha}\rangle\langle e_{j,\alpha}|\Big) \\ &= \sum_j \lambda_j P_{\lambda_j}.\end{aligned}$$

We call the formula $A = \sum_j \lambda_j P_{\lambda_j}$ the spectral decomposition of A into diagonal form. In particular,

$$A = \begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_m\end{pmatrix} \quad \text{in the eigenvector orthonormal basis,}$$

and where degenerate eigenvalues appear multiple times on the diagonal.

Examples 2.16. 1. When $A = A^*$, $A = \sum_j \lambda_j P_{\lambda_j}$ with $\lambda_j \in \mathbb{R}$ for all j.

2. When U is unitary, eigenvalues are complex with norm 1, so we write

$$U = \sum_j e^{i\theta_j}|e_{j,\alpha}\rangle\langle e_{j,\alpha}| = \sum_j e^{i\theta_j} P_{e^{i\theta_j}}$$

with $\theta_j \in \mathbb{R}$ for all j.

3. For a general normal operator T, $T^*T = TT^*$, we can write each non-zero eigenvalue in polar form, $\lambda_j = r_j e^{i\theta_j}$ with $r_j = |\lambda_j| > 0$ and $\theta_j \in \mathbb{R}$. Hence we can decompose

$$T = \sum_j r_j e^{i\theta_j}|e_{j,\alpha}\rangle\langle e_{j,\alpha}| = \sum_j r_j e^{i\theta_j} P_{\lambda_j}.$$

The determinant gives a map from linear operators on $\mathcal{H}$ to $\mathbb{C}$. Another complex-valued map on linear operators that we will use is the trace.

Definition 2.17. The trace of a linear operator A on $\mathcal{H}$ with orthonormal basis $\{e_j\}_{j=1}^n$ is given by

$$\mathrm{Tr}(A) = \sum_{j=1}^n \langle e_j \mid Ae_j\rangle \in \mathbb{C}.$$

We leave it as an exercise to show that Tr is independent of the choice of orthonormal basis and is such that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ for any linear operators A and B on $\mathcal{H}$.

3 Postulates of quantum mechanics

We will now apply linear algebra to assemble the basics of quantum theory. We will do this in a rather mathematical way, where we consider 'Postulates'/axioms, which we take to be given and consider the logical and physical implications from this.

For the reader who has studied quantum mechanics, throughout this document $\hbar = 1$.

3.1 Pure states and observables

As previously mentioned, quantum mechanics takes place on a complex Hilbert space $\mathcal{H}$. In what follows, we will only consider finite-dimensional spaces. This restriction can be lifted, but it is substantially more technically demanding.

Postulate 1 (Observables and pure states)**.** The elements $|\psi\rangle \in \mathcal{H}$ with $\|\psi\| = 1$ are the pure states of a quantum mechanical system. An observable is a physically measurable quantity which is represented by a self-adjoint (Hermitian) operator $A = A^*$ on $\mathcal{H}$.
The mean-value/expectation of the observable A in the pure state $|\psi\rangle$ is the quantity

$$\langle A \rangle_\psi = \langle \psi \mid A\psi \rangle \in \mathbb{R}.$$

Put another way, an isolated quantum system is described by a fixed Hilbert space $\mathcal{H}$. The possible states that the system can reside in are given by unit vectors $|\psi\rangle \in \mathcal{H}$ and the observables one can measure are the self-adjoint operators. The condition $\|\psi\| = 1$ is a normalisation so that, for example, $\langle \mathbf{1} \rangle_\psi = 1$ for all states $|\psi\rangle$.

Remark 3.1. A crucial element of quantum theory is that measurement is always done with respect to a state $|\psi\rangle$. Given $A = A^*$, different states $|\psi\rangle$ and $|\phi\rangle$ may give different expectation values $\langle A \rangle_\psi \neq \langle A \rangle_\phi$.

If A is an observable, then by the spectral decomposition, $A = \sum_j \lambda_j P_{\lambda_j}$ and

$$\begin{aligned}\langle A \rangle_\psi &= \Big\langle \psi \Big| \Big(\sum_j \lambda_j P_{\lambda_j} \Big) \psi \Big\rangle = \sum_j \lambda_j \langle \psi \mid P_{\lambda_j} \psi \rangle \\ &= \sum_j \lambda_j \langle \psi \mid P_{\lambda_j}^* P_{\lambda_j} \psi \rangle = \sum_j \lambda_j \langle P_{\lambda_j} \psi \mid P_{\lambda_j} \psi \rangle \\ &= \sum_j \lambda_j \big\| P_{\lambda_j} \psi \big\|^2.\end{aligned}$$

Because $\|\psi\| = 1$ and P_{λ_j} is the projection onto a subspace, $\|P_{\lambda_j}\psi\|^2 \in [0,1]$, which we think of as the probability that the measurement of A with respect to $|\psi\rangle$ returns the eigenvalue λ_j.

Postulate 2 (Measurement probability)**.** The possible measurement values of an observable A are given by the spectrum $\sigma(A) \subset \mathbb{R}$.
For a pure state $|\psi\rangle$, the probability that a measurement of A with respect to $|\psi\rangle$ returns the value $\lambda \in \sigma(A)$ is given by

$$\mathbf{P}_\psi(\lambda) = \big\| P_\lambda \psi \big\|^2 = \langle P_\lambda \psi \mid P_\lambda \psi \rangle = \langle P_\lambda \rangle_\psi.$$

For a given observable A and pure state $|\psi\rangle$, the map $\mathbf{P}_\psi : \sigma(A) \to [0,1]$, $\mathbf{P}_\psi(\lambda) = \|P_\lambda \psi\|^2$ indeed defines a probability measure,

$$\sum_{\lambda \in \mathbb{C}} \mathbf{P}_\psi(\lambda) = \sum_\lambda \sum_\alpha \langle \psi \mid e_{\lambda,\alpha} \rangle \langle e_{\lambda,\alpha} \mid \psi \rangle = \sum_{\lambda,\alpha} \big| \psi_{\lambda,\alpha} \big\|^2 = \|\psi\|^2 = 1,$$

where $\{e_{\lambda,\alpha}\}$ is the orthonormal basis of eigenvectors of A and $\psi_{\lambda,\alpha} = \langle e_{\lambda,\alpha} \mid \psi\rangle$ the complex coefficients of ψ in this basis expansion.

Remarks 3.2. 1. If $|\psi\rangle$ and $|\phi\rangle$ are pure states, then so is $a|\psi\rangle + b|\phi\rangle$ for any $a, b \in \mathbb{C}$ such that $\|a\psi + b\phi\| = 1$. This the *superposition* of quantum states.

2. If $|\psi\rangle$ is a pure state, so is $e^{i\theta}|\psi\rangle$ and $\langle A\rangle_\psi = \langle A\rangle_{e^{i\theta}\psi}$. So quantum states are invariant under a global phase. On the other hand, the states $a|\psi\rangle + b|\phi\rangle$ and $a|\psi\rangle + e^{i\theta}b|\phi\rangle$ may give different measurement outcomes. So the *relative* phase of quantum states does matter.

Proposition 3.3. *If a quantum system is prepared in the state $|\psi\rangle$, then the probability to observe it with respect to the state $|\phi\rangle$ is given by $|\langle\psi \mid \phi\rangle|^2$.*

Proof. The projection $P_\phi = |\phi\rangle\langle\phi|$ is the projection onto the span of $|\phi\rangle$. Therefore, to measure $|\phi\rangle$ in the state $|\psi\rangle$ we are precisely measuring $\mathbf{P}_\psi(1)$ of the observable P_ϕ. Because $|\phi\rangle$ is the 1-eigenvector of P_ϕ, we are therefore measuring

$$\mathbf{P}_\psi(1) = \langle\psi \mid \big(|\phi\rangle\langle\phi|\big)\psi\rangle = \big|\langle\psi \mid \phi\rangle\big|^2. \qquad \square$$

One can describe measurement in quantum mechanics is several ways that, when combined with the other postulates, are equivalent to each other. For completeness, we also state measurement with respect to an orthonormal basis.

Postulate (Measurement – orthonormal basis version)**.** Each orthonormal basis $\{|e_j\rangle\}_{j=1}^n \subset \mathcal{H}$ describes a measurement as follows. Given any state

$$|\psi\rangle = \sum_{j=1}^n |e_j\rangle\langle e_j \mid \psi\rangle, \qquad \sum_{j=1}^n \big|\langle e_j \mid \psi\rangle\big|^2 = 1,$$

then the state $|\psi\rangle$ after measurement is $|e_j\rangle$ with probability $|\langle e_j \mid \psi\rangle|^2$.

Remark 3.4 (Observable vs orthonormal basis description of measurementements). Because any self-adjoint operator $A = A^*$ is diagonalisable, there is a corresponding orthonormal basis $\{|e_{j,\alpha}\rangle\}$ such that $A = \sum_{j,\alpha} = \lambda_j |e_{j,\alpha}\rangle\langle e_{j,\alpha}|$. We can therefore decompose any pure state

$$\psi\rangle = \sum_{j,\alpha} |e_{j,\alpha}\rangle\langle e_{j,\alpha} \mid \psi\rangle$$

and measurement of $|\psi\rangle$ in this orthonormal basis will return $|e_{j,\alpha}\rangle$ with probability $\big|\langle e_{j,\alpha} \mid \psi\rangle\big|^2$. The knowledge that $\{|e_{j,\alpha}\rangle\}$ comes from a spectral decomposition gives us extra information. Namely, we have eigenvalues $\{\lambda_j\}_{j=1}^m$ that label every basis element. If λ_j has linearly independent eigenvalues, then $\dim(P_{\lambda_j}\mathcal{H}) = l > 1$. In this case $|e_{j,1}\rangle, \dots |e_{j,l}\rangle$ are all assigned the same value λ_j. So the probability that a measurement of $|\psi\rangle$ in the orthonormal basis $\{|e_{j,\alpha}\rangle\}$ will return any of these vectors is given by

$$\begin{aligned}\sum_{\alpha=1}^l \big|\langle e_{j,\alpha} \mid \psi\rangle\big|^2 &= \sum_{\alpha=1}^l \langle\psi \mid e_{j,\alpha}\rangle\langle e_{j,\alpha} \mid \psi\rangle = \Big\langle \psi \,\Big|\, \sum_{\alpha=1}^l \big(|e_{j,\alpha}\rangle\langle e_{j,\alpha}|\big)\psi \Big\rangle \\ &= \langle\psi \mid P_{\lambda_j}\psi\rangle = \langle P_{\lambda_j}\rangle_\psi.\end{aligned}$$

Hence, the probability for the measurement of $|\psi\rangle$ to return any of the vectors $|e_{j,\alpha}\rangle$ in a fixed eigenspace $P_{\lambda_j}\mathcal{H}$ (using the orthonormal basis approach to measurement) is the same as the probability to measure the eigenvalue λ_j of A with respect to the state $|\psi\rangle$ (observable approach to measurement).

Conversely, given any orthonormal basis $\{|e_j\rangle\}_{j=1}^n$, we can define a self-adjoint observable $A = \sum_j j|e_j\rangle\langle e_j|$. Then the probability to measure the value $j \in \sigma(A)$ in the state $|\psi\rangle$ is given by

$$\mathbf{P}_\psi(j) = \langle P_j\rangle_\psi = \langle \psi \mid (|e_j\rangle\langle e_j|)\psi\rangle = \left|\langle e_j \mid \psi\rangle\right|^2,$$

which is the same as the probability of returning $|e_j\rangle$ if we were to measure $|\psi\rangle$ in the orthonormal basis $\{|e_j\rangle\}$. To summarise, the two approaches to measurement represent the same probabilistic data.

There is a more general approach to measurement that is with respect to a set of measurement operators $\{M_k\}_{k=1}^m \subset \mathcal{L}(\mathcal{H})$, where $\sum_k M_k^* M_k = \mathbf{1}_\mathcal{H}$. In this picture, given the pure state $|\psi\rangle$, the probability that a result $k = 1, \ldots, m$ occurs is given by $\|M_k\psi\|^2 = \langle M_k\psi \mid M_k\psi\rangle$. We recover the previous picture in the case where $M_k = |e_k\rangle\langle e_k|$ for an orthonormal basis $\{|e_k\rangle\}$ or $M_k = P_{\lambda_k}$, a spectral projection onto a fixed eigenspace of a self-adjoint operator.

The expectation $\langle A\rangle_\psi$ is probabilistic quantity. We can similarly consider the standard deviation/uncertainly of a given observable A with respect to $|\psi\rangle$.

Definition 3.5. The uncertainty/standard deviation of an observable A in the state $|\psi\rangle$ is

$$\Delta_\psi(A) = \sqrt{\left\langle (A - \langle A\rangle_\psi \mathbf{1})^2\right\rangle_\psi} = \sqrt{\langle \psi \,|\, (A - \langle A\rangle_\psi \mathbf{1})^2 \psi\rangle}.$$

Example 3.6. Suppose that $|\psi\rangle$ is an eigenvector of A, $A|\psi\rangle = \lambda|\psi\rangle$. Then $\langle A\rangle_\psi = \lambda$ and

$$\Delta_\psi(A) = \sqrt{\langle \psi \,|\, (A - \langle A\rangle_\psi \mathbf{1})^2 \psi\rangle} = \sqrt{\langle \psi \,|\, (A - \lambda\mathbf{1})^2 \psi\rangle} = 0.$$

Exercise 3.1. Show that

$$\Delta_\psi(A) = 0 \quad \Longleftrightarrow \quad A|\psi\rangle = \langle A\rangle_\psi |\psi\rangle.$$

The uncertainty $\Delta_\psi(A)$ gives an indication of the spread of the possible values that may occur when we measure A in the state $|\psi\rangle$. Indeed, If $\Delta_\psi(A) \neq 0$, then different measurements of A in the state $|\lambda\rangle$ may give different outcomes. The expectation $\langle A\rangle_\psi$ gives us the average/mean of these measurements, but if $\Delta_\psi(A)$ is large, then we may obtain measurements very far from $\langle A\rangle_\psi$ with non-zero probability.

This uncertainty of measurement is a key feature/property of quantum mechanics, as is evidenced by the famous result below.

Proposition 3.7 (The uncertainty relation). *For any pure state and observables A and B,*

$$\Delta_\psi(A)\Delta_\psi(B) \geq \frac{1}{2}\left|\langle [A, B]\rangle_\psi\right|$$

Example 3.8. The most famous example of the uncertainty relation is the *Heisenberg* uncertainty relation, which considers the position X and momentum D observables on the (infinite-dimensional) Hilbert space $L^2(\mathbb{R})$, where

$$(X\psi)(x) = x\psi(x), \qquad (D\psi)(x) = -i\psi'(x), \qquad [X, D] = i$$

and so $\Delta_X(\psi)\Delta_\psi(D) \geq \frac{1}{2}$. This means if we measure the position with high certainty, then we cannot simultaneously measure the momentum with high certainty.

Given an observable $A = \sum_j \lambda_j P_{\lambda_j}$ and state $|\psi\rangle$, then we will measure the value λ_j with probability $\mathbf{P}_\psi(\lambda_j) = \|P_{\lambda_j}\psi\|^2$. If we measure the same state again, we will *always* get the value λ_j. That is, the measurement alters the state $|\psi\rangle$.

Postulate 3 (Projection / collapse of 'wave function')**.** If a measurement of the observable A in the pure state $|\psi\rangle$ yields the eigenvalue λ, then the measurement has caused the transition

$$\underbrace{|\psi\rangle}_{\text{before measurement}} \xrightarrow{\text{measure } A} \underbrace{\frac{P_\lambda|\psi\rangle}{\|P_\lambda\psi\|}}_{\text{after measurement}} .$$

We see that measurement fundamentally changes the quantum state. We can also perform operations on a quantum state that are not as drastic. Indeed, if $|\psi\rangle$ is a state and $U : \mathcal{H} \to \mathcal{H}$ is unitary, then $\|U\psi\| = \|\psi\| = 1$ and so $U|\psi\rangle$ is also a state. We can therefore use unitary operators to consider the evolution of a state as time progresses.

Postulate 4 (Time evolution – Hiesenberg picture)**.** The time-evolved state $|\psi(t)\rangle$, $t \in \mathbb{R}$, from an initial state $|\psi_0\rangle$ is given by $|\psi(t)\rangle = U(t)|\psi_0\rangle$, wherer $U(t)$ is a unitary operator such that

$$i\frac{d}{dt}U(t) = H(t)U(t), \qquad U(0) = \mathbf{1}_{\mathcal{H}}, \tag{3.1}$$

where $H(t)$ is a self-adjoint operator, the Hamiltonian, that describes the observable energy of the quantum system at time t. The operator $U(t)$ is called the time-evolution unitary.

Exercise 3.2. Show that any solution $U(t)$ to the Equation (3.1) is unitary and unique.

The Hamiltonian $H(t)$ both specifies the possible energies of a quantum system as well as how it evolves over time. As such, the Hamiltonian is often the most important observable to consider for a given quantum system.

There is an equivalent formulation of Postulate 4 that works on the evolution of the states themselves.

Postulate (Time evolution – Schrödinger picture)**.** The time-evolved state $|\psi(t)\rangle$, $t \in \mathbb{R}$, from an initial state $|\psi_0\rangle$ is the solution to the Schrödinger equation

$$i\frac{d}{dt}|\psi(t)\rangle = H(t)|\psi(t)\rangle, \qquad |\psi(0)\rangle = |\psi_0\rangle,$$

where $H(t)$ is the Hamiltonian operator at time t.

While the Heisenberg picture of the time-evolution is a little more complicated (involving a differential equation of linear operators rather than vectors), it is the more natural picture from the perspective of quantum information and computing, where we perform unitary operations to transmit information encoded in a quantum state. Because the Heisenberg and Schrödinger picture are equivalent and describe the same process, we also call Equation (3.1) the Schrödinger equation.

Example 3.9. Suppose that we have a time-*independent* Hamiltonian, $H(t) = H = H^*$ for all t. Taking the spectral decomposition $H = \sum_j E_j P_{E_j}$ where $\{E_1, \ldots, E_m\}$ are the energy eigenvalues, we then consider

$$U(t) = \sum_j e^{-itE_j} P_{E_j}.$$

We leave it as an exercise to show that $U(t)$ is unitary. We then directly check the Schrödinger equation,

$$\begin{aligned} i\frac{d}{dt}U(t) &= \sum_j E_j e^{-itE_j} P_j \\ &= \Big(\sum_j E_j P_{E_j}\Big)\Big(\sum_k e^{-itE_k} P_{E_k}\Big) \qquad (\text{as } P_{E_j}P_{E_k} = \delta_{j,k}P_{E_j}) \\ &= HU(t). \end{aligned}$$

Also $U(0) = \sum_j P_{E_j} = \mathbf{1}_{\mathcal{H}}$ and so $U(t)$ is indeed the time evolution. We can also write

$$U(t) = e^{-itH} := \sum_j e^{-itE_j} P_{E_j} = \sum_{n=0}^{\infty} \frac{(-itH)^n}{n!},$$

where the last equation is a computation using the spectral decomposition of H (exercise?).

If $H(t)$ is time-*dependent*, finding the time-evolution $U(t)$ is quite challenging in general. One might guess that $U(t) = \exp\big(-i\int_0^t H(s)\mathrm{d}s\big)$ would solve the equation. But this does not always work as Hamiltonians at different times $H(t)$ and $H(t')$ might not commute, $[H(t), H(t')] \neq 0$. In such a setting, one instead uses the so-called time-ordered exponential, which is beyond the scope of these notes.

Conversely, one may have a family of unitary operators $U(t)$ and may wish to find a Hamiltonian $H(t)$ such that $U(t)$ solves the corresponding Schrödinger equation. When $U(t) = U$ is constant in time, this can be done by an taking an appropriate logarithm. Though we remark that the some care is needed when taking the complex logarithm to avoid discontinuities.

Example 3.10. Suppose that $\mathcal{H} = \mathbb{C}^2$ and $H = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. We see that $H^2 = \mathbf{1}_2$ and so a computation using the Taylor expansion gives that

$$U(t) = \exp(-it\sigma_x) = \cos(t)\mathbf{1}_2 - i\sigma_x \sin(t) = \begin{pmatrix} \cos(t) & -i\sin(t) \\ -i\sin(t) & \cos(t) \end{pmatrix}.$$

We consider the state $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We first compute

$$\langle \sigma_x \rangle_{|0\rangle} = \langle 0 \mid \sigma_x 0\rangle = \Big\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Bigm| \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Big\rangle = 0$$

and so $\sigma_x - \langle \sigma_x \rangle_{|0\rangle}\mathbf{1}_2 = \sigma_x$ and

$$\Delta_{|0\rangle}(\sigma_x) = \big\langle 0 |(\sigma_x - \langle \sigma_x \rangle_{|0\rangle}\mathbf{1}_2)^2| 0\big\rangle = \langle 0 \mid \mathbf{1}_2 0\rangle = 1.$$

Because $\Delta_{|0\rangle}(\sigma_x) \neq 0$, we see that we can't sharply measure the observable σ_x in the the state $|0\rangle$.

3.2 Mixed states

A pure state is determined by a fixed element $|\psi\rangle \in \mathcal{H}$. We now consider an ensemble of states $\{|\psi_j\rangle\}_{j\in J} \subset \mathcal{H}$ with probabilities $\{p_j\}_{j\in J} \subset [0,1]$ such that $p_j \in [0,1]$ denotes the probability that the system is in the (pure) state $|\psi_j\rangle$. We also assume that this collection is 'complete' in the sense that $\sum_j p_j = 1$.

Suppose we are in the setting that the collection $\{|\psi_j\rangle\}_j$ forms an orthonormal basis of $\mathcal{H}$. Then we can define the operator

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|,$$

which will be self-adjoint and positive with eigenvalues $\{p_j\}_j \subset [0,1]$. For such an operator

$$\mathrm{Tr}(\rho) = \sum_k \Big\langle \psi_k \Big| \Big(\sum_j p_j |\psi_j\rangle\langle\psi_j| \Big) \psi_k \Big\rangle = \sum_j p_j = 1.$$

We take the above properties and turn them into an abstract definition.

Postulate 5 (Mixed states)**.** A general quantum mechanical system is described by a linear operator $\rho : \mathcal{H} \to \mathcal{H}$ such that

1. $\rho^* = \rho$ and $\rho \geq 0$, i.e. all eigenvalues are non-negative,
2. $\mathrm{Tr}(\rho) = 1$.

We call ρ the density operator/matrix of the quantum state.

Remark 3.11. Given a Hilbert space $\mathcal{H}$, we denote by

$$\mathrm{Dens}(\mathcal{H}) = \big\{\rho : \mathcal{H} \to \mathcal{H} \mid \rho = \rho^*,\ \rho \geq 0,\ \mathrm{Tr}(\rho) = 1\big\}$$

the set of density operators. A quantum state is therefore determined by an element $\rho \in \mathrm{Dens}(\mathcal{H})$.

Example 3.12. If $|\psi\rangle$ is a pure state, then $\rho_\psi = P_\psi = |\psi\rangle\langle\psi|$ is a density operator.

We note the following useful property. The proof is an exercise.

Lemma 3.13. *If $\rho \in \mathrm{Dens}(\mathcal{H})$ and $U : \mathcal{H} \to \mathcal{H}$ is unitary, then $U\rho U^* \in \mathrm{Dens}(\mathcal{H})$.*

Suppose that $\{|\psi_j\rangle\}_{j\in J} \subset \mathcal{H}$ is an orthonormal basis and $\{p_j\}_{j\in J} \subset [0,1]$ a collection of probabilities the system is in state $|\psi_j\rangle$. One should be careful to not confuse

$$\underbrace{\rho = \sum_{j\in J} p_j |\psi_j\rangle\langle\psi_j| \in \mathrm{Dens}(\mathcal{H}),}_{\text{mixed state}} \qquad \underbrace{|\psi\rangle = \sum_{j\in J} \sqrt{p_j}|\psi_j\rangle \in \mathcal{H}.}_{\text{pure state}}$$

Let us rewrite the previous postulates considered for pure states in the more general language of mixed states.

Postulate (Density operator postulates)**.** Let $\mathcal{H}$ be a Hilbert space describing a quantum mechanical system and $\rho \in \mathrm{Dens}(\mathcal{H})$ a quantum state.

1. The expectation of an observable $A = A^*$ in ρ is given by
$$\langle A \rangle_\rho = \mathrm{Tr}(\rho A) \in \mathbb{R}.$$
2. The probability that a measurement of the observable $A = A^*$ in the state ρ returns the value $\lambda \in \sigma(A)$ is given by
$$\mathbf{P}_\rho(\lambda) = \mathrm{Tr}(\rho P_\lambda) \in [0,1], \qquad P_\lambda = \sum_\alpha |e_{\lambda,\alpha}\rangle\langle e_{\lambda,\alpha}| \text{ eigenspace projection.}$$
3. If we measure the observable A in the state ρ and obtain the value $\lambda \in \sigma(A)$, this measurement has caused the transition
$$\underbrace{\rho}_{\text{before measurement}} \xrightarrow{\text{measure } A} \underbrace{\frac{P_\lambda \rho P_\lambda}{\mathrm{Tr}(\rho P_\lambda)}}_{\text{after measurement}}.$$
4. If $\rho_0 \in \mathrm{Dens}(\mathcal{H})$ is a state at time $t = 0$, then the time-evolved state is given by $\rho(t) = U(t)\rho_0 U(t)^*$, where $U(t)$ is solves the Schrödinger equation,
$$i\frac{d}{dt}U(t) = H(t)U(t), \qquad U(0) = \mathbf{1}_{\mathcal{H}}.$$
5. The uncertainty of an observable $A = A^*$ in the state ρ is the quantity
$$\Delta_\rho(A) = \sqrt{\big\langle (A - \langle A\rangle_\rho \mathbf{1})^2 \big\rangle_\rho}.$$

Exercise 3.3. 1. Check that for the case $\rho = \rho_\psi = |\psi\rangle\langle\psi|$, the postulates for ρ are equivalent to the postulates for the pure state $|\psi\rangle$.

2. Check that the postulates for $\rho \in \mathrm{Dens}(\mathcal{H})$ are well-defined. Namely, show that for an observable $A = A^*$ with eigenvalue $\lambda \in \sigma(A)$,
$$\langle A\rangle_\rho = \mathrm{Tr}(\rho A) \in \mathbb{R}, \qquad \mathbf{P}_\rho(\lambda) = \mathrm{Tr}(\rho P_\lambda) \in [0,1], \qquad \frac{P_\lambda \rho P_\lambda}{\mathrm{Tr}(\rho P_\lambda)} \in \mathrm{Dens}(\mathcal{H}).$$

The Spectral Theorem (Theorem 2.15) gives us a canonical form for density operators.

Proposition 3.14. *Let $\rho \in \mathrm{Dens}(\mathcal{H})$ be a density operator. Then there exist $\{p_j\}_{j\in J} \subset [0,1]$ and and orthonormal basis $\{|\psi_j\rangle\}_{j\in J}$ such that*
$$\rho = \sum_{j\in J} p_j |\psi_j\rangle\langle\psi_j|, \qquad \sum_{j\in J} p_j = 1.$$

Proof. Because ρ is self-adjoint, there is a spectral decomposition $\rho = \sum_{k,\alpha} \lambda_k |e_{k,\alpha}\rangle\langle e_{k,\alpha}|$. Recalling that the trace can be computed by the sum of eigenvalues, if $\rho \geq 0$ and $\mathrm{Tr}(\rho) = 1$, it must follow that $\lambda_k \in [0,1]$ for all k. We then define take the orthonormal basis $\{|\psi_j\rangle\}_j$ to be $\{|e_{k,\alpha}\rangle\}_{k,\alpha}$. That is, if λ_k is non-degenerate, we take each eigenvalue $\{e_{k,1}, \ldots, e_{k,m}\}$ as a separate ψ_j. Similarly, we define $p_j = \lambda_j$, where we repeat eigenvalues when there is degeneracy. This gives us the desired decomposition. □

Hence from our abstract picture of mixed states $\rho \in \mathrm{Dens}(\mathcal{H})$, we can recover the probabilistic picture of a collection of pure states $\{|\psi_j\rangle\}_j$ with $p_j = \lambda_j \in \sigma(\rho)$ the probability the system is in the state $|\psi_j\rangle$.

Proposition 3.14 gives a 'canonical' decomposition of a density matrix into an ensemble of pure states $\{|\psi_j\rangle\} \subset \mathcal{H}$ with probabilities $\{p_j\} \subset [0,1]$. However, this decomposition is *not* unique.

Example 3.15. Suppose that $\{|\psi\rangle, |\phi\rangle\} \subset \mathbb{C}^2$ are an orthonormal basis (e.g. $|\psi\rangle = |0\rangle = \binom{1}{0}$ and $|\phi\rangle = |1\rangle = \binom{0}{1}$) and consider the density matrix

$$\rho = \frac{2}{3}|\psi\rangle\langle\psi| + \frac{1}{3}|\phi\rangle\langle\phi|.$$

We may be tempted to assert that ρ describes a system that is in the state $|\psi\rangle$ with probability $\frac{2}{3}$ and state $|\phi\rangle$ with probability $\frac{1}{3}$. But this might not be the case. Define the states

$$|\Psi_+\rangle = \sqrt{\tfrac{2}{3}}|\psi\rangle + \sqrt{\tfrac{1}{3}}|\phi\rangle, \qquad |\Psi_-\rangle = \sqrt{\tfrac{2}{3}}|\psi\rangle - \sqrt{\tfrac{1}{3}}|\phi\rangle.$$

Then the density operator describing quantum system prepared in the state $|\Psi_+\rangle$ and $|\Psi_-\rangle$ with probability $\frac{1}{2}$ each is given by

$$\begin{aligned}\frac{1}{2}|\Psi_+\rangle\langle\Psi_+| + \frac{1}{2}|\Psi_-\rangle\langle\Psi_-| &= \frac{1}{2}\big(\sqrt{\tfrac{2}{3}}|\psi\rangle + \sqrt{\tfrac{1}{3}}|\phi\rangle\big)\big(\sqrt{\tfrac{2}{3}}\langle\psi| + \sqrt{\tfrac{1}{3}}\langle\phi|\big) \\ &\quad + \frac{1}{2}\big(\sqrt{\tfrac{2}{3}}|\psi\rangle - \sqrt{\tfrac{1}{3}}|\phi\rangle\big)\big(\sqrt{\tfrac{2}{3}}\langle\psi| - \sqrt{\tfrac{1}{3}}\langle\phi|\big) \\ &= \frac{2}{3}|\psi\rangle\langle\psi| + \frac{1}{3}|\phi\rangle\langle\phi| = \rho.\end{aligned}$$

That is, the same density operator ρ can potentially describe many different quantum ensembles.

The following result gives a characterisation of what ensembles describe the same density operator.

Theorem 3.16 ([3, Theorem 2.6]). *The orthonormal bases $\{|\psi_j\rangle\}_{j=1}^n$ and $\{|\phi_k\rangle\}_{k=1}^n$ in $\mathcal{H}$ with probabilities $\{p_j\}_{j=1}^n$ for and $\{q_k\}_{k=1}^n$ generate the same density operator $\rho = \sum_j p_j|\psi_j\rangle\langle\psi_j| = \sum_k q_k|\phi_k\rangle\langle\phi_k|$ if and only if there is a unitary operator $U \in \mathcal{L}(\mathcal{H})$ with matrix coefficients $\{U_{jk}\}_{j,k=1}^n$ such that*

$$\sqrt{p_j}|\psi_j\rangle = \sum_{k=1}^n U_{jk}\sqrt{q_k}|\phi_k\rangle.$$

For one part of the proof, the unitary operator is defined via its matrix coefficients, $U_{jk} = \sqrt{\frac{p_j}{q_k}}\langle\phi_k \mid \psi_j\rangle$. The results can also be extended to the case of sets $\{|\psi_j\rangle\}_{j=1}^m$ and $\{|\phi_k\rangle\}_{k=1}^n$ with $m < n$. In this case, we extend the first set to $\{|\psi_1\rangle, \dots, |\psi_m\rangle, 0, 0, \dots, 0\}$ with $m - n$ zeros added.

Exercise 3.4.

1. If $\rho \in \mathrm{Dens}(\mathcal{H})$ and $|\psi\rangle \in \mathcal{H}$, show that $\langle\psi \mid (\rho - \rho^2)\psi\rangle \geq 0$.
2. Show that $\rho \in \mathrm{Dens}(\mathcal{H})$ is a pure state $\rho = |\psi\rangle\langle\psi|$ if and only if $\rho^2 = \rho$.
3. Show that $\rho \in \mathrm{Dens}(\mathcal{H})$ describes a non-pure state if $\mathrm{Tr}(\rho^2) < 1$.

Example 3.17. Suppose that $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j| \in \mathrm{Dens}(\mathcal{H})$ and $A = A^*$ is an observable. We assume that A has no degenerate eigenvalues and consider the spectral decomposition of A, $A = \sum_k \lambda_k |e_k\rangle\langle e_k|$. We compute

$$\begin{aligned}\mathbf{P}_\rho(\lambda_k) &= \big\langle |e_k\rangle\langle e_k| \big\rangle_\rho = \mathrm{Tr}\left(\rho|e_k\rangle\langle e_k|\right)\\ &= \sum_i \langle e_i \mid (\rho|e_k\rangle\langle e_k) e_i\rangle = \langle e_k \mid \rho e_k\rangle\\ &= \sum_j \langle e_k \mid p_j\psi_j\rangle\langle\psi_j \mid e_k\rangle = \sum_j \big|\langle\psi_j \mid e_k\rangle\big|^2.\end{aligned}$$

By linearity, we also have shown that

$$\langle A\rangle_\rho = \sum_k \lambda_k \big\langle |e_k\rangle\langle e_k|\big\rangle_\rho = \sum_{k,j} \lambda_k p_j \big|\langle\psi_j \mid e_k\rangle\big|^2.$$

We leave the case that A has degenerate eigenvalues as an exercise.

Remark 3.18. Once again, the quantum state $\rho \in \mathrm{Dens}(\mathcal{H})$ is not affected by a global phase. If $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ and we consider $|\psi_j\rangle \mapsto |e^{i\theta_j}\psi_j\rangle$, then

$$\rho \mapsto \sum_j p_j |e^{i\theta_j}\psi_j\rangle\langle e^{i\theta_j}\psi_j| = \sum_j p_j e^{i\theta_j} e^{-i\theta_j} |\psi_j\rangle\langle\psi_j| = \sum_j p_j |\psi_j\rangle\langle\psi_j| = \rho.$$

3.3 The Qubit space $\mathbb{C}^2$

Let us now apply some of our acquired knowledge of the basics of quantum mechanics to the very simple but import space of single Qubits, $\mathbb{C}^2$. Recall that a classical bit is an element $x \in \{0, 1\}$. The element x can be considered as a logical check, where 0 represents yes/true and 1 represents no/false.

Definition 3.19. A Qubit is a pure state $|\psi\rangle \in \mathbb{C}^2$. That is, $|\psi\rangle$ is a unit vector, $\|\psi\| = 1$.

Example 3.20 (Useful orthonormal bases of $\mathbb{C}^2$)**.** We highlight a few orthonormal bases of $\mathbb{C}^2$ and set some notation.

1. The most canonical orthonormal basis, sometimes the canonical basis, can be described via the ±1 eigenvectors of $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$,

$$|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \sigma_z|y\rangle = (-1)^y|y\rangle \;\text{ for } y \in \{0, 1\}.$$

   We can therefore write
$$\sigma_z = \sum_j \lambda_j P_{\lambda_j} = |0\rangle\langle 0| - |1\rangle\langle 1|.$$

2. We will also make use of the orthonormal basis that comes from the ±1 eigenvectors of $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, where we use the special notation,

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) - \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \sigma_x|\pm\rangle = (\pm 1)|\pm\rangle.$$

The vectors $|0\rangle$ and $|1\rangle$ are examples of Qubits, but recalling the principle of superposition, linear combinations of pure states also give pure state. Therefore the vector

$$|x\rangle = a|0\rangle + b|1\rangle, \quad a, b \in \mathbb{C},\ |a|^2 + |b|^2 = 1$$

is also a *single* Qubit. It is *not* a sum of two Qubits. Therefore, unlike the case of a classical bit $x \in \{0, 1\}$, where there are two options, there are (uncountably infinitely) many possibilities for a single Qubit $|\psi\rangle \in \mathbb{C}^2$.

Example 3.21. Let's consider σ_z as an observable (it is self-adjoint). Any observation of σ_z with respect to a pure state/Qubit $|\psi\rangle$ will yield the value $+1$ or -1. For $|\psi\rangle = a|0\rangle + b|1\rangle$ with $|a|^2 + |b|^2 = 1$, we see that

$$\mathbf{P}_{|\psi\rangle}(+1) = \langle |0\rangle\langle 0| \rangle_{|\psi\rangle} = \big|\langle 0 \mid \psi\rangle\big|^2 = \big|a\langle 0 \mid 0\rangle + b\langle 0 \mid 1\rangle\big|^2 = |a|^2.$$

Similar computations give that $\mathbf{P}_{|\psi\rangle}(-1) = |b|^2$ and $\langle \sigma_z \rangle_{|\psi\rangle} = |a|^2 - |b|^2$.
If we take a measurement of σ_z in the state $|\psi\rangle$ and return the value $+1$, then by the Projection Postulate (Postulate 3),

$$|\psi\rangle \xrightarrow{\text{measure } +1} \frac{P_{+1}|\psi\rangle}{\|P_{+1}|\psi\rangle\|},$$

where

$$P_{+1}|\psi\rangle = |0\rangle\langle 0|(a|0\rangle + b|1\rangle) = a|0\rangle, \qquad \|P_{+1}|\psi\rangle\| = |a|.$$

Therefore the measurement of $+1$ collapses $|\psi\rangle$ to the state

$$\frac{a}{|a|}|0\rangle = \operatorname{sgn}(a)|0\rangle = e^{ik\pi}|0\rangle \sim |0\rangle,$$

where $k \in \mathbb{Z}$ and we have used that $|\psi\rangle \sim e^{i\alpha}|\psi\rangle$ by the global phase invariance of states. Similarly, if we measure σ_z in with respect to $|\psi\rangle$ and yield the value -1, then $|\psi\rangle$ collapses to the state $|1\rangle$.

We also have a geometric interpretation of generic Qubits $|\psi\rangle = a|0\rangle + b|1\rangle$ by considering a parametrisation of the coefficients $a, b \in \mathbb{C}$ with $|a|^2 + |b|^2 = 1$.

Exercise 3.5. Show that up to global phase equivalence $|\psi\rangle \sim e^{i\alpha}|\psi\rangle$ any Qubit can be written as the pure state

$$|\psi(\theta, \phi)\rangle = e^{-i\frac{\phi}{2}} \cos\left(\tfrac{\theta}{2}\right)|0\rangle + e^{i\frac{\phi}{2}} \sin\left(\tfrac{\theta}{2}\right)|1\rangle, \quad \theta, \phi \in \mathbb{R}.$$

Suppose that $\mathbf{a} = (a_1, a_2, a_3) \in \mathbb{R}^3$. We define the matrix

$$\mathbf{a} \cdot \sigma = a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 = \begin{pmatrix} a_3 & a_1 - ia_2 \\ a_1 + ia_2 & -a_3 \end{pmatrix},$$

which is a self-adjoint. We also recall the spherical coordinates in $\mathbb{R}^3$, where any non-zero vector $\mathbf{x} \in \mathbb{R}^3$ with $\|\mathbf{x}\| = r > 0$ can be written

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\sin(\theta)\cos(\phi) \\ r\sin(\theta)\sin(\phi) \\ r\cos(\theta) \end{pmatrix}, \quad \theta \in [0, \pi],\ \phi \in [0, 2\pi).$$

We are interested in the unit sphere, where $r = 1$, where we define

$$\hat{n}(\theta, \phi) = \begin{pmatrix} \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{pmatrix}, \qquad \hat{n}(\theta, \phi) \cdot \sigma = \begin{pmatrix} \cos(\theta) & \sin(\theta)e^{-i\phi} \\ \sin(\theta)e^{-i\phi} & -\cos(\theta) \end{pmatrix}.$$

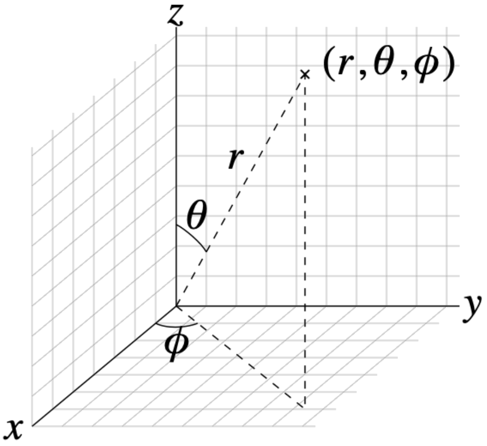


Figure 1: Spherical coordinates, from Wikipedia.

Note that $\hat{n}(\theta,\phi)\cdot\sigma$ is self-adjoint and unitary, $(\hat{n}(\theta,\phi)\cdot\sigma)^2=\mathbf{1}_2$. We can explicitly write down the ±1 eigenvectors of $\hat{n}(\theta,\phi)\cdot\sigma$, where

$$\begin{aligned} |\uparrow_{\hat{n}}\rangle &= \begin{pmatrix} e^{\frac{-i\phi}{2}}\cos(\frac{\theta}{2}) \\ e^{\frac{i\phi}{2}}\sin(\frac{\theta}{2}) \end{pmatrix} = e^{-i\frac{\phi}{2}}\cos\left(\tfrac{\theta}{2}\right)|0\rangle + e^{i\frac{\phi}{2}}\sin\left(\tfrac{\theta}{2}\right)|1\rangle, & |\downarrow_{\hat{n}}\rangle &= \begin{pmatrix} -e^{\frac{-i\phi}{2}}\sin(\frac{\theta}{2}) \\ e^{\frac{i\phi}{2}}\cos(\frac{\theta}{2}) \end{pmatrix}, \\ \hat{n}(\theta,\phi)\cdot\sigma|\uparrow_{\hat{n}}\rangle &= (+1)|\uparrow_{\hat{n}}\rangle, & \hat{n}(\theta,\phi)\cdot\sigma|\downarrow_{\hat{n}}\rangle &= (-1)|\downarrow_{\hat{n}}\rangle. \end{aligned}$$

The vector $\hat{n}(\theta,\phi)$ represents an element in $S_{\mathbb{R}^3}$, the unit sphere in $\mathbb{R}^3$. The corresponding vector $|\uparrow_{\hat{n}}\rangle$ is the 'spin-up' state for the spin in direction $\hat{n}(\theta,\phi)$.

> **Proposition 3.22.** *For any pure state $|\psi\rangle\in\mathbb{C}^2$, there is a point $\hat{n}(\theta,\phi)\in S_{\mathbb{R}^3}$ such that $|\psi\rangle=|\uparrow_{\hat{n}}\rangle$. Hence there an equivalence between Qubits $|\psi\rangle$ and points on the* Bloch sphere *$S_{\mathbb{R}^3}$.*

We therefore have a geometric interpretation of Bloch states. For example, one finds that

$$\hat{n}(0,0)\cdot\sigma=\sigma_z \quad\Longrightarrow\quad |\uparrow_{\hat{n}(0,0)}\rangle=|0\rangle,\quad |\downarrow_{\hat{n}(0,0)}\rangle=|1\rangle,$$

so the the Qubits $|0\rangle$ and $|1\rangle$ represent the north and south poles of the Bloch sphere. Similalry,

$$\hat{n}(\tfrac{\pi}{2},0)\cdot\sigma=\sigma_x \quad\Longrightarrow\quad |\uparrow_{\hat{n}(\frac{\pi}{2},0)}\rangle=|+\rangle,\quad |\downarrow_{\hat{n}(\frac{\pi}{2},0)}\rangle=|-\rangle,$$

so $|\pm\rangle$ are located at the intersection of the Bloch sphere with the x-axis in $\mathbb{R}^3$.

One may also ask if a geometric interpretation is possible for mixed states on $\mathbb{C}^2$ described by a density operator $\rho\in\mathrm{Dens}(\mathbb{C}^2)$.

> **Proposition 3.23.** *Let $\overline{B_{\mathbb{R}^3}(\mathbf{0},1)}=\{\mathbf{x}\in\mathbb{R}^3 : \|x\|\leq 1\}$ denote the closed unit ball in $\mathbb{R}^3$. For every $\mathbf{x}\in\overline{B_{\mathbb{R}^3}(\mathbf{0},1)}$, there is a density operator*
>
> $$\rho_{\mathbf{x}}=\frac{1}{2}\big(\mathbf{1}_2+\mathbf{x}\cdot\sigma\big)=\frac{1}{2}\begin{pmatrix}1+x_3 & x_1-ix_2\\ x_1+ix_2 & 1-x_3\end{pmatrix}\in\mathrm{Dens}(\mathbb{C}^2)$$
>
> *Furthermore, any $\rho\in\mathrm{Dens}(\mathbb{C}^2)$ can be written as $\rho_{\mathbf{x}}$ for some $\mathbf{x}\in\overline{B_{\mathbb{R}^3}(\mathbf{0},1)}$. The operator $\rho_{\mathbf{x}}$ is a pure state $\rho_{\mathbf{x}}=|\psi\rangle\langle\psi|$ if and only if $\|\mathbf{x}\|=1$.*

That is, any state in $\mathbb{C}^2$ is described by an element in the unit ball of $\mathbb{R}^3$. The boundary of this ball is the Bloch sphere that describes pure states/Qubits.

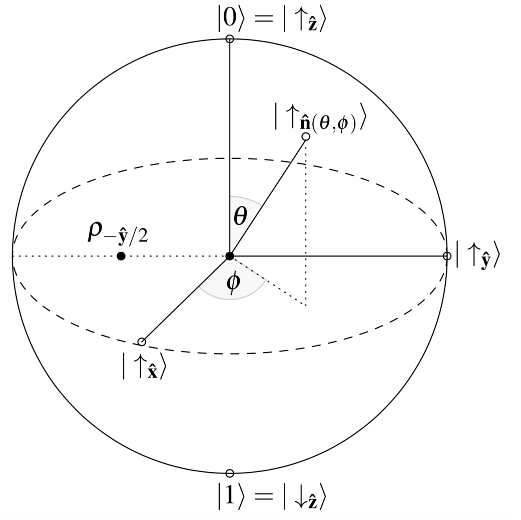

$|0\rangle = |\uparrow_{\hat{\mathbf{z}}}\rangle$
$|\uparrow_{\hat{\mathbf{n}}(\theta,\phi)}\rangle$
θ
$\rho_{-\hat{\mathbf{y}}/2}$
$|\uparrow_{\hat{\mathbf{y}}}\rangle$
ϕ
$|\uparrow_{\hat{\mathbf{x}}}\rangle$
$|1\rangle = |\downarrow_{\hat{\mathbf{z}}}\rangle$

Figure 2: The Bloch sphere $S_{\mathbb{R}^3}$ [4, Fig. 2.1]

Proof. We take a generic density matrix ρ, where the condition that $\rho = \rho^*$ and $\mathrm{Tr}(\rho) = 1$ means that ρ is of the form

$$\rho = \begin{pmatrix} a & \bar{b} \\ b & c \end{pmatrix}, \quad b \in \mathbb{C},\, a, c \in \mathbb{R}, \quad a + c = 1.$$

We write $a = \frac{1}{2}(1 + x_3)$ for $x_3 \in \mathbb{R}$, which implies that $c = \frac{1}{2}(1 - x_3)$. Then

$$\rho = \frac{1}{2}\begin{pmatrix} 1 + x_3 & 2\bar{b} \\ 2b & 1 - x_3 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 + x_3 & x_1 - ix_2 \\ x_1 + ix_2 & 1 - x_3 \end{pmatrix} = \rho_{\mathbf{x}},$$

where $2b = x_1 + ix_2 \in \mathbb{C}$. Now the matrix $\rho_{\mathbf{x}}$ has eigenvalues $\frac{1}{2}(1 \pm \|\mathbf{x}\|)$. The condition that $\rho \geq 0$ means that all eigenvalues must be non-negative. Therefore

$$\frac{1}{2}(1 \pm \|\mathbf{x}\|) \geq 0 \quad \Longleftrightarrow \quad \|\mathbf{x}\| \leq 1.$$

Finally, one computes that

$$\rho_{\mathbf{x}}^2 = \frac{1}{4}\big((1 + \|\mathbf{x}\|^2)\mathbf{1}_2 + 2\mathbf{x} \cdot \sigma\big),$$

which implies that $\rho_{\mathbf{x}}^2 = \rho_{\mathbf{x}}$ (and so $\rho_{\mathbf{x}}$ is pure) if and only if $\|\mathbf{x}\| = 1$. □

Let us now consider operators on the Qubit space $\mathbb{C}^2$. Recall that if $|\psi\rangle$ is a pure state, then so is $U|\psi\rangle$ for any unitary U, $UU^* = U^*U = \mathbf{1}$. The quantum circuits and algorithms that we will consider will be constructed from unitary operators on Qubit space (and multiple Qubit space). Hence we would like to understand well the unitary operators on $\mathbb{C}^2$.

Example 3.24 (Hadmard transform)**.** We have already seen the Hadmard transformation H, where $H|y\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle + (-1)^y|1\rangle\big)$ for $y \in \{0, 1\}$. Written in matrix form (with respect to the $\{|0\rangle, |1\rangle\}$ orthonormal basis),

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

We can also write H as a sum,

$$H = \frac{1}{\sqrt{2}}\big(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| - |1\rangle\langle 1|\big) = \frac{1}{\sqrt{2}} \sum_{x,y\in\{0,1\}} (-1)^{xy} |x\rangle\langle y|.$$

It is an easy check that $H = H^*$ and $H^2 = \mathbf{1}_2$, so H is unitary and self-adjoint on $\mathbb{C}^2$.

Exercise 3.6. Let $A = A^*$ be a linear operator on a Hilbert space $\mathcal{H}$ such that $A^2 = \mathbf{1}$. Show that for all $\alpha \in \mathbb{R}$,

$$\text{(a)} \quad e^{i\alpha A} = \sum_{n=0}^{\infty} \frac{(itA)^n}{n!} = \cos(\alpha)\mathbf{1}_2 + i\sin(\alpha)A, \qquad\qquad \text{(b)} \quad e^{i\alpha A} \text{ unitary.}$$

If we consider the matrix representation,

$$U = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ unitary}, \quad \Longrightarrow \quad |a|^2 + |b|^2 = |c|^2 + |d|^2 = 1,\ a\overline{c} + b\overline{d} = 0, \tag{3.2}$$

where $a, b, c, d \in \mathbb{C}$. For pure states, we could relate the condition $|a|^2 + |b|^2 = 1$ to the Bloch sphere. Indeed, we have a similar description of unitary operators. Recall

$$\hat{n}(\theta,\phi) = \begin{pmatrix} \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{pmatrix}, \qquad \hat{n}(\theta,\phi)\cdot\sigma = \begin{pmatrix} \cos(\theta) & \sin(\theta)e^{-i\phi} \\ \sin(\theta)e^{i\phi} & -\cos(\theta) \end{pmatrix},$$

where $\hat{n}(\theta,\phi)\cdot\sigma$ is unitary and self-adjoint. We can therefore use the previous exercise to define the spin rotations,

$$D_{\hat{n}}(\alpha) := e^{-i\frac{\alpha}{2}\hat{n}(\theta,\phi)\cdot\sigma} = \exp\big(-i\frac{\alpha}{2}\hat{n}(\theta,\phi)\cdot\sigma\big),$$

which is unitary for all $\alpha \in \mathbb{R}$.

Exercise 3.7. Show that for all $\alpha, \beta \in \mathbb{R}$,

$$D_{\hat{n}}(\alpha)D_{\hat{n}}(\beta) = D_{\hat{n}}(\alpha+\beta)$$

It turns out that the spin rotations $D_{\hat{n}}(\alpha)$ describe all unitary operators in $\mathbb{C}^2$ up to a global phase.

Theorem 3.25. *If U is unitary on $\mathbb{C}^2$, then there exist $\alpha, \xi \in \mathbb{R}$ and $\hat{n}(\theta,\phi) \in S_{\mathbb{R}^3}$, the Bloch sphere, such that*

$$U = e^{i\xi} D_{\hat{n}}(\alpha).$$

The proof of this statement is long and a little tedious, one does a careful study of the relations in Equation (3.2) to find a spin-rotation that describes a generic unitary U up to the global phase $e^{i\xi}$. We leave the details as an exercise to an enthusiastic reader.

4 Tensor products and entanglement

So far we have studied a mathematics framework that allows us to consider states of quantum mechanical systems, such as Qubits $|x\rangle = a|0\rangle + b|1\rangle \in \mathbb{C}^2$. We now turn our attention to composite systems, where a Hilbert space $\mathcal{H}$ is comprised of sub-systems $\mathcal{H}_A$ (Alice's system) and $\mathcal{H}_B$ (Bob's system).

The mathematical construction that allows us to consider such composite systems is the tensor product. Loosely speaking, given Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, the tensor product $\mathcal{H} \otimes \mathcal{K}$ is the Hilbert space that is generated by the Cartesian product $\mathcal{H} \times \mathcal{K}$ of the sets $\mathcal{H}$ and $\mathcal{K}$.

Tensor products and their properties lie at the heart of truly quantum effects, where 'long-range entanglement' leads to phenomena that has no analogue in classical methods or using a classical computer.

4.1 Tensor products of Hilbert spaces

Suppose that $\mathcal{H}$ and $\mathcal{K}$ are finite-dimensional Hilbert spaces. Our aim is to construct a new Hilbert space from pairs of elements $|\psi\rangle \in \mathcal{H}$ and $|\eta\rangle \in \mathcal{K}$. That is, the pair $(\psi, \eta) \in \mathcal{H} \times \mathcal{K}$ (as sets) gives an element $|\psi\rangle \otimes |\eta\rangle \in \mathcal{H} \otimes \mathcal{K}$.

Properties (Properties of the tensor $\otimes$)**.** We want the map $(\psi, \eta) \mapsto |\psi\rangle \otimes |\eta\rangle \in \mathcal{H} \otimes \mathcal{K}$ to satisfy the following properties:

(i) If $a \in \mathbb{C}$, then for all $|\psi\rangle \in \mathcal{H}$ and $|\eta\rangle \in \mathcal{K}$,

$$(a|\psi\rangle) \otimes |\eta\rangle = a(|\psi\rangle \otimes |\eta\rangle) = |\psi\rangle \otimes (a|\eta\rangle)$$

(ii) If $|\psi\rangle, |\psi'\rangle \in \mathcal{H}$ and $|\eta\rangle, \eta'\rangle \in \mathcal{K}$, then

$$\big(|\psi\rangle + |\psi'\rangle\big) \otimes |\eta\rangle = |\psi\rangle \otimes |\eta\rangle + |\psi'\rangle \otimes |\eta\rangle, \quad |\psi\rangle \otimes \big(|\eta\rangle + |\eta'\rangle\big) = |\psi\rangle \otimes |\eta\rangle + |\psi\rangle \otimes |\eta'\rangle.$$

The above properties imply that we have a *bilinear map* $\mathcal{H} \times \mathcal{K} \to \mathcal{H} \otimes \mathcal{K}$. The precise definition of the tensor product vector space $\mathcal{H} \otimes \mathcal{K}$ is a little technical, though intuitively it is the largest vector space that is bilinear in $\mathcal{H}$ and $\mathcal{K}$.

Definition/Theorem 4.1. *The complex vector space $\mathcal{H} \otimes \mathcal{K}$ is the universal space such that for any complex vector space E with a bilinear map $B : \mathcal{H} \times \mathcal{K} \to E$, there is a unique linear map $L : \mathcal{H} \otimes \mathcal{K} \to E$ such that $L(|\psi\rangle \otimes |\eta\rangle) = B(|\psi\rangle, |\eta\rangle)$ for all $(|\psi\rangle, |\eta\rangle) \in \mathcal{H} \times \mathcal{K}$.*

For convenience, we will also write $|\psi \otimes \eta\rangle = |\psi\rangle \otimes |\eta\rangle$.

Definition 4.2. The inner product of $\mathcal{H} \otimes \mathcal{K}$ is given by

$$\langle \psi_1 \otimes \eta_1 \mid \psi_2 \otimes \eta_2 \rangle_{\mathcal{H} \otimes \mathcal{K}} = \langle \psi_1 \mid \psi_2 \rangle_{\mathcal{H}} \, \langle \eta_1 \mid \eta_2 \rangle_{\mathcal{K}}.$$

In particular, $\|\psi \otimes \eta\| = \|\psi\| \, \|\eta\|$.

A consequence of our inner product on $\mathcal{H} \otimes \mathcal{K}$ is that $|\mathbf{0}_{\mathcal{H}} \otimes \eta\rangle$ and $|\psi \otimes \mathbf{0}_{\mathcal{K}}\rangle$ are zero vectors in $\mathcal{H} \otimes \mathcal{K}$. Suppose that $\{|e_j\rangle\}_{j=1}^n$ and $\{|f_k\rangle\}_{k=1}^m$ are orthonormal bases of $\mathcal{H}$ and $\mathcal{K}$ respectively. Then because the tensor product is bilinear, we can decompose

$$
\begin{aligned}
|\psi\rangle \otimes |\eta\rangle &= \Big(\sum_{j=1}^{n} \psi_j |e_j\rangle \Big) \otimes \Big(\sum_{k=1}^{m} \eta_k |f_k\rangle \Big) \\
&= \sum_{j=1}^{n} \sum_{k=1}^{m} \psi_j \eta_k \big(|e_j\rangle \otimes |f_k\rangle \big) = \sum_{j=1}^{n} \sum_{k=1}^{m} a_{jk} |e_j \otimes f_k\rangle,
\end{aligned}
$$

where $a_{jk} \in \mathbb{C}$ for all j and k.

Exercise 4.1. Show that $\big\{ |e_j \otimes f_k\rangle \big\}_{j=1}^{n}{}_{k=1}^{m}$ is an orthonormal basis of $\mathcal{H} \otimes \mathcal{K}$ and the dimension of $\mathcal{H} \otimes \mathcal{K}$ is nm.

Every 'ket' $|\psi \otimes \eta\rangle \in \mathcal{H} \otimes \mathcal{K}$ has a 'bra' $\langle \psi \otimes \eta| \in (\mathcal{H} \otimes \mathcal{K})^*$, where

$$
|\psi \otimes \eta\rangle = \sum_{j,k} a_{jk} |e_j \otimes f_k\rangle \quad \Longrightarrow \quad \langle \psi \otimes \eta| = \sum_{j,k} \overline{a_{jk}} \langle e_j \otimes f_k|.
$$

Properties. Given a general element $\sum_{j,k} a_{jk} |e_j \otimes f_k\rangle \in \mathcal{H} \otimes \mathcal{K}$, one may ask if we can find vectors $|\psi\rangle \in \mathcal{H}$ and $|\eta\rangle \in \mathcal{K}$ such that

$$
\sum_{j,k} a_{jk} |e_j \otimes f_k\rangle = |\psi\rangle \otimes |\eta\rangle.
$$

The answer is **NO** in general. This lack of simple decomposition has many interesting (but counter-intuitive) implications.

Exercise 4.2. Show that $\frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big) = \frac{1}{\sqrt{2}}\big(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle\big) \in \mathbb{C}^2 \otimes \mathbb{C}^2$ can not be written as a single product $|\psi\rangle \otimes |\phi\rangle$ with $|\psi\rangle \in \mathbb{C}^2$, $|\phi\rangle \in \mathbb{C}^2$.

Definition 4.3. A pure state $|\phi\rangle \in \mathcal{H} \otimes \mathcal{K}$ is called a product state if there exists $|\psi\rangle \in \mathcal{H}$ and $|\eta\rangle \in \mathcal{K}$ such that $|\phi\rangle = |\psi \otimes \eta\rangle$. Otherwise $|\phi\rangle \in \mathcal{H} \otimes \mathcal{K}$ is called an entangled state.

Tensor products can be iterated. The proof of the following is an exercise.

Lemma 4.4. *If $\mathcal{H}_A$, $\mathcal{H}_B$ and $\mathcal{H}_C$ are Hilbert spaces, then there is an isomorphism*

$$
\big(\mathcal{H}_A \otimes \mathcal{H}_B\big) \otimes \mathcal{H}_C = \mathcal{H}_A \otimes \big(\mathcal{H}_B \otimes \mathcal{H}_C\big)
$$

The triple tensor has the inner product

$$
\langle \psi_A \otimes \psi_B \otimes \psi_C \mid \phi_A \otimes \phi_B \otimes \phi_C\rangle = \langle \psi_A \mid \phi_A\rangle \, \langle \psi_B \mid \phi_B\rangle \, \langle \psi_C \mid \phi_C\rangle.
$$

And if $\{|e_j\rangle\}_{j=1}^{n_A}$, $\{|f_k\rangle\}_{k=1}^{n_B}$, $\{|g_l\rangle\}_{l=1}^{n_C}$, then

$$
|\phi_A \otimes \phi_B \otimes \phi_C\rangle = \sum_{j,k,l} a_{jkl} |e_j \otimes f_k \otimes g_l\rangle, \qquad a_{jkl} \in \mathbb{C}
$$

Example 4.5 (2-Qubit space)**.** Recalling the space of Qubits $\mathbb{C}^2 = \operatorname{span}\{|0\rangle, |1\rangle\}$, we can consider $\mathbb{C}^2 \otimes \mathbb{C}^2$, which we call the space of 2-Qubits. We can label the orthonormal basis of this vector space by hand,

$$\big\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\big\} = \big\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\big\} = \big\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\big\},$$

where in particular we can identify $\operatorname{span}\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\} = \mathbb{C}^4$, where

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad |2\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

We also notice the following:

$$0 = 00 \text{ (binary)}, \quad 1 = 01 \text{ (binary)}, \quad 2 = 10 \text{ (binary)}, \quad 3 = 11 \text{ (binary)}.$$

Exercise 4.3. Show that the Bell states

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big), \qquad |\Phi^-\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle - |11\rangle\big),$$
$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}\big(|01\rangle + |10\rangle\big), \qquad |\Psi^-\rangle = \frac{1}{\sqrt{2}}\big(|01\rangle - |10\rangle\big),$$

form an orthonormal basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$.

To extend what we do for iterated tensor products of $\mathbb{C}^2$, we recall the binary decomposition of numbers more generally.

Lemma 4.6. *For any natural number $n \geq 1$, any number $x \in \{0, 1, \dots, 2^n - 1\}$ can be decomposed*

$$x = x_{n-1}2^{n-1} + x_{n-2}2^{n-2} + \dots + x_1 2^1 + x_0 2^0 = \sum_{j=0}^{n-1} x_j 2^j,$$

where $x_j \in \{0, 1\}$ for all $j = 0, 1, \dots, n-1$.

We can therefore write $x = x_{n-1}x_{n-2} \cdots x_1 x_0$ as the binary representation of $x \in \{0, 1, \dots, 2^n - 1\}$.

Examples 4.7.

1. $$1001101110 = 2^9 + 2^6 + 2^5 + 2^3 + 2^2 + 2^1 = 622.$$

2. $$10011110001111000001000111010100000111000000 = 10873802146240$$

3. The largest known prime number (at the time of writing) is $2^{136279841} - 1$. This number can therefore be written in binary as a string of $136,279,841$ ones.

Example 4.8 (n-Qubit space)**.** We now consider the space of n-Qubits, which is defined to be the n-fold tensor product of $\mathbb{C}^2$ with itself.

$$(\mathbb{C}^2)^{\otimes n} = \underbrace{\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{n \text{ times}} = \bigotimes_{j=1}^{n} \mathbb{C}^2.$$

We know that this space is 2^n-dimensional. We will use the binary decomposition of numbers to label this basis. Namely, a basis element $|x\rangle = |x_{n-1} \cdots x_1 x_0\rangle$, where $x \in \{0, 1, \ldots, 2^n - 1\}$ and $x = \sum\limits_{j=1}^{n-1} x_j 2^j$ is its binary decomposition with $x_j \in \{0, 1\}$ for all $j = 0, 1, \ldots, n-1$. Hence we can equate $(\mathbb{C}^2)^{\otimes n} \cong \mathbb{C}^{2^n}$ by identifying the basis elements,

$$|x\rangle = |x_{n-1} \cdots x_1 x_0\rangle = |x_{n-1}\rangle \otimes |x_{n-2}\rangle \otimes \cdots \otimes |x_1\rangle \otimes |x_0\rangle.$$

Making the identification with $\mathbb{C}^{2^n}$,

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \, |1\rangle = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \, |2^n - 1\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Putting this another way, we have the equivalent bases for n-Qubit space

$$\{|x\rangle\}_{x=0}^{2^n-1} \sim \{|x_{n-1} \cdots x_1 x_0\rangle\}_{x_j \in \{0,1\}} \sim \{|x_{n-1} \otimes \cdots x_1 \otimes x_0\rangle\}_{x_j \in \{0,1\}}.$$

This presentation also gives us an efficient way to write down inner-products, where for any $x, y \in \{0, 1, \ldots, 2^{n-1} - 1\}$,

$$\begin{aligned} \langle x \mid y\rangle = \langle x_{n-1} \cdots x_1 x_0 \mid y_{n-1} \cdots y_1 y_0\rangle = \prod_{j=0}^{n-1} \langle x_j \mid y_j\rangle = \begin{cases} 1, & x_j = y_j \text{ for all } j, \\ 0 & \text{otherwise,} \end{cases} \\ = \delta_{x,y}. \end{aligned}$$

4.2 Linear operators and the partial trace

The space $\mathcal{H} \otimes \mathcal{K}$ is a Hilbert space, so our theory of linear operators also applies to linear maps $T : \mathcal{H} \otimes \mathcal{K} \to \mathcal{H} \otimes \mathcal{K}$, $T \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$. Of course, we would like to know if linear operator $A \in \mathcal{L}(\mathcal{H})$ and $B \in \mathcal{L}(\mathcal{K})$ can be combined into a linear operator on the tensor product.

Definition 4.9. If $A \in \mathcal{L}(\mathcal{H})$ and $B \in \mathcal{L}(\mathcal{K})$, $A \otimes B : \mathcal{H} \otimes \mathcal{K} \to \mathcal{H} \otimes \mathcal{K}$ is the linear operator such that $(A \otimes B)(|\psi\rangle \otimes |\eta\rangle) = A|\psi\rangle \otimes B|\eta\rangle$ for all $|\psi\rangle \in \mathcal{H}$ and $|\eta\rangle \in \mathcal{K}$. More generally,

$$(A \otimes B)\Big(\sum_{j=1}^{n}\sum_{k=1}^{m} a_{jk}(|e_j\rangle \otimes |f_k\rangle)\Big) = \sum_{j,k} a_{jk}(A|e_j\rangle \otimes B|f_k\rangle) = \sum_{j,k} a_{jk}|Ae_j \otimes Bf_k\rangle.$$

Exercise 4.4. Show that $(A \otimes B)^* = A^* \otimes B^*$ and $\operatorname{Tr}(A \otimes B) = \operatorname{Tr}(A)\operatorname{Tr}(B)$.

Like vectors in the tensor product $\mathcal{H} \otimes \mathcal{K}$, some but not all linear operators $T \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$ can be decomposed as a tensor $A \otimes B$.

Exercise 4.5. 1. Let $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}$, $|\eta_1\rangle, |\eta_2\rangle \in \mathcal{K}$. Show that

$$|\psi_1 \otimes \eta_1\rangle\langle\psi_2 \otimes \eta_2| = \big(|\psi_1\rangle\langle\psi_2|\big) \otimes \big(|\eta_1\rangle\langle\eta_2|\big).$$

2. Show that the matrix (with respect to the standard orthonormal basis of $\mathbb{C}^4$)

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in \mathcal{L}(\mathbb{C}^4)$$

can *not* be written in the form $A \otimes B$ with $A, B \in \mathcal{L}(\mathbb{C}^2)$.

Let us consider the matrix of a general operator $T \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$ with respect to the orthonormal basis $\{|e_j \otimes f_k\rangle\}_{j=1}^{n}{}_{k=1}^{m}$, where

$$T = \sum_{j,j'=1}^{n} \sum_{k,k'=1}^{m} |e_j \otimes f_k\rangle T_{jkj'k'} \langle e_{j'} \otimes f_{k'}|, \qquad T_{jkj'k'} = \langle e_j \otimes f_k \mid T(e_{j'} \otimes f_{k'})\rangle.$$

In the case that $T = A \otimes B$ with $A \in \mathcal{L}(\mathcal{H})$, $B \in \mathcal{L}(\mathcal{K})$,

$$(A \otimes B)_{jkj'k'} = \langle e_j \otimes f_k \mid Ae_{j'} \otimes Bf_{k'}\rangle = \langle e_j \mid Ae_{j'}\rangle_{\mathcal{H}} \langle f_k \mid Bf_{k'}\rangle_{\mathcal{K}} = A_{jj'}B_{kk'}.$$

To write $(A \otimes B)$ as a $(nm \times nm)$-matrix, we need to fix a labelling, where $\mathcal{H} \otimes \mathcal{K} \cong \mathbb{C}^{nm}$. We choose a labeling analogous to what we have done for multi-Qubit systems (Example 4.8),

$$|e_1 \otimes f_1\rangle \leftrightarrow \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots, \ |e_1 \otimes f_m\rangle \leftrightarrow \begin{pmatrix} 0 \\ \vdots \\ 1 \ (m\text{th slot}) \\ \vdots \\ 0 \end{pmatrix}, \ |e_2 \otimes f_1\rangle \leftrightarrow \begin{pmatrix} 0 \\ \vdots \\ 1 \ ((m+1)\text{st slot}) \\ \vdots \\ 0 \end{pmatrix},$$

$$\ldots, \ |e_j \otimes f_k\rangle \leftrightarrow \begin{pmatrix} 0 \\ \vdots \\ 1 \ (((j-1)m+k)\text{th slot}) \\ \vdots \\ 0 \end{pmatrix}, \ \ldots, \ |e_n \otimes f_m\rangle \leftrightarrow \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{4.1}$$

Following this labeling, we can write

$$
\begin{aligned}
A \otimes B &= \begin{pmatrix} A_{11}B & \cdots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{n1}B & \cdots & A_{nn}B \end{pmatrix} \\
&= \begin{pmatrix} A_{11}\begin{pmatrix} B_{11} & \cdots & B_{1m} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mm} \end{pmatrix} & \cdots & A_{1n}\begin{pmatrix} B_{11} & \cdots & B_{1m} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mm} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ A_{n1}\begin{pmatrix} B_{11} & \cdots & B_{1m} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mm} \end{pmatrix} & \cdots & A_{nn}\begin{pmatrix} B_{11} & \cdots & B_{1m} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mm} \end{pmatrix} \end{pmatrix} \\
&= \begin{pmatrix} A_{11}B_{11} & \cdots & A_{11}B_{1m} & \cdots & A_{n1}B_{1m} & \cdots & A_{1n}B_{1m} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{11}B_{m1} & \cdots & A_{11}B_{mm} & \cdots & A_{1n}B_{m1} & \cdots & A_{1n}B_{mm} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ A_{n1}B_{11} & \cdots & A_{n1}B_{1m} & \cdots & A_{nn}B_{11} & \cdots & A_{nn}B_{1m} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{n1}B_{m1} & \cdots & A_{n1}B_{mm} & \cdots & A_{nn}B_{m1} & \cdots & A_{nn}B_{mm} \end{pmatrix} \qquad (4.2)
\end{aligned}
$$

Definition 4.10. Following the labeling of the basis of $\mathcal{H} \otimes \mathcal{K}$ in Equation (4.1), the matrix product $A \otimes B$ in Equation (4.2) is called the Kronecker product of A and B.

Example 4.11. Following the Kronecker product,

$$
\begin{pmatrix} 2 & -1 \\ 0 & 3 \end{pmatrix} \otimes \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 2\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} & (-1)\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \\ 0\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} & 3\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 2 & 4 & -1 & -2 \\ 6 & 8 & -3 & -4 \\ 0 & 0 & 3 & 6 \\ 0 & 0 & 9 & 12 \end{pmatrix}.
$$

Like vector spaces, we can iterate the tensor product of operators, if $A_i \in \mathcal{L}(\mathcal{H}_i)$ for all $i = 1, \dots l$, we get a linear operator

$$
A_1 \otimes A_2 \otimes \cdots \otimes A_l = \bigotimes_{i=1}^{l} A_i \in \mathcal{L}\Big(\bigotimes_{i=1}^{l} \mathcal{H}_i\Big).
$$

Exercise 4.6. Let $H \in \mathcal{L}(\mathbb{C}^2)$ be the Hadamard operator, $H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, $H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.

1. Compute $H^{\otimes 2}$ explicitly, both in terms of the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ and the Kronecker product of matrices.

2. Show that $H^{\otimes n} \in \mathcal{L}\big((\mathbb{C}^2)^{\otimes n}\big) \cong \mathcal{L}\big(\mathbb{C}^{2^n}\big)$ can be written

$$
H^{\otimes n} = \frac{1}{2^{n/2}} \sum_{x,y=0}^{2^n-1} (-1)^{x \cdot y} |x\rangle\langle y|, \qquad (-1)^{x \cdot y} = (-1)^{x_0 y_0} \cdots (-1)^{x_{n-1} y_{n-1}},
$$

where $|x\rangle = |x_{n-1}\cdots x_1x_0\rangle$ and $|y\rangle = |y_{n-1}\cdots y_1y_0\rangle$ is the binary decomposition from Example 4.8.

Suppose that $A\otimes B = (A\otimes \mathbf{1}_{\mathcal{K}})(\mathbf{1}_{\mathcal{H}}\otimes B) \in \mathcal{L}(\mathcal{H}\otimes\mathcal{K})$. The operators $A\otimes\mathbf{1}_{\mathcal{K}}$ and $\mathbf{1}_{\mathcal{H}}\otimes B$ do not change one part of the tensor product vector space, but the operator still acts on the full tensor product. In contrast, given $A\otimes B$, we can 'trace out' one of the tensors to obtain a linear operator on the reduced Hilbert space,

$$A\otimes B \xmapsto{\mathrm{Tr}^{\mathcal{K}}} A\,\mathrm{Tr}_{\mathcal{K}}(B)\in\mathcal{L}(\mathcal{H}), \qquad A\otimes B \xmapsto{\mathrm{Tr}^{\mathcal{H}}} \mathrm{Tr}_{\mathcal{H}}(A)B\in\mathcal{L}(\mathcal{K}).$$

That is, we take the trace in one of the components of the tensor product, but ignore the other. Such an operation is called the partial trace. An important distinction between the trace and partial trace is its range

$$\text{operator} \xrightarrow{\text{Trace}} \text{complex number}, \qquad\qquad \text{operator} \xrightarrow{\text{Partial trace}} \text{operator}.$$

What is perhaps surprising is that the partial trace operation is also possible for more general operators $M\in\mathcal{L}(\mathcal{H}\otimes\mathcal{K})$ that might not have a decomposition as $A\otimes B$. To define this operation, we first take a decomposition into matrix coefficients with respect to the orthonormal basis $\{|e_j\otimes f_k\rangle\}_{j=1}^{n}{}_{k=1}^{m}$,

$$M = \sum_{j,j'=1}^{n}\sum_{k,k'=1}^{m} M_{jkj'k'}\big(|e_j\rangle\langle e_{j'}|\otimes|f_k\rangle\langle f_{k'}|\big).$$

Suppose we were to change the 'ket-bra' $|f_k\rangle\langle f_{k'}|$ to a 'bra-ket' $\langle f_{k'} \mid f_k\rangle = \delta_{k,k'}$. Then we would obtain the sum

$$\sum_{j,j'=1}^{n}\sum_{k,k'=1}^{m} M_{jkj'k'}\big(|e_j\rangle\langle e_{j'}|\otimes\delta_{k,k'}\big) = \sum_{j,j'=1}^{n}\sum_{k=1}^{m} M_{jkj'k}|e_j\rangle\langle e_{j'}|,$$

which is now a linear operator on $\mathcal{H}$. Similarly, we could switch $|e_j\rangle\langle e_{j'}|$ to $\langle e_{j'} \mid e_j\rangle = \delta_{j,j'}$ to obtain

$$\sum_{j=1}^{n}\sum_{k,k'=1}^{m} M_{jkjk'}|f_k\rangle\langle f_{k'}| \in\mathcal{L}(\mathcal{K}).$$

Theorem 4.12. *Let $M\in\mathcal{L}(\mathcal{H}\otimes\mathcal{K})$ and $\{|e_j\otimes f_k\rangle\}_{j=1}^{n}{}_{k=1}^{m}$ an orthonormal basis of $\mathcal{H}\otimes\mathcal{K}$. Define the partial trace operators*

$$\mathrm{Tr}^{\mathcal{K}}(M) = \sum_{j,j'=1}^{n}\sum_{k=1}^{m} M_{jkj'k}|e_j\rangle\langle e_{j'}|\in\mathcal{L}(\mathcal{H}),$$
$$\mathrm{Tr}^{\mathcal{H}}(M) = \sum_{j=1}^{n}\sum_{k,k'=1}^{m} M_{jkjk'}|f_k\rangle\langle f_{k'}|\in\mathcal{L}(\mathcal{K}).$$

Then $\mathrm{Tr}^{\mathcal{K}}(M)$ and $\mathrm{Tr}^{\mathcal{H}}(M)$ do not depend on the choice of orthonormal bases, $\{|e_j\rangle\}_{j=1}^{m}\subset\mathcal{H}$ and $\{|f_k\rangle\}_{k=1}^{m}\subset\mathcal{K}$, and are the unique operators such that

$$\mathrm{Tr}_{\mathcal{H}}\big(A\,\mathrm{Tr}^{\mathcal{K}}(M)\big) = \mathrm{Tr}_{\mathcal{H}\otimes\mathcal{K}}\big((A\otimes\mathbf{1}_{\mathcal{K}})M\big), \quad \textit{for all } A\in\mathcal{L}(\mathcal{H}),$$
$$\mathrm{Tr}_{\mathcal{K}}\big(\mathrm{Tr}^{\mathcal{H}}(M)B\big) = \mathrm{Tr}_{\mathcal{H}\otimes\mathcal{K}}\big(M(\mathbf{1}_{\mathcal{H}}\otimes B)\big), \quad \textit{for all } B\in\mathcal{L}(\mathcal{K}).$$

We will prove one of the defining relations and leave the other statements of the theorem as an exercise.

Proof. Let $A \in \mathcal{L}(\mathcal{H})$. Then we can calculate

$$\begin{aligned}
\operatorname{Tr}_{\mathcal{H}\otimes\mathcal{K}}\big((A\otimes \mathbf{1}_{\mathcal{K}})M\big) &= \sum_{c=1}^{n}\sum_{d=1}^{m}\big\langle e_c\otimes f_d \mid (A\otimes \mathbf{1}_{\mathcal{K}})M(e_c\otimes f_d)\big\rangle \\
&= \sum_{c=1}^{n}\sum_{d=1}^{m}\Big\langle e_c\otimes f_d \,\Big|\, (A\otimes \mathbf{1}_{\mathcal{K}})\Big(\sum_{j,j'}\sum_{k,k'} M_{jkj'k'}|e_j\otimes f_k\rangle\langle e_{j'}\otimes f_{k'}|\Big)e_c\otimes f_d\Big\rangle \\
&= \sum_{j,j'}\sum_{k,k'}\langle e_{j'}\otimes f_{k'} \mid Ae_j\otimes f_k\rangle M_{jkj'k'} \\
&= \sum_{j,j'}\sum_{k}\langle e_{j'} \mid Ae_j\rangle M_{jkj'k}
\end{aligned}$$

and compare to

$$\begin{aligned}
\operatorname{Tr}_{\mathcal{H}}\big(A\operatorname{Tr}^{\mathcal{K}}(M)\big) &= \sum_{c=1}^{n}\langle e_c \mid A\operatorname{Tr}^{\mathcal{K}}(M)e_c\rangle \\
&= \sum_{c=1}^{n}\sum_{j,j'}\sum_{k}\langle e_c \mid Ae_j\rangle\langle e_{j'} \mid e_c\rangle M_{jkj'k} \\
&= \sum_{j,j'}\sum_{k}\langle e_{j'} \mid Ae_j\rangle M_{jkj'k} = \operatorname{Tr}_{\mathcal{H}\otimes\mathcal{K}}\big((A\otimes \mathbf{1}_{\mathcal{K}})M\big).
\end{aligned}$$

The other defining relation is similar. □

Example 4.13. Let us check that the partial trace for tensor products $A \otimes B$ really reduces to the trace over one of the spaces. If $M = A \otimes B$, then $M_{j,k,j',k'} = A_{j,j'}B_{k,k'}$. Then

$$\begin{aligned}
\operatorname{Tr}^{\mathcal{K}}(M) &= \sum_{j,j'=1}^{n}\sum_{k=1}^{m} A_{jj'}B_{kk}|e_j\rangle\langle e_{j'}| \\
&= \Big(\sum_{j,j'=1}^{n} A_{jj'}|e_j\rangle\langle e_{j'}|\Big)\sum_{k=1}^{m} B_{kk} = A\operatorname{Tr}_{\mathcal{K}}(B)
\end{aligned}$$

The relation $\operatorname{Tr}^{\mathcal{H}}(A\otimes B) = \operatorname{Tr}_{\mathcal{H}}(A)B$ is similar.

Example 4.14 (Partial trace in 2-Qubit space)**.** We consider M on $\mathbb{C}^4 \cong \mathbb{C}^2 \otimes \mathbb{C}^2$ with basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, where

$$M = \sum_{x,x'\in\{0,1\}}\sum_{y,y'\in\{0,1\}} M_{xyx'y'}|xy\rangle\langle x'y'| = \begin{pmatrix} M_{0000} & M_{0001} & M_{0010} & M_{0011} \\ M_{0100} & M_{0101} & M_{0110} & M_{0111} \\ M_{1000} & M_{1001} & M_{1010} & M_{1011} \\ M_{1100} & M_{1101} & M_{1110} & M_{1111} \end{pmatrix}.$$

To distinguish the two tensor components, we write $\mathbb{C}^4 \cong \mathbb{C}^2_A \otimes \mathbb{C}^2_B$. Then

$$\begin{aligned}
\operatorname{Tr}^{B}(M) &= \sum_{x,x'\in\{0,1\}}\sum_{y\in\{0,1\}} M_{xyx'y}|x\rangle\langle x'| \\
&= (M_{0000}+M_{0101})|0\rangle\langle 0| + (M_{0010}+M_{0111})|0\rangle\langle 1| + (M_{1000}+M_{1101})|1\rangle\langle 0| \\
&\quad + (M_{1010}+M_{1111})|1\rangle\langle 1|.
\end{aligned}$$

Written using matrices,

$$\begin{pmatrix} M_{0000} & M_{0001} & M_{0010} & M_{0011} \\ M_{0100} & M_{0101} & M_{0110} & M_{0111} \\ M_{1000} & M_{1001} & M_{1010} & M_{1011} \\ M_{1100} & M_{1101} & M_{1110} & M_{1111} \end{pmatrix} \xmapsto{\mathrm{Tr}^B} \begin{pmatrix} M_{0000} + M_{0101} & M_{0010} + M_{0111} \\ M_{1000} + M_{1101} & M_{1010} + M_{1111} \end{pmatrix}.$$

Hence the partial trace is composed from the trace of M in different blocks.

Exercise 4.7. 1. Compute the matrix form of $\mathrm{Tr}^A(\mathbb{C})$ for the matrix $M \in \mathcal{L}(\mathbb{C}^4)$.

2. For any $M \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$, show that

$$\mathrm{Tr}_{\mathcal{H}}\left(\mathrm{Tr}^{\mathcal{K}}(M)\right) = \mathrm{Tr}_{\mathcal{K}}\left(\mathrm{Tr}^{\mathcal{H}}(M)\right) = \mathrm{Tr}_{\mathcal{H}\otimes\mathcal{K}}(M) \in \mathbb{C}.$$

3. If $M = M^* \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$, show that $\mathrm{Tr}^{\mathcal{H}}(M)$ and $\mathrm{Tr}^{\mathcal{K}}(M)$ are self-adjoint.

4.3 Quantum mechanics of composite systems

Let us return to quantum mechanics and the relevance of the tensor product operation.

Postulate 6 (Composite systems)**.** The Hilbert space of pure states of a composite system consisting of the subsystems $\mathcal{H}$ and $\mathcal{K}$ is described by the tensor product $\mathcal{H} \otimes \mathcal{K}$.

The space $\mathcal{H} \otimes \mathcal{K}$ is a Hilbert space, so all previous postulates and constructions regarding states, observables and measurement apply, where

$$|\psi\rangle = \sum_{j=1}^{n}\sum_{k=1}^{m} a_{jk}|e_j \otimes f_k\rangle, \qquad \sum_{j=1}^{n}\sum_{k=1}^{m} |a_{jk}|^2 = 1$$

are the pure states and

$$\rho = \sum_{l=1}^{nm} p_l |g_l\rangle\langle g_l|, \qquad \{g_l\}_{l=1}^{nm} \text{ orthonormal basis}, \qquad p_l \geq 0, \qquad \sum_{l=1}^{nm} p_l = 1$$

are the density operators describing mixed states. Of course, what is additionally of interest to us is understanding the connection of the composite systems $\mathcal{H} \otimes \mathcal{K}$ in terms of the sub-systems $\mathcal{H}$ and $\mathcal{K}$.

Example 4.15. As a warm-up let's consider $\sigma_z \otimes \sigma_z$ on $\mathbb{C}^2 \otimes \mathbb{C}^2$, where

$$\begin{aligned}(\sigma_z \otimes \sigma_z)\frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle) &= \frac{1}{\sqrt{2}}\big(\sigma_z|0\rangle \otimes \sigma_z|0\rangle \pm \sigma_z|1\rangle \otimes \sigma_z|1\rangle\big) \\ &= \frac{1}{\sqrt{2}}\big(|0\rangle \otimes 0\rangle \pm (-1)|1\rangle \otimes (-1)|1\rangle\big) \\ &= \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle).\end{aligned}$$

Hence both $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and $\frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$ and $+1$-eigenvectors of $\sigma_z \otimes \sigma_z$. We see that

$$\begin{aligned}P_1(\sigma_z \otimes \sigma_z) &= \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|) + \frac{1}{2}(|00\rangle - |11\rangle)(\langle 00| - \langle 11|) \\ &= |00\rangle\langle 00| + |11\rangle\langle 11|.\end{aligned}$$

In particular, $|00\rangle$ and $|11\rangle$ are also $+1$ eigenvectors of $\sigma_z \otimes \sigma_z$ and $\frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle) \in \operatorname{span}\{|00\rangle, |11\rangle\}$.

Exercise 4.8. Suppose that $\rho_{\mathcal{H}} \in \operatorname{Dens}(\mathcal{H})$ and $\rho_{\mathcal{K}} \in \operatorname{Dens}(\mathcal{K})$ are density operators. Show that $\rho_{\mathcal{H}} \otimes \rho_{\mathcal{K}} \in \operatorname{Dens}(\mathcal{H} \otimes \mathcal{K})$.

The tensor product of operators allows us to consider measurables $A \otimes B$ and density operators $\rho_{\mathcal{H}} \otimes \rho_{\mathcal{K}}$ from operators on the sub-systems $\mathcal{H}$ and $\mathcal{K}$. We would also like to go the other direction. Given a state or observable in the composite system $\mathcal{H} \otimes \mathcal{K}$, what can be measured/observed in the subsystem $\mathcal{H}$ or $\mathcal{K}$? It is here that the partial trace is a useful operation to take us from the composite system to a subsystem.

Example 4.16 (Partial measurements). Suppose that $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$ is a pure state and $A = A^* = \sum_j \lambda_j P_{\lambda_j} \in \mathcal{L}(\mathcal{H})$ is an observable on the subsystem $\mathcal{H}$. We can understand A in the composite system via the tensor product $A \otimes \mathbf{1}_{\mathcal{K}}$. We leave it as an exercise to show that $\sigma(A \otimes \mathbf{1}_{\mathcal{K}}) = \sigma(A)$ and so the possible outcomes of a measurement of A in $\mathcal{H}$ are the same as the possible outcomes of a measurement of $A \otimes \mathbf{1}_{\mathcal{K}}$ in $\mathcal{H} \otimes \mathcal{K}$. We can therefore ask what is the probability of measuring $\lambda \in \sigma(A \otimes \mathbf{1}_{\mathcal{K}})$ in the pure state $\psi \in |\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$. The spectral projection of $A \otimes \mathbf{1}_{\mathcal{K}}$ onto the λ-eigenspace is given by $P_\lambda \otimes \mathbf{1}_{\mathcal{K}}$ and so

$$\begin{aligned}\mathbf{P}_\psi(\lambda) &= \big\|(P_\lambda \otimes \mathbf{1}_{\mathcal{K}})\psi\big\|^2 = \langle P_\lambda \otimes \mathbf{1}_{\mathcal{K}}\rangle_\psi = \operatorname{Tr}_{\mathcal{H}\otimes\mathcal{K}}\big((P_\lambda \otimes \mathbf{1}_{\mathcal{K}})|\psi\rangle\langle\psi|\big) \\ &= \operatorname{Tr}_{\mathcal{H}}\big(P_\lambda \operatorname{Tr}^{\mathcal{K}}(|\psi\rangle\langle\psi|)\big),\end{aligned}$$

where we have used the defining property of the partial trace. Hence the probability of measuring $\lambda \in \sigma(A \otimes \mathbf{1}_{\mathcal{K}})$ in the state $|\psi\rangle$ can be determined by the spectral projection $P_\lambda \in \mathcal{L}(\mathcal{H})$ and the partial trace $\operatorname{Tr}^{\mathcal{K}}(|\psi\rangle\langle\psi|) \in \mathcal{L}(\mathcal{H})$.

We suppose that we take such a measurement and observe the value λ. Then by the projection postulate,

$$|\psi\rangle \xrightarrow{\text{Measure } \lambda} \frac{(P_\lambda \otimes \mathbf{1}_{\mathcal{K}})|\psi\rangle}{\|(P_\lambda \otimes \mathbf{1}_{\mathcal{K}})\psi\|}.$$

If $|\psi\rangle = |\phi_{\mathcal{H}}\rangle \otimes |\phi_{\mathcal{K}}\rangle$ is a product state, then

$$(P_\lambda \otimes \mathbf{1}_{\mathcal{K}})(|\phi_{\mathcal{H}}\rangle \otimes |\phi_{\mathcal{K}}\rangle) = P_\lambda|\phi_{\mathcal{H}}\rangle \otimes |\phi_{\mathcal{K}}\rangle, \quad \Longrightarrow \quad \frac{(P_\lambda \otimes \mathbf{1}_{\mathcal{K}})|\psi\rangle}{\|(P_\lambda \otimes \mathbf{1}_{\mathcal{K}})\psi\|} = \Big(\frac{P_\lambda|\phi_{\mathcal{H}}\rangle}{\|P_\lambda \phi_{\mathcal{H}}\|}\Big) \otimes |\phi_{\mathcal{K}}\rangle$$

and a measurement of an operator acting on the first tensor will not affect the second tensor.

Example 4.17. We expand upon the previous example. Consider a pure state $\sum_{j,k} a_{jk}|e_j \otimes f_k\rangle \in \mathcal{H} \otimes \mathcal{K}$ and an observable $A \in \mathcal{L}(\mathcal{H})$ in the subsystem $\mathcal{H}$. We can extend A to $A \otimes \mathbf{1}_{\mathcal{K}}$ to obtain an observable on $\mathcal{H} \otimes \mathcal{K}$, where

$$\begin{aligned}\big\langle A \otimes \mathbf{1}_{\mathcal{K}}\big\rangle_\psi &= \sum_{j,j'=1}^{n} \sum_{k,k'=1}^{n} \langle a_{j'k'} e_{j'} \otimes f_{k'} \mid (A \otimes \mathbf{1}_{\mathcal{K}}) a_{jk} e_j \otimes f_k\rangle \\ &= \sum_{j,j'} \sum_{k,k'} \overline{a_{j'k'}} a_{jk} \langle e_j' \mid A e_j\rangle_{\mathcal{H}} \langle f_{k'} \mid f_k\rangle_{\mathcal{K}} \\ &= \sum_{j,j'} \sum_{k} \overline{a_{j'k}} a_{jk} \langle e_j' \mid A e_j\rangle_{\mathcal{H}}.\end{aligned}$$

Using the partial trace, we can understand this expectation via quantities defined on $\mathcal{H}$ only,

$$\big\langle A \otimes \mathbf{1}_{\mathcal{K}}\big\rangle_\psi = \operatorname{Tr}_{\mathcal{H}\otimes\mathcal{K}}\big((A \otimes \mathbf{1}_{\mathcal{K}})|\psi\rangle\langle\psi|\big) = \operatorname{Tr}_{\mathcal{H}}\big(A \operatorname{Tr}^{\mathcal{K}}(|\psi\rangle\langle\psi|)\big) = \operatorname{Tr}_{\mathcal{H}}\big(A \operatorname{Tr}^{\mathcal{K}}(\rho_\psi)\big),$$

where the second equality is the defining property of $\mathrm{Tr}^{\mathcal{K}}$ (see Theorem 4.12). We leave it as an exercise to check that $\mathrm{Tr}^{\mathcal{K}}(\rho_\psi) \in \mathrm{Dens}(\mathcal{H})$. We find that restricting a pure state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$ to $\mathcal{H}$ will give a mixed state $\mathrm{Tr}^{\mathcal{K}}(|\psi\rangle\langle\psi|)$ such that $\langle A \otimes \mathbf{1}_{\mathcal{K}}\rangle_\psi = \langle A \rangle_{\mathrm{Tr}^{\mathcal{K}}(|\psi\rangle\langle\psi|)}$. An analogous statement holds for the restriction to $\mathcal{K}$ and observables $\mathbf{1}_{\mathcal{K}} \otimes B$.

Similarly, if $\rho \in \mathrm{Dens}(\mathcal{H} \otimes \mathcal{K})$, then

$$\langle A \otimes \mathbf{1}_{\mathcal{K}}\rangle_\rho = \mathrm{Tr}_{\mathcal{H}\otimes\mathcal{K}}\big((A \otimes \mathbf{1}_{\mathcal{K}})\rho\big) = \mathrm{Tr}_{\mathcal{H}}\big(A\,\mathrm{Tr}^{\mathcal{K}}(\rho)\big)$$

and one can show $\mathrm{Tr}^{\mathcal{K}}(\rho) \in \mathrm{Dens}(\mathcal{H})$, whence $\langle A \otimes \mathbf{1}_{\mathcal{K}}\rangle_\rho = \langle A \rangle_{\mathrm{Tr}^{\mathcal{K}}(\rho)}$.

Exercise 4.9. Suppose that $\rho \in \mathrm{Dens}(\mathcal{H} \otimes \mathcal{K})$, $A \in \mathcal{B}(\mathcal{H})$ and $B \in \mathcal{B}(\mathcal{K})$. Show that

$$\mathrm{Tr}^{\mathcal{K}}(\rho) \in \mathrm{Dens}(\mathcal{H}), \quad \mathrm{Tr}^{\mathcal{H}}(\rho) \in \mathrm{Dens}(\mathcal{K}), \quad \langle A \otimes \mathbf{1}_{\mathcal{K}}\rangle_\rho = \langle A \rangle_{\mathrm{Tr}^{\mathcal{K}}(\rho)}, \quad \langle \mathbf{1}_{\mathcal{H}} \otimes B\rangle_\rho = \langle B \rangle_{\mathrm{Tr}^{\mathcal{H}}(\rho)}.$$

Definition/Theorem 4.18. *If $\rho \in \mathrm{Dens}(\mathcal{H} \otimes \mathcal{K})$ is a density operator, we call the operators*

$$\rho_{\mathcal{H}} = \mathrm{Tr}^{\mathcal{K}}(\rho) \in \mathrm{Dens}(\mathcal{H}), \qquad \rho_{\mathcal{K}} = \mathrm{Tr}^{\mathcal{H}}(\rho) \in \mathrm{Dens}(\mathcal{K})$$

the reduced density operators of the composite system that describe the state ρ if only the subsystem $\mathcal{H}$ or $\mathcal{K}$ is observed. In particular, for all $A = A^* \in \mathcal{L}(A)$ and $B = B^* \in \mathcal{L}(B)$,

$$\langle A \rangle_{\rho_{\mathcal{H}}} = \langle A \otimes \mathbf{1}_{\mathcal{K}}\rangle_\rho, \qquad \langle B \rangle_{\rho_{\mathcal{K}}} = \langle \mathbf{1}_{\mathcal{H}} \otimes B\rangle_\rho.$$

4.4 Application 1: Superdense coding

We consider a relatively simple example that shows the power of tensor products and entangled states. We suppose there are two parties, Alice and Bob, who share a 2-Qubit $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \in \mathbb{C}^2 \otimes \mathbb{C}^2$. (Note that $|\psi\rangle = |\Phi^+\rangle$, one of the Bell states in $\mathbb{C}^2 \otimes \mathbb{C}^2$.) Alice and Bob are now very far apart but have access to one Qubit each: Alice can act on $\mathcal{H}_A = \mathbb{C}^2$ and Bob can act on $\mathcal{H}_B = \mathbb{C}^2$, but neither have access to the full space $\mathcal{H}_A \otimes \mathcal{H}_B = \mathbb{C}^2 \otimes \mathbb{C}^2$.

Alice wishes to send 2 classical bits of information to Bob, that is, one of the elements $\{00, 01, 10, 11\}$. To do this, she designs a schema: she transforms the Qubit $|\psi\rangle$ to $|\tilde{\psi}\rangle = (A \otimes \mathbf{1})|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, where the unitary element A will depend on what classical 2-bit she wishes to send. Recall that

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

and Alice uses the schema

$$\begin{aligned}
&\text{Send } 00: \quad |\psi\rangle \mapsto (\mathbf{1} \otimes \mathbf{1})|\psi\rangle = |\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle), \quad \text{(do nothing)},\\
&\text{Send } 01: \quad |\psi\rangle \mapsto (\sigma_z \otimes \mathbf{1})|\psi\rangle = \frac{1}{\sqrt{2}}(\sigma_z|0\rangle \otimes |0\rangle + \sigma_z|1\rangle \otimes |1\rangle) = \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle - |1\rangle \otimes |1\rangle)\\
&\text{Send } 10: \quad |\psi\rangle \mapsto (\sigma_x \otimes \mathbf{1})|\psi\rangle = \frac{1}{\sqrt{2}}(|1\rangle \otimes |0\rangle + |0\rangle \otimes |1\rangle),\\
&\text{Send } 11: \quad |\psi\rangle \mapsto \left(\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \otimes \mathbf{1}\right)|\psi\rangle = \frac{1}{\sqrt{2}}(-|1\rangle \otimes |0\rangle + |0\rangle \otimes |1\rangle) = \frac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle).
\end{aligned}$$

Alice has therefore transformed $|\psi\rangle$ into $|\tilde{\psi}\rangle$, which is one of the four Bell states in $\mathbb{C}^2 \otimes \mathbb{C}^2$,

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big), \qquad |\Phi^-\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle - |11\rangle\big),$$

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}\big(|01\rangle + |10\rangle\big), \qquad |\Psi^-\rangle = \frac{1}{\sqrt{2}}\big(|01\rangle - |10\rangle\big).$$

which form an orthonormal basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$.

Now, if Alice sends her part of the 2-Qubit to Bob, then Bob has access to the whole space $\mathcal{H}_A \otimes \mathcal{H}_B$ and can make a measurement using $|\tilde{\psi}\rangle$. In particular, Bob can make a measurement of the projections onto each Bell state vector,

$$P_{\Phi^\pm} = |\Phi^\pm\rangle\langle\Phi^\pm|, \qquad P_{\Psi^\pm} = |\Psi^\pm\rangle\langle\Psi^\pm|,$$

Because $|\tilde{\psi}\rangle$ is one of the Bell states, one of these projections will return a value 1 for the measurement and all the others will be 0. Bob will therefore know what transformation Alice has done to shared 2-Qubit $|\psi\rangle$ and will learn the classical 2-bit 00, 01, 10, or 11. To summarise, Alice has transmitted *two* classical bits of information to Bob via a shared 2-Qubit in which Alice only acted on her subspace $\mathcal{H}_A$. Such information transmission can not be done classically.

> **Exercise 4.10.** Suppose that $|\psi\rangle$ is one of the Bell states, $|\Phi^\pm\rangle$ or $|\Psi^\pm\rangle$. Show that
>
> $$\langle (A \otimes \mathbf{1}) \rangle_\psi = \langle (A \otimes \mathbf{1}) \rangle_{\Phi^\pm} = \langle (A \otimes \mathbf{1}) \rangle_{\Psi^\pm}.$$

The exercise shows that if a third party, Eve, intercepts Alice's transmission of $(A \otimes \mathbf{1})|\psi\rangle$ to Bob, then Eve cannot distinguish which state $(A \otimes \mathbf{1})|\psi\rangle$ is in as each expectation will return the same value. This also shows that the information that Alice sends to Bob is 'secure' as a third party must possess the entire 2-Qubit (from both the sender and receiver) to determine the transmission.

4.5 Application 2: Quantum teleportation

We again consider Alice and Bob, who share the 2-Qubit $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and are far away from each other. Alice wishes to send an arbitrary but fixed Qubit $|\psi\rangle = a|0\rangle + b|1\rangle \in \mathbb{C}^2$ to Bob, where $a, b \in \mathbb{C}$ and $|a|^2 + |b|^2 = 1$. Taking a transformation or measurement using $|\psi\rangle$ will change it, so Alice needs to be careful to make sure the same Qubit $|\psi\rangle$ is transmitted to Bob. The procedure to do this, called quantum teleportation, can be split up into several steps.

Step 0. Alice puts $|\psi\rangle$ into a composite system with the shared state $|\Phi^+\rangle$, i.e. we take the tensor product

$$\begin{aligned}|\psi_0\rangle &= |\psi\rangle \otimes |\Phi^+\rangle = (a|0\rangle + b|1\rangle) \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ &= \frac{1}{\sqrt{2}}\big(a|000\rangle + a|011\rangle + b|100\rangle + b|111\rangle\big) \\ &= \frac{1}{\sqrt{2}}\big(a|00\rangle \otimes |0\rangle + a|01\rangle \otimes |1\rangle + b|10\rangle \otimes |0\rangle + b|11\rangle \otimes |1\rangle\big),\end{aligned}$$

where the last vector is in $\mathcal{H}_A \otimes \mathcal{H}_B = \mathbb{C}^4 \otimes \mathbb{C}^2$, the tensor product of spaces that Alice and Bob have control over.

Step 1. Alice can manipulate $|\psi_0\rangle$ in the space $\mathcal{H}_A = \mathbb{C}^4$. She uses the CNOT (controlled not) gate U_{CN}, where in the orthonormal basis $\{|x, y\rangle\}_{x,y \in \{0,1\}}$ of $\mathbb{C}^4$, $U_{CN}|x, y\rangle = |x, y \oplus x\rangle$ (with addition modulo 2). Explicitly,

$$U_{CN}|00\rangle = |00\rangle, \qquad U_{CN}|01\rangle = |01\rangle, \qquad U_{CN}|10\rangle = |11\rangle, \qquad U_{CN}|11\rangle = |10\rangle.$$

Written as a matrix in the computational basis

$$U_{CN} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Acting by U_{CN} on $\mathcal{H}_A$, we obtain

$$\begin{aligned}|\psi_1\rangle = (U_{CN}\otimes \mathbf{1}_2)|\psi_0\rangle &= \frac{1}{\sqrt{2}}(U_{CN}\otimes \mathbf{1}_2)\big(a|00\rangle\otimes|0\rangle + a|01\rangle\otimes|1\rangle + b|10\rangle\otimes|0\rangle + b|11\rangle\otimes|1\rangle\big)\\ &= \frac{1}{\sqrt{2}}\big(a|00\rangle\otimes|0\rangle + a|01\rangle\otimes|1\rangle + b|11\rangle\otimes|0\rangle + b|10\rangle\otimes|1\rangle\big).\end{aligned}$$

Step 2. We then apply the Hadamard gate to the first Qubit; that is, we consider $(H\otimes\mathbf{1}_2\otimes\mathbf{1}_2)|\psi_1\rangle$. First we write

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\big(a|0\rangle(|00\rangle + |11\rangle) + b|1\rangle(|10\rangle + 01\rangle)\big)$$

where we are using the shorthand notation $|x\rangle|y\rangle = |xy\rangle = |x\rangle\otimes|y\rangle$. Applying the Hadamard gate,

$$\begin{aligned}|\psi_2\rangle = (H\otimes\mathbf{1}_2\otimes\mathbf{1}_2)|\psi_1\rangle &= \frac{1}{2}\big(a(|0\rangle+|1\rangle)(|00\rangle+|11\rangle) + b(|0\rangle-|1\rangle)(|10\rangle+|01\rangle)\big)\\ &= \frac{1}{2}\big(a(|000\rangle+|011\rangle+|100\rangle+|111\rangle) + b(|010\rangle+|001\rangle-|110\rangle-|101\rangle)\big)\\ &= \frac{1}{2}\big(|00\rangle(a|0\rangle+b|1\rangle) + |01\rangle(a|1\rangle+b|0\rangle) + |10\rangle(a|0\rangle-b|1\rangle) + |11\rangle(a|1\rangle-b|0\rangle)\big),\end{aligned}$$

where we have written the vector as an element in $\mathcal{H}_A\otimes\mathcal{H}_B = \mathbb{C}^4\otimes\mathbb{C}^2$.

Step 3. Alice can now take a measurement *in* $\mathcal{H}_A = \mathbb{C}^4$. In particular, she can measure $P_{|00\rangle}$, $P_{|01\rangle}$, $P_{|10\rangle}$ and $P_{|11\rangle}$, the projection onto the span of each orthonormal basis vector. One of these measurements will give the value 1 and all the others will be 0. From this measurement, the part of $|\psi_2\rangle$ in $\mathcal{H}_A$ will collapse to whichever orthonormal basis element was detected.

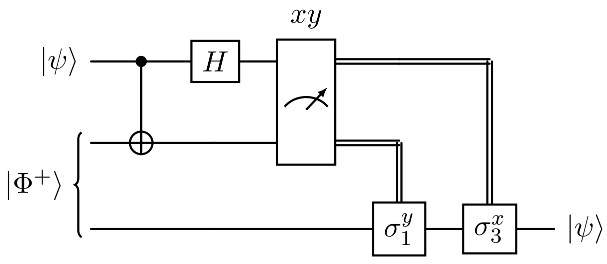
$$\begin{aligned}&\text{Detect } P_{|00\rangle}: && |\psi_2\rangle \longrightarrow |\psi_3\rangle = |00\rangle(a|0\rangle+b|1\rangle),\\ &\text{Detect } P_{|01\rangle}: && |\psi_2\rangle \longrightarrow |\psi_3\rangle = |01\rangle(a|1\rangle+b|0\rangle),\\ &\text{Detect } P_{|10\rangle}: && |\psi_2\rangle \longrightarrow |\psi_3\rangle = |10\rangle(a|0\rangle-b|1\rangle),\\ &\text{Detect } P_{|11\rangle}: && |\psi_2\rangle \longrightarrow |\psi_3\rangle = |11\rangle(a|1\rangle-b|0\rangle).\end{aligned}$$

Alice then needs to communicate the result of this measurement to Bob (which prevents faster-than-light transmission of information).

Step 4. Bob is given the result of the measurement by Alice. He can then perform a unitary transformation on the Qubit $|\psi_3\rangle$ in $\mathcal{H}_B$ to obtain the original Qubit $|\psi\rangle = a|0\rangle + b|1\rangle$ that Alice wanted to send.

$$\begin{aligned}00: &\quad |\psi_3\rangle = |00\rangle(a|0\rangle+b|1\rangle) = |00\rangle\otimes|\psi\rangle, \quad \text{(do nothing)},\\ 01: &\quad (\mathbf{1}\otimes\sigma_1)|\psi_3\rangle = \mathbf{1}|01\rangle\otimes\sigma_1(a|1\rangle+b|0\rangle) = |01\rangle\otimes(a|0\rangle+b|1\rangle) = |01\rangle\otimes|\psi\rangle,\\ 10: &\quad (\mathbf{1}\otimes\sigma_3)|\psi_3\rangle = \mathbf{1}|10\rangle\otimes\sigma_3(a|0\rangle-b|1\rangle) = |10\rangle\otimes(a|0\rangle+b|1\rangle) = |10\rangle\otimes|\psi\rangle,\\ 11: &\quad (\mathbf{1}\otimes\sigma_3\sigma_1)|\psi_3\rangle = \mathbf{1}|11\rangle\otimes\sigma_3\sigma_1(a|1\rangle-b|0\rangle) = |11\rangle\otimes\sigma_z(a|0\rangle-b|1\rangle) = |11\rangle\otimes|\psi\rangle.\end{aligned}$$

To write this compactly, if the outcome is xy with $x,y\in\{0,1\}$, then Bob applies $\sigma_3^x\sigma_1^y$ to recover $|\psi\rangle$. Diagrammatically, we can represent the procedure as follows



where a wire with double-lines denotes a classical information channel.

Let us explain further why quantum teleportation does not violate special relativity, which says that information cannot travel faster than the speed of light. Before Alice takes her measurement in $\mathcal{H}_A$, the system is in the state

$$\frac{1}{2}\big(|00\rangle(a|0\rangle + b|1\rangle) + |01\rangle(a|1\rangle + b|0\rangle) + |10\rangle(a|0\rangle - b|1\rangle) + |11\rangle(a|1\rangle - b|0\rangle)\big)$$
$$= \frac{1}{2}\big(|\psi_{00}\rangle + |\psi_{01}\rangle + |\psi_{10}\rangle + |\psi_{11}\rangle\big).$$

Taking the measurement will collapse this state to $|\psi_{xy}\rangle$ for some $x, y \in \{0, 1\}$. Each outcome has probability $\frac{1}{4}$ of occurring. From Bob's perspective, the state after the measurement by Alice will be in one of four unknown states, each with probability $\frac{1}{4}$. Therefore, before Bob receives Alice's transmission, he can describe the composite system via the density operator

$$\rho = \frac{1}{4}\big(|\psi_{00}\rangle\langle\psi_{00}| + |\psi_{01}\rangle\langle\psi_{01}| + |\psi_{10}\rangle\langle\psi_{10}| + |\psi_{11}\rangle\langle\psi_{11}|\big) \in \mathrm{Dens}(\mathcal{H}_A \otimes \mathcal{H}_B).$$

By taking the reduced density operator $\rho_B = \mathrm{Tr}^{\mathcal{H}_A}(\rho) \in \mathrm{Dens}(\mathcal{H}_B)$, Bob can try and find out more about the state before Alice's message arrives. However, we find the following.

> **Exercise 4.11.** Show that $\rho_B = \mathrm{Tr}^{\mathcal{H}_A}(\rho) = \frac{1}{2}\mathbf{1}$.

The exercise shows that Bob does not learn anything about the state $|\psi\rangle$ by considering ρ_B. Instead he must wait for the result of Alice's measurement to be relayed to him via a classical information channel, whose transmission is limited by the speed of light.

If a classical communication channel is required between Alice and Bob, why does Alice simply not take a measurement using $|\psi\rangle$ to determine its nature and then communicate this result to Bob? The problem is that $|\psi\rangle$ is an arbitrary Qubit and so can correspond to any point on the Bloch sphere. So there are uncountably many possibilities for the state $|\psi\rangle$ and it would take an infinite amount of time to determine to communicate this information to Bob. So the use of the shared 2-Qubit $|\Phi^+\rangle$ and entanglement was crucial.

4.6 Schmidt decomposition

We start with a few more results from linear algebra.

> **Exercise 4.12.** For any linear operator $A \in \mathcal{L}(\mathcal{H})$, show that $\mathrm{Ker}(A) = \mathrm{Ker}(A^*A)$, where $\mathrm{Ker}(A) = \{|\psi\rangle \in \mathcal{H} \mid A|\psi\rangle = \mathbf{0}\}$.

The following should be compared to the polar form of complex numbers $z = e^{i\theta}|z|$.

> **Lemma 4.19** (Polar decomposition). *Let $\mathcal{H}$ be a finite-dimensional Hilbert space and $A \in \mathcal{L}(\mathcal{H})$, then $A = U|A|$, where $U \in \mathcal{L}(\mathcal{H})$ is unitary and $|A| \in \mathcal{L}(\mathcal{H})$ is positive.*

Proof (Proof sketch). We define $|A| = \sqrt{A^*A}$. Namely, A^*A is self-adjoint and so has a spectral decomposition $A^*A = \sum\limits_{j,\alpha} \mu_j |e_{j,\alpha}\rangle\langle e_{j,\alpha}|$ with $\mu_j \geq 0$. We define

$$|A| = \sqrt{A^*A} = \sum_{j,\alpha} \sqrt{\mu_j}|e_{j,\alpha}\rangle\langle e_{j,\alpha}|.$$

We then define, $|f_{j,\alpha}\rangle = \frac{1}{\sqrt{\mu_j}}A|e_{j,\alpha}\rangle$ for all j such that $\mu_j \neq 0$. Using that $|e_{j,\alpha}\rangle$ are eigenvectors of A^*A, one can check that $\{|f_{j,\alpha}\rangle\}$ is an orthonormal set, which can then be completed into an orthonormal basis of $\mathcal{H}$, which we also denote by $\{|f_{j,\alpha}\rangle\}$. Defining the operator

$$U = \sum_{k,\beta} |f_{k,\beta}\rangle\langle e_{k,\beta}|, \qquad U \text{ unitary},$$

we find that for $\mu_j \neq 0$,

$$U|A||e_{j,\alpha}\rangle = U\sqrt{\mu_j}|e_{j,\alpha}\rangle = \sum_{k,\beta} |f_{k,\beta}\rangle\langle e_{k,\beta} \mid \sqrt{\mu_j} e_{j,\alpha}\rangle = \sqrt{\mu_j}|f_{j,\alpha}\rangle = A|e_{j,\alpha}\rangle.$$

For the case that $\mu_j = 0$, $|e_{j,k}\rangle \in \operatorname{Ker}(A^*A) = \operatorname{Ker}(\sqrt{A^*A}) = \operatorname{Ker}(A)$ and so $A|e_{j,\alpha}\rangle = 0 = U|A|e_{j,\alpha}\rangle$. Because $U|A|$ and A agree on an orthonormal basis, it follows that $A = U|A|$. □

When $A : \mathbb{C}^n \to \mathbb{C}^m$ is a linear operator between spaces of different dimension (a non-square matrix), then a polar decomposition $A = U|A|$ is still possible, where $|A| = \sqrt{A^*A} \in M_{n\times n}(\mathbb{C})$ and $U \in M_{m\times n}(\mathbb{C})$ is such that $U^*U = \mathbf{1}_n$.

Lemma 4.20 (Singular value decomposition)**.** *Let $\mathcal{H}$ be finite-dimensional and $A \in \mathcal{L}(\mathcal{H})$. Then there exists unitary operators U and V and a positive and diagonalisable operator D such that $A = UDV$*

Note that the operator A does not have to be diagonalisable.

Proof. The result is a corollary of (and is equivalent to) the polar decomposition. Writing $A = U'|A|$, we can diagonalise $U'|A| = U'WDW^*$, where D is diagonal and W is unitary. We then let $U = U'W$ and $V = W^*$, so $A = UDV$. □

Exercise 4.13. Extend the polar and singular value decomposition to non-square matrices.

Returning to tensor products, given a pure state $|\psi\rangle \in \mathcal{H}\otimes\mathcal{K}$, taking an orthonormal basis of $\mathcal{H}$ and $\mathcal{K}$, we can decompose

$$|\psi\rangle = \sum_{j,k} a_{jk}|e_j\rangle \otimes |f_k\rangle.$$

Using the singular value decomposition, we can reduce this double-sum to a single sum of basis vectors *that depend on the original state* $|\psi\rangle$. This is the Schmidt decomposition of a pure state in the tensor product. We will state and prove this result for the case of $\mathcal{H}$ and $\mathcal{K}$ have the same dimension (so $\mathcal{H} \cong \mathcal{K}$).

Proposition 4.21 (Schmidt decomposition)**.** *Let $|\psi\rangle \in \mathcal{H}\otimes\mathcal{H}$. Then there are orthonormal sets $\{|\tilde{e}_l\rangle\}$ and $\{|\tilde{f}_l\rangle\}$ in $\mathcal{H}$ and non-negative numbers $\lambda_l \geq 0$ such that*

$$|\psi\rangle = \sum_l \lambda_l |\tilde{e}_l\rangle \otimes |\tilde{f}_l\rangle.$$

Proof. The vector $|\psi\rangle$ has a generic decomposition $|\psi\rangle = \sum_{j,k} a_{jk}|e_j \otimes f_k\rangle$ with $a_{jk} \in \mathbb{C}$ for all j,k. The coefficients $\{a_{jk}\}_{j,k}$ then specify a matrix A for which we can take the singular value decomposition $A = UDV$. We can then expand

$$a_{jk} = (UDV)_{jk} = \sum_{l,m} U_{jl}D_{lm}V_{mk} = \sum_l U_{jl}D_{ll}V_{lk} = \sum_l \sqrt{\mu_l}U_{jl}V_{lk},$$

where we have used that $D = W^*|A|W$ is diagonal with eigenvalues $\sqrt{\mu_j} \geq 0$, i.e. $D_{lm} = \delta_{l,m}\sqrt{\mu_l}$.

Substituting a_{jk} back into our decomposition of $|\psi\rangle$,

$$\begin{aligned}|\psi\rangle &= \sum_{j,k,l} \sqrt{\mu_l}\, U_{jl} V_{lk} |e_j\rangle \otimes |f_k\rangle = \sum_l \sqrt{\mu_l} \Big(\sum_j U_{jl} |e_j\rangle \Big) \otimes \Big(\sum_k V_{lk} |f_k\rangle \Big) \\ &=: \sum_l \lambda_l |\tilde{e}_l\rangle \otimes |\tilde{f}_l\rangle.\end{aligned}$$

Because U and V are unitary operators, it follows that $|\tilde{e}_l\rangle = \sum_j U_{jl}|e_j\rangle$ and $|\tilde{f}_l\rangle = \sum_k V_{lk}|f_k\rangle$ are orthonormal sets. □

We emphasise that the decomposition is via orthonormal sets $\{|\tilde{e}_l\rangle\}$ and $\{|\tilde{f}_l\rangle\}$, which depend on the state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}$. So the decomposition is quite different to a generic decomposition with respect to an orthonormal basis.

Exercise 4.14. Prove the Schmidt decomposition for $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$, where $\mathcal{H}$ and $\mathcal{K}$ have different dimension.

If $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$ is a product state $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is a product state, then the Schmidt decomposition is not necessary as $|\psi\rangle$ is already written as a single product of states. Therefore the number of terms that appear in the Schmidt decomposition give us an insight into the level of entanglement of state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$.

Definition 4.22. For a pure state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$ with Schmidt decomposition $\sum_l \lambda_l |\tilde{e}_l\rangle \otimes |\tilde{f}_l\rangle$, the number of non-zero values of λ_l is called the Schmidt number for $|\psi\rangle$.

Exercise 4.15. Show that a pure state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}$ has Schmidt number 1 if and only if it is a product state.

For a pure state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, the Schmidt decomposition also gives us an easy way to compute the reduced density matrices $\rho_A = \mathrm{Tr}^{\mathcal{H}_B}(|\psi\rangle\langle\psi|)$ and $\rho_B = \mathrm{Tr}^{\mathcal{H}_A}(|\psi\rangle\langle\psi|)$. Namely,

$$\begin{aligned}\rho_B = \mathrm{Tr}^{\mathcal{H}_B}(|\psi\rangle\langle\psi|) &= \mathrm{Tr}^{\mathcal{H}_A} \Big(\sum_{l,l'} \lambda_l \lambda_{l'} \left(|\tilde{e}_l\rangle\langle\tilde{e}_{l'}| \otimes |\tilde{f}_l\rangle\langle\tilde{f}_{l'}| \right) \Big) \\ &= \sum_l \lambda_l^2 \, |\tilde{f}_l\rangle\langle\tilde{f}_l|\end{aligned}$$

and similarly $\rho_A = \sum_l \lambda_l^2 \, |\tilde{e}_l\rangle\langle\tilde{e}_l|$. We see that ρ_A and ρ_B are in diagonal form with the *same* eigenvalues λ_l^2.

A closely related procedure to the Schmidt decomposition of a pure state is the purification of a non-pure state. Given a state $\rho_A \in \mathrm{Dens}(\mathcal{H}_A)$, can we find an auxiliary Hilbert space $\mathcal{H}_B$ and a pure state $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ such that $\mathrm{Tr}^{\mathcal{H}_B}(|\psi_{AB}\rangle\langle\psi_{AB}|) = \rho_A$? That is, we would like to consider the possibly non-pure state ρ_A as coming from a restriction of a pure state via the partial trace. It is an exercise to show that such a purification of $\rho_A \in \mathrm{Dens}(\mathcal{H}_A)$ to $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is possible. The new Hilbert space $\mathcal{H}_B$ that appears in the purification is largely a mathematical object and may or may not have any physical relevance.

4.7 The EPR experiment and Bell's inequality

Let us briefly reflect on some of the implications of the theory that we have developed thus far. It matches our day to day experience that if no one is around, the position of pen on a desk or a book in a bookshelf has a definite value or property that is independent of whether it is observed or not. This is different to the theory of quantum mechanics that we have introduced. Instead, a physical property of an object (e.g. the momentum, spin in the y-direction) does not have a definite value. Instead, this must be observed by a measurement that will return certain values with certain probabilities.

The quantum mechanical description of reality was challenged by a thought experiment by the physicists Einstein, Podolsky and Rosen (EPR). A simplified version of the thought experiment concerns the entangled 2-Qubit $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. We again consider the case that Alice and Bob share this 2-Qubit and are very far from each other. Suppose that we consider the observable $\sigma_z \otimes \sigma_z$. A simple computation will gives that

$$(\sigma_z \otimes \sigma_z)\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) = \frac{1}{\sqrt{2}}\big((-1)|01\rangle - (-1)|10\rangle\big) = -\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

and so $|\Psi^-\rangle$ is a (-1)-eigenvector of $\sigma_z \otimes \sigma_z$. So if Alice were to perform a measurement of σ_z on her Qubit, if she observes ± 1, then she knows with certainty that Bob must measure the value ∓ 1. Therefore she knows the outcome of Bob's measurement without the need for any communication channel between the two parties.

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \xrightarrow{\text{Alice measures } +1} |01\rangle \text{ (Bob must measure -1)},$$
$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \xrightarrow{\text{Alice measures -1}} |10\rangle \text{ (Bob must measure +1)}.$$

This basic result can be generalised further.

Exercise 4.16. Let $|\uparrow_{\hat{n}}\rangle = e^{\frac{-i\phi}{2}}\cos(\frac{\theta}{2})|0\rangle + e^{\frac{i\phi}{2}}\sin(\frac{\theta}{2})|1\rangle$ and $|\downarrow_{\hat{n}}\rangle = -e^{\frac{-i\phi}{2}}\sin(\frac{\theta}{2})|0\rangle + e^{\frac{i\phi}{2}}\cos(\frac{\theta}{2})|1\rangle$ denote the ± 1 eigenvalues of $\hat{n}(\theta,\phi)\cdot\sigma$. Show that the pure state

$$\frac{1}{\sqrt{2}}|\uparrow_{\hat{n}}\downarrow_{\hat{n}}\rangle - |\downarrow_{\hat{n}}\uparrow_{\hat{n}}\rangle = \frac{1}{\sqrt{2}}\big(|\uparrow_{\hat{n}}\rangle \otimes |\downarrow_{\hat{n}}\rangle - |\downarrow_{\hat{n}}\rangle \otimes |\uparrow_{\hat{n}}\rangle\big) \sim |\Psi^-\rangle$$

where $|\psi_1\rangle \sim |\psi_2\rangle$ if $|\psi_2\rangle = e^{i\alpha}|\psi_1\rangle$ for some $\alpha \in \mathbb{R}$. In particular, $\frac{1}{\sqrt{2}}|\uparrow_{\hat{n}}\downarrow_{\hat{n}}\rangle - |\downarrow_{\hat{n}}\uparrow_{\hat{n}}\rangle$ is a (-1)-eigenvector of $(\hat{n}(\theta,\phi)\cdot\sigma)\otimes(\hat{n}(\theta,\phi)\cdot\sigma)$ and $\big\langle(\hat{n}(\theta,\phi)\cdot\sigma)\otimes(\hat{n}(\theta,\phi)\cdot\sigma)\big\rangle_{|\Psi^+\rangle} = -1$.

The exercise shows that any measurement of ± 1 of $(\hat{n}(\theta,\phi)\cdot\sigma)\otimes(\hat{n}(\theta,\phi)\cdot\sigma)$ in Alice's Qubit will ensure that Bob measures ∓ 1. EPR argued that Alice's measurement will give Bob's Qubit an intrinsic property independent of measurement and this is inconsistent with quantum mechanics.

We can further probe this question by considering a 'classical' and 'quantum' version of the same experiment. The classical picture goes as follows. We have 4 quantities Q, R, S, T, which each have an intrinsic and set physical property ± 1 that can be measured. We emphasise that value ± 1 is already set before any experiment takes place and one simply makes an observation to determine this number. A third party, Eve, prepares the quantities Q, R, S, T and sends Q, R to Alice and S, T to Bob to measure (where Alice and Bob are very far apart). Because each value is either ± 1, some basic algebra will show that

$$QS + RS + RT - QT = (Q + R)S + (R - Q)T = \pm 2.$$

Depending on how objects Q, R, S, T are prepared, the outcome of each combination of ± 1 for all 4 objects has a certain probability. (Note: this is still a classical system, where the probability is determined by how Eve prepares each object, rather than the probability of a measurement outcome.) This allows us to take a classical expectation, where one finds that

$$\mathbf{E}\big(QS + RS + RT - QT\big) = \mathbf{E}(QS) + \mathbf{E}(RS) + \mathbf{E}(RT) - \mathbf{E}(QT) \leq 2,$$

the so-called Bell or CHSH (Clauser–Horne–Shimony–Holt) inequality. By repeating the experiment enough times, Alice and Bob are able to determine the expectation of their measurement and determine the left hand side of the Bell inequality with good accuracy.

We now consider a quantum mechanical version of the above experiment. The third party Eve prepares the 2-Qubit state $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. The first Qubit is sent to Alice and the second is sent to Bob, where the two take (quantum mechanical) measurements of the following observables

$$Q = \sigma_z, \qquad R = \sigma_x, \qquad S = \frac{-1}{\sqrt{2}}(\sigma_z + \sigma_x), \qquad T = \frac{1}{\sqrt{2}}(\sigma_z - \sigma_x).$$

Note that Q, R, S, T are all self-adjoint unitaries with eigenvalues ± 1, so our set of observables matches the classical experiment. In a quantum mechanical description, if Alice measures Q and Bob measures S, then observable of interest in the composite system is $Q \otimes S \in \mathcal{L}(\mathbb{C}^2 \otimes \mathbb{C}^2)$.

> **Exercise 4.17.** Show that
>
> $$\langle Q \otimes S\rangle_{\Psi^-} = \frac{1}{\sqrt{2}}, \qquad \langle R \otimes S\rangle_{\Psi^-} = \frac{1}{\sqrt{2}}, \qquad \langle R \otimes T\rangle_{\Psi^-} = \frac{1}{\sqrt{2}}, \qquad \langle Q \otimes T\rangle_{\Psi^-} = -\frac{1}{\sqrt{2}},$$

The exercise shows that

$$\langle Q \otimes S\rangle_{\Psi^-} + \langle R \otimes S\rangle_{\Psi^-} + \langle R \otimes T\rangle_{\Psi^-} - \langle Q \otimes T\rangle_{\Psi^-} = \frac{4}{\sqrt{2}} = 2\sqrt{2},$$

a violation of the Bell inequality. This incompatibility between classical and quantum approaches to the same experiment was noted by Bell in the 1960s. In recent decades, experimental technology and precision has reached a point that the EPR experiment can be carried out in such a way that the expectations of all quantities are precise enough that one can determine if there is abound of 2 or $2\sqrt{2}$, for example using polarised photons for Qubits. The experimental evidence is overwhelmingly in favour of a quantum mechanical description, see [6] for example. In contemporary physics, we consider phenomena that violate a version of Bell's inequality to be truly quantum phenomena. In particular, tensor products and entanglement are required to explain such physical systems.

There are, of course, many theoretical (and philosophical) implications of the violation of Bell's inequality. In the context of this course, it is enough to recognise that a quantum mechanical description of reality, which involves entanglement and a probabilistic approach to measurement and physical properties, is experimentally supported. The postulates of quantum mechanics do not always match our daily experience with reality, but our day-to-day experience occurs at a particular length scale and might not be applicable what happens at the scale of nanometres or galaxies.

5 Quantum circuits

In this chapter, we turn to the question of using quantum mechanics as a tool of *computation.* Specifically, given an n-Qubit system $\mathcal{H} = (\mathbb{C}^2)^{\otimes n} \cong \mathbb{C}^{2^n}$, we aim to study various operations/algorithms that can be implemented using the quantum mechanical techniques that we have previously studied.

An unsatisfying definition of a quantum circuit is any unitary operator U acting on $(\mathbb{C}^2)^{\otimes n}$. Such a definition does not help with the question of how to build or realise such a circuit in the real world. Instead, our aim is to understand how computational questions and unitary operations on Qubits can be decomposed into a discrete collection of basic operations, which are easier to understand and potentially realisable in a laboratory.

5.1 Classical logic gates and circuits

Before we begin our study of quantum gates and circuits, it is worth doing a brief review of the classical setting. Here information is transmitted by a string of classical bits, an element $(x_1, \dots, x_n) \in \{0,1\}^n$. We can think of each $x_j \in \{0,1\}$ as representing a logical check, TRUE/FALSE or YES/NO.

Definition 5.1. A classical logical gate is a map

$$\begin{aligned} &g : \{0,1\}^n \to \{0,1\}^m, \\ &(x_1, \dots, x_n) \overset{g}{\mapsto} g(x_1, \dots, x_n) = \big(g_1(x_1, \dots, x_n), \dots, g_m(x_1, \dots, x_n)\big). \end{aligned}$$

We say that g is reversible if it is a bijection.

We review some elementary examples.

Examples 5.2.

1. The identity gate
$$\mathbf{ID} : \{0,1\} \to \{0,1\}, \qquad \mathbf{ID}(x) = x.$$

2. Addition modulo 2,
$$\{0,1\}^2 \xrightarrow{\oplus} \{0,1\}, \qquad (x,y) \mapsto x + y \bmod 2 =: x \oplus y.$$

3. The NOT gate,
$$\mathbf{NOT} : \{0,1\} \to \{0,1\}, \qquad \mathbf{NOT}(x) = 1 \oplus x,$$
where $\mathbf{NOT}(0) = 1$ and $\mathbf{NOT}(1) = 0$.

4. The AND gate,
$$\mathbf{AND} : \{0,1\}^2 \to \{0,1\}, \qquad \mathbf{AND}(x,y) = xy,$$
where $\mathbf{AND}(0,0) = \mathbf{AND}(0,1) = \mathbf{AND}(1,0) = 0$ and $\mathbf{AND}(1,1) = 1$.

5. The OR gate
$$\mathbf{OR} : \{0,1\}^2 \to \{0,1\}, \qquad \mathbf{OR}(x,y) = x \oplus y \oplus xy,$$
where $\mathbf{OR}(0,0) = 0$ and $\mathbf{OR}(0,1) = \mathbf{OR}(1,0) = \mathbf{OR}(1,1) = 1$.

6. The exclusive OR gate,
$$\mathbf{XOR} : \{0,1\}^2 \to \{0,1\}, \qquad \mathbf{XOR}(x,y) = x \oplus y.$$

7. The COPY gate,

$$\mathbf{COPY} : \{0,1\} \to \{0,1\}^2, \qquad \mathbf{COPY}(x) = (x,x).$$

8. The TOFFOLI gate,

$$\mathbf{TOF} : \{0,1\}^3 \to \{0,1\}^3, \qquad \mathbf{TOF}(x_1,x_2,x_3) = (x_1,x_2,x_1x_2 \oplus x_3).$$

To build (classical) circuits, we need a procedure to construct new (classical) bits and gates.

Properties. To build new bits and gates from existing ones, we allow ourselves the following operations,

1. (Padding) We can increase a string of bits $\{0,1\}^n \to \{0,1\}^{n+l}$ by inserting $y_1,\dots,y_l \in \{0,1\}$ at chosen points in the string.
2. (Restriction/Reordering) For $m \le n$, we can reorder or restrict bits $\{0,1\}^n \to \{0,1\}^m$ by $(x_1,\dots,x_n) \mapsto (x_{j_1},\dots,x_{j_m})$.

Similarly, given gates $g_1,\dots,g_K$, we can build new gates by the following operations,

1. (Composition) For any j,k we can consider $g_j \circ g_k$, padding and restricting as necessary so that there are no domain issues.
2. (Cartesian products). For any $g_j : \{0,1\}^{n_j} \to \{0,1\}^{m_j}$ and $g_k : \{0,1\}^{n_k} \to \{0,1\}^{m_k}$, we have the gate

$$g_j \times g_k : \{0,1\}^{n_j+n_k} \to \{0,1\}^{m_j+m_k} = \{0,1\}^{m_j} \times \{0,1\}^{m_k},$$
$$(g_j \times g_k)\big(x_1,\dots,x_{n_j},x_{n_j+1},\dots,x_{n_j+n_k}\big) = \big(g_j(x_1,\dots,x_{n_j}), g_k(x_{n_j+1},\dots,x_{n_j+n_k})\big).$$

Given the classical gates $g_1,\dots,g_K$, a (finite) combination of the operations above is called a logical circuit.

Definition 5.3. We say that a set of classical gates $\{g_1,\dots,g_K\}$ is universal if any classical gate g is a logical circuit of $\{g_1,\dots,g_K\}$. That is, g can be constructed from padding, restriction/reordering, composition and Cartesian products of $\{g_1,\dots,g_K\}$.

We leave the proof of the following as an exercise to an enthusiastic reader.

Theorem 5.4 ([4, Example 5.5, Theorem 5.6])**.** *1. The TOFFOLI gate* **TOF** *is universal and reversible.*

2. The TOFFOLI gate can be constructed from **ID**, **AND**, **XOR** *and* **COPY**.

Example 5.5. The composition

$$\mathbf{TOF}(1,1,x) = (1,1,1\oplus x) \xrightarrow{\text{restrict}} 1 \oplus x = \mathbf{NOT}(x)$$

and so $\mathbf{NOT}(x) = r_3 \circ \mathbf{TOF}(1,1,x) = r_3 \circ \mathbf{TOF} \circ p_{(1,1)}(x)$, where $p_{(1,1)}(x) = (1,1,x)$ is the padding operation.

Hence, the classical gates **ID**, **AND**, **XOR** and **COPY** are sufficient to construct any logical circuit.

5.2 Single Qubit gates

A very rough description of a quantum circuit is as follows, we start with a n-Qubit $|\psi_{in}\rangle \in (\mathbb{C}^2)^{\otimes n}$ and perform some operations to obtain an output n-Qubit $|\psi_{out}\rangle \in (\mathbb{C}^2)^{\otimes n}$ from which we can perform a measurement or further operations. Hence, a quantum circuit should map pure states to pure states and so must be implemented by a unitary transformation $U \in \mathcal{L}\big((\mathbb{C}^2)^{\otimes n}\big)$ on n-Qubit space. Being unitary, it therefore follows that any quantum logical circuit is reversible (which is not true for classical gates).

Our task is therefore to understand unitary operators on $(\mathbb{C}^2)^{\otimes n}$. Of course, understanding arbitrary unitary operators does not necessarily help us implement such a transformation in the real world. So an important aspect of quantum circuits is to understand how to decompose unitary operators into simpler pieces. In particular, we will show that any unitary operator can be decomposed into a product of unitaries on the single Qubit space $\mathbb{C}^2$ and a few so-called controlled operations. With this in mind, we first review some basic but important unitary transformations on $\mathbb{C}^2$.

Definition 5.6. For any $k = 1, 2 \ldots$, we call a unitary operator $U \in \mathcal{L}\big((\mathbb{C}^2)^{\otimes k}\big)$ a quantum gate on k-Qubits.

Examples 5.7 (Single Qubit gates)**.** We review some common unitary operations on $\mathbb{C}^2$. We also their description using circuit diagrams. When an operator is written in terms of a matrix, this is always with respect to the canonical/computational basis $\{|0\rangle, |1\rangle\}$.

1. The Phase-shift gate, $P(\alpha)$ for $\alpha \in \mathbb{R}$,

$$P(\alpha) = |0\rangle\langle 0| + e^{i\alpha}|1\rangle\langle 1|, \qquad P(\alpha) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{pmatrix}, \qquad -\boxed{P(\alpha)}-$$

2. The Pauli matrices $\sigma_x = \sigma_1 = X$, $\sigma_y = \sigma_2 = Y$ and $\sigma_z = \sigma_3 = Z$,

$$X = |0\rangle\langle 1| + |1\rangle\langle 0|, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad -\boxed{X}- ,$$

$$Y = -i|0\rangle\langle 1| + i|1\rangle\langle 0|, \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad -\boxed{Y}- ,$$

$$Z = |0\rangle\langle 0| - |1\rangle\langle 1|, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad -\boxed{Z}- .$$

3. The Hadamard gate

$$H = \frac{1}{\sqrt{2}} \sum_{x,y \in \{0,1\}} (-1)^{xy} |x\rangle\langle y|, \qquad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad -\boxed{H}- .$$

4. Rotation around a unit vector $\hat{n} \in S_{\mathbb{R}^3} \subset \mathbb{R}^3$, the Bloch sphere,

$$D_{\hat{n}}(\alpha) = e^{-\frac{i\alpha}{2}(\hat{n}\cdot\sigma)} = \begin{pmatrix} \cos(\frac{\alpha}{2}) - i\sin(\frac{\alpha}{2})n_z & -i\sin(\frac{\alpha}{2})(n_x - in_y) \\ -i\sin(\frac{\alpha}{2})(n_x + in_y) & \cos(\frac{\alpha}{2}) + i\sin(\frac{\alpha}{2})n_z \end{pmatrix}, \qquad -\boxed{D_{\hat{n}}(\alpha)}-$$

5. It will often suffice to consider rotations around the x, y and z-axes, where we use the notation $R_x(\alpha) = D_{\hat{x}}(\alpha)$, $R_y(\alpha) = D_{\hat{y}}(\alpha)$, $R_z(\alpha) = D_{\hat{z}}(\alpha)$. Written as matrices,

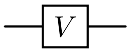
$$R_x(\alpha) = \begin{pmatrix} \cos(\frac{\alpha}{2}) & -i\sin(\frac{\alpha}{2}) \\ -i\sin(\frac{\alpha}{2}) & \cos(\frac{\alpha}{2}) \end{pmatrix}, \; R_y(\alpha) = \begin{pmatrix} \cos(\frac{\alpha}{2}) & -\sin(\frac{\alpha}{2}) \\ -\sin(\frac{\alpha}{2}) & \cos(\frac{\alpha}{2}) \end{pmatrix}, \; R_z(\alpha) = \begin{pmatrix} e^{-\frac{i\alpha}{2}} & 0 \\ 0 & e^{\frac{i\alpha}{2}} \end{pmatrix}.$$

   We also remark that $R_\bullet(\alpha) = e^{-\frac{i\alpha}{2}\sigma_\bullet}$ for $\bullet = x, y, z$.

6. The identity is also a unitary operation and it is marked on a circuit diagram by straight line (i.e. nothing is done to the Qubit),

$$\mathbf{1}|x\rangle = |x\rangle,$$.

7. An arbitrary unitary V on $\mathbb{C}^2$ is marked by a box on the circuit diagram,

$$V = \begin{pmatrix} v_{00} & v_{01} \\ v_{10} & v_{11} \end{pmatrix},$$ 

 .

Exercise 5.1. Show that for any $\alpha, \beta \in \mathbb{R}$,

(a) $P(\alpha) = e^{i\frac{\alpha}{2}} R_z(\alpha) R_y(0) R_z(0)$, (b) $X = e^{i\frac{\pi}{2}} R_z(\beta) R_y(\pi) R_z(\pi + \beta)$,

(c) $H = e^{-i\frac{\pi}{2}} R_z(0) R_y(\frac{\pi}{2}) R_z(-\pi)$.

Exercise 5.2. 1. Show that there exists $\xi, \alpha, \beta, \gamma \in \mathbb{R}$ such that any single Qubit gate U can be written as

$$U = e^{i\xi} R_z(\alpha) R_y(\beta) R_z(\gamma).$$

2. If U is single Qubit gate, show that there is a $\xi \in \mathbb{R}$ and unitary operators A, B, C constructed from R_y and R_z such that $ABC = \mathbf{1}$ and

$$U = e^{i\xi} AXBXC.$$

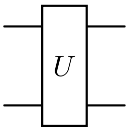
5.3 Controlled gates

Recall that the n-Qubit space has the canonical/computational basis via the binary representation of numbers $0, 1, \ldots, 2^n - 1$.

$$\big\{|x\rangle\big\}_{x=0}^{2^n-1} \sim \big\{|x_{n-1}\cdots x_1 x_0\rangle\big\}_{x_0,\ldots,x_{n-1}\in\{0,1\}} \text{ orthonormal basis of } \mathbb{C}^{2^n} \cong \big(\mathbb{C}^2\big)^{\otimes n},$$

where $x = x_{n-1}2^{n-1} + \cdots + x_1 2 + x_0$. To understand the action of a unitary U on an arbitrary n-Qubit $|\psi\rangle = \sum_x a_x |x\rangle$, it suffices to consider the action of U on this orthonormal basis.

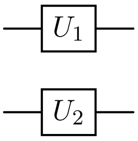
To begin, let's consider the space of 2-Qubits with basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. Given any unitary operator U, we can diagrammatically write its action on a 2-Qubit as



 ,

where the two lines going in/out represent the input/output Qubits. When we can factorise $U =$

$U_1 \otimes U_2$ with U_1 and U_2 unitary operators on single Qubit space, this gives a diagram



.

As much as possible, we would like to build more general quantum gates U via single Qubit gates. Of course, not every operator $T \in \mathcal{L}(\mathbb{C}^4)$ can be factorised as $A \otimes B$, but we can still obtain a larger class of operations via so-called controlled gates.

Example 5.8 (Controlled-NOT)**.** The single Qubit gate $X = \sigma_1$ acts as a quantum NOT gate as $X|x\rangle = |1 \oplus x\rangle$ for $x \in \{0, 1\}$ (recall that we use the convention $x \oplus y$ for addition modulo 2). We can extend this operation to a controlled-NOT gate on 2-Qubit space

$$C(X)|x_1x_0\rangle = C(X)(|x_1\rangle|x_0\rangle) = |x_1\rangle|x_1 \oplus x_0\rangle = |x_1\rangle\, X^{x_1}|x_0\rangle,$$

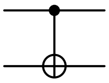
where we use the shorthand notation $|x_1\rangle|x_0\rangle = |x_1\rangle \otimes |x_0\rangle$. The operation of X on the second Qubit only occurs if the first Qubit is set to $|1\rangle$, so this action is controlled by the first Qubit. That is, the first Qubit is the control Qubit and the second Qubit is the target Qubit for the NOT operation. Explicitly,

$$C(X)|00\rangle = |00\rangle, \qquad C(X)|01\rangle = |01\rangle, \qquad C(X)|10\rangle = |11\rangle, \qquad C(X)|11\rangle = |10\rangle.$$

This operation can be written as a matrix (in the canonical/computational basis),

$$C(X) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

we also use the special diagrammatic notation



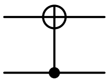
.

We could also consider a controlled not operation where the second Qubit is the control and the first Qubit is the target, $C_1(X)|x_1x_0\rangle = X^{x_0}|x_1\rangle|x_0\rangle$, i.e.,

$$C_1(X)|00\rangle = |00\rangle, \qquad C_1(X)|01\rangle = |11\rangle, \qquad C_1(X)|10\rangle = |10\rangle, \qquad C_1(X)|11\rangle = |01\rangle.$$

Written as a matrix and a diagram,

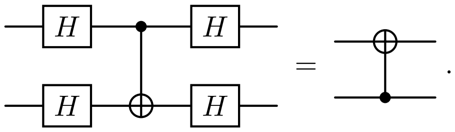
$$C_1(X) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \sim$$



.

With some care, we can interchange the role of the target and control Qubit when applying a controlled NOT gate.

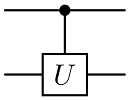
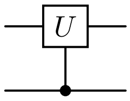
Exercise 5.3. Write the following equality of diagrams as an equality of operators and show it is true,



Example 5.9 (Controlled-U gate)**.** We can extend the idea of a controlled NOT gate to a controlled gate for any single Qubit gate U. Namely, we apply U to a target Qubit if the control Qubit is $|1\rangle$ and do nothing if the control Qubit is $|0\rangle$. Written mathematically,

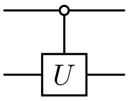
$$C(U)|x_1x_0\rangle = |x_1\rangle\, U^{x_1}|x_0\rangle, \qquad C_1(U)|x_1x_0\rangle = U^{x_0}|x_1\rangle\,|x_0\rangle.$$

Written as a circuit,

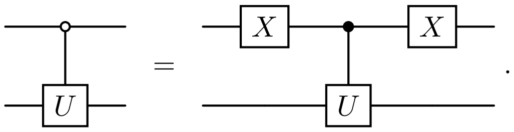
$C(U)\ \sim$  , $C_1(U)\ \sim$ 

Example 5.10 (0-Controlled-U gate)**.** The controlled gate $C(U)$ is such that we apply U to the second Qubit if the first Qubit is $|1\rangle$. This choice is rather arbitrary and we can instead perform operations if a Qubit is set to $|0\rangle$. Hence we define

$${}_0C^1(U)|x_1x_0\rangle = |x_1\rangle\, U^{1-x_1}|x_0\rangle, \qquad {}_0C^1(U)\ \sim$$

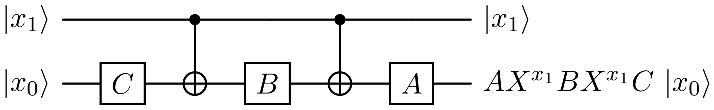


On a circuit diagram, we use a closed circle/dot to indicate a controlled operation subject to the Qubit being set to $|1\rangle$ and an open circle/dot to indicate a controlled operation with the Qubit set to $|0\rangle$. The choice of controlled-U vs 0-controlled-U is indeed quite minimal as we can always write



We would like to try and break down the controlled-U gate $C(U)$ into simpler steps. First, recall from part (2) of Exercise 5.2 that that there is some $\xi \in \mathbb{R}$ and unitary operators A, B, C such that $ABC = \mathbf{1}_2$ and $U = e^{i\xi}AXBXC$. We first consider $U' = AXBXC$ and the operation

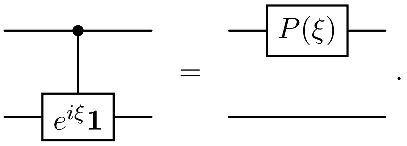
$$|x_1x_0\rangle \mapsto |x_1\rangle\, AX^{x_1}BX^{x_1}C|x_0\rangle,$$



where we note that a circuit diagram goes left-to-right. That is, we apply C on the second Qubit first. If $x_1 = 0$, then $AX^{x_1}BX^{x_1}C = ABC = \mathbf{1}$ and so we do nothing to the second Qubit. If $x_1 = 1$, then we apply $AXBXC = U'$ to the second Qubit as needed. Finally, we need to apply the phase $e^{i\xi}$ in a controlled way, $|x_1x_0\rangle \mapsto |x_1\rangle\, e^{ix_1\xi}|x_0\rangle$. A computation will show that this operation is the same as applying $\left(\begin{smallmatrix} 1 & 0 \\ 0 & e^{i\xi}\end{smallmatrix}\right) \otimes \mathbf{1}_2 = P(\xi) \otimes \mathbf{1}_2$. Indeed, for any $x \in \{0, 1\}$,

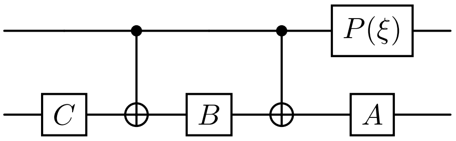
$$\begin{pmatrix} 1 & 0 \\ 0 & e^{i\xi}\end{pmatrix} \otimes \mathbf{1}_2|0x\rangle = |0x\rangle, \qquad \begin{pmatrix} 1 & 0 \\ 0 & e^{i\xi}\end{pmatrix} \otimes \mathbf{1}_2|1x\rangle = e^{i\xi}|1x\rangle.$$

So written as a circuit diagram,



Putting everything together, we can describe the controlled-U gate via the following operation

$C(U)\ \sim$ (5.1)



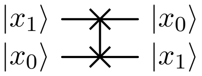
Why have we done this procedure? The unitaries A, B and C such that $ABC = \mathbf{1}$ and $U = e^{i\xi}AXBXC$ are constructed from rotations $R_y(\alpha)$ and $R_z(\beta)$. So we have shown that any controlled-U gate can be written as a combinations of controlled-NOT gates, R_y, R_z and the matrix $\left(\begin{smallmatrix}1 & 0\\ 0 & e^{i\xi}\end{smallmatrix}\right)$. More generally, our aim is to find a set of gates that are *universal* for n-Qubits: any unitary operator on n-Qubit space can be decomposed into a combination of these basic gates.

Example 5.11 (Swap gate)**.** We consider the unitary operation $S|x_1x_0\rangle = |x_0x_1\rangle$ that swaps the order of the two Qubits. In more detail

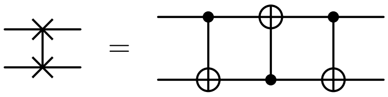
$$S|00\rangle = |00\rangle, \quad S|01\rangle = |10\rangle, \quad S|10\rangle = |01\rangle, \quad S|11\rangle = |11\rangle, \qquad S = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}.$$

As a diagram, we write

$S\ \sim$



Exercise 5.4. Show that S can be written as a combination of controlled-NOT gates,



Let us now generalise controlled gates to a system of n Qubits. As a first step, we need to specify how many Qubits are control Qubits and how many are target Qubits. Taking k control Qubits and l target Qubits with $k + l = n$, we can split $|x\rangle = |x_{n_1}\cdots x_{n-k}\rangle\,|x_{n-k-1}\cdots x_1x_0\rangle = |x_{(k)}\rangle|x_{(l)}\rangle$. Then for any l-Qubit gate U, we define

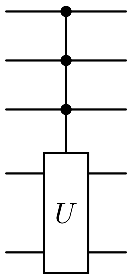
$$C^{k,l}(U)|x_{(k)}\rangle|x_{(l)}\rangle = |x_{(k)}\rangle\, U^{x_{(k)}}|x_{(l)}\rangle, \qquad U^{x_{(k)}} = U^{x_{n-1}\cdots x_{n-k}} = \begin{cases}U, & x_{n-1} = \ldots = x_{n-k} = 1,\\ \mathbf{1}, & \text{otherwise.}\end{cases}$$

Hence, we only apply U if *all* control Qubits are in the state $|1\rangle$. We can similarly define for a k-Qubit gate U,

$$C_{k,l}(U)|x_{(k)}\rangle|x_{(l)}\rangle = U^{x_{(l)}}|x_{(k)}\rangle\,|x_{(l)}\rangle, \qquad U^{x_{(l)}} = U^{x_0\cdots x_{l-1}}.$$

For example, when $k = 3$, $l = 2$, we write this as a diagram

$C^{3,2}(U) \sim$



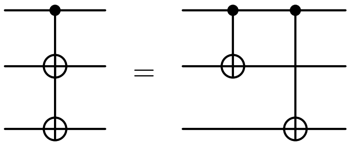
.

Note that a controlled operation involves all Qubits, but this includes cases where we wish to do nothing to a Qubit. For example, if U is an m-Qubit gate but where the last Qubit is unaffected (the general case can be done by rearranging Qubits by the isomorphism $\mathcal{H} \otimes \mathcal{K} \cong \mathcal{K} \otimes \mathcal{H}$). Then there is some other $(m-1)$-Qubit gate $\tilde{U}$ such that

$$U|x_{m-1} \cdots x_1 x_0\rangle = \tilde{U}|x_{m-1} \cdots x_2 x_1\rangle\,|x_0\rangle$$

and $U = \tilde{U} \otimes \mathbf{1}_2$.

Similarly, if $C^{k,l}(U)$ is a controlled gate and there is a factorisation $U = U_1 \otimes U_2$, then we can write $C^{k,l}(U) = C^{k,l}(U_1 \otimes \mathbf{1})C^{k,l}(\mathbf{1} \otimes U_2)$. As a simple example,



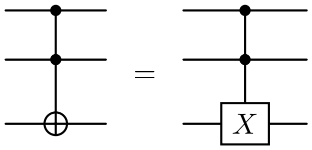
.

Example 5.12 (Quantum TOFFOLI gate). The TOF operation is also well-defined as a 3-Qubit gate,

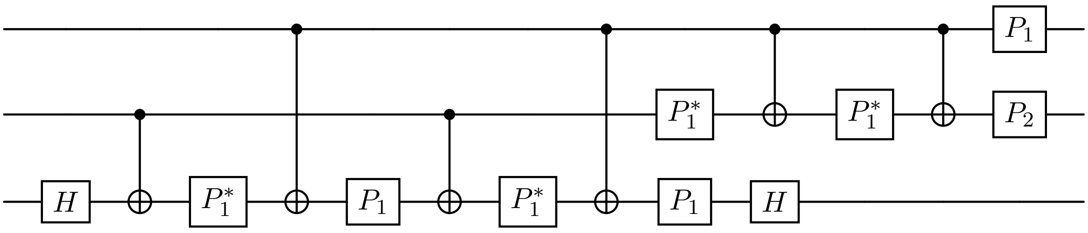
$$Q\text{-TOF} : |x_2 x_1 x_0\rangle = |x_2\rangle\,|x_1\rangle\,|x_2 x_1 \oplus x_0\rangle.$$

We can understand this gate as a controlled operation with two control Qubits. If x_2 or $x_1 = 0$, then nothing is done to the Qubit. If $x_2 = x_1 = 1$, then we perform a NOT operation on the third Qubit, $|x_0\rangle \mapsto |1 \oplus x_0\rangle$. Hence the quantum TOFFOLI gate can be considered as a controlled-NOT operation with 2 control Qubits.

Q-TOF =



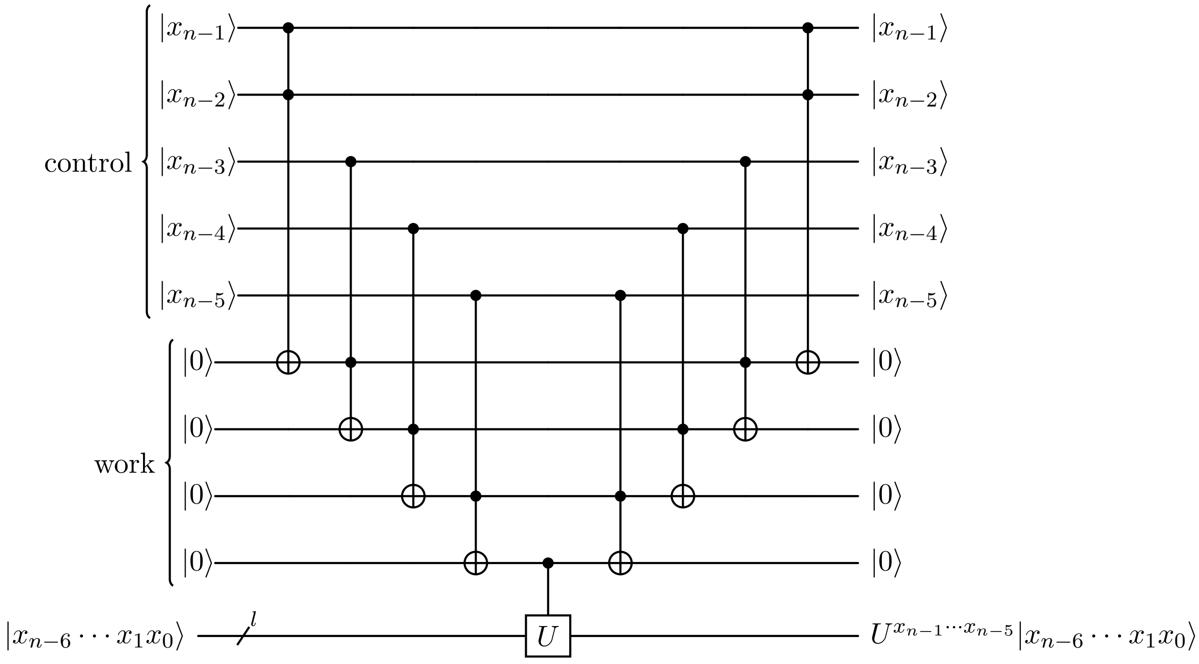
Exercise 5.5. Show that the quantum TOFFOLI gate can be decomposed as follows:



where $P_1 = P(\frac{\pi}{4})$ and $P_2 = P(\frac{\pi}{2})$.

Properties. By adding some additional 'work' Qubits, we can decompose $C^{k,l}(U)$ as a combination of Q-TOF gates and $C^{1,l}(U)$. We work on the $(k + (k-1) + l)$-Qubit space. Our aim is to use the additional $(k-1)$-Qubits to decompose a controlled gate of k-Qubits into steps.

We assume that the extra $(k-1)$ work Qubits are all set to $|0\rangle$ and fix the control Qubits $|x_{n-1}\cdots x_{n-k}\rangle$. Using a Q-TOF gate for $|x_{n-1}x_{n-2}\rangle$, we can transform the first work Qubit to $|x_{n-1}x_{n-2}\rangle$. We then apply Q-TOF to $|x_{n-3}\rangle$ and the first work Qubit $|x_{n-1}x_{n-2}\rangle$ to get $|x_{n-1}x_{n-2}x_{n-3}\rangle$ for the second work Qubit. Continuing this procedure, the $(k-1)$th work Qubit gets transformed to $|x_{n-1}\cdots x_{n-k}\rangle$. So we can apply $C^{1,l}(U)$ to the final work Qubit and target Qubits. Because Q-TOF is invertible (with itself as its inverse), we can then undo all the operations on the work Qubits. A diagram for $k = 5$ is shown below.



5.4 Principles of circuit diagrams

Our circuit diagram notation for quantum gates has largely been introduced by example. For completeness, let us therefore more explicitly state how we construct and use these diagrams.

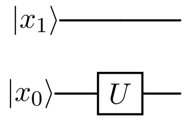
Properties. 1. **Unless otherwise stated, the input Qubits are elements of the canonical/computational basis of n-Qubit space**. There are exceptions such as the example of quantum teleportation, but these cases are explicitly labelled as such. Furthermore, we can always transform our input Qubits into other elements of other orthonormal bases by applying the appropriate transformation (e.g. applying the Hadamard gate to transform $\{|0\rangle, |1\rangle\}$ to $\{|+\rangle, |-\rangle\}$).

Hence, the input of a n-Qubit appearing in a circuit diagram is one of the elements $\{|x\rangle\}_{x=0}^{2^n-1} \sim \{|x_{n-1}\cdots x_1x_0\rangle\}_{x_j\in\{0,1\}}$.

2. **Each Qubit is represented by a straight line**. For multiple Qubit systems, these lines are stacked vertically.

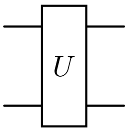
$$|x\rangle{=}|x_1x_0\rangle{=}\begin{cases} |x_1\rangle\rule[0.5ex]{3em}{0.4pt} \\ \\ |x_0\rangle\rule[0.5ex]{3em}{0.4pt} \end{cases} \sim\ |x_1\rangle \otimes |x_0\rangle.$$

Each individual Qubit may be labelled, but often will not be. Operations/gates that affect only one Qubit will only appear on the corresponding line,



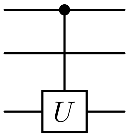
$$\sim (\mathbf{1}_2 \otimes U)|x\rangle = |x_1\rangle \otimes U|x_0\rangle.$$

Gates that act on multiple Qubits therefore appear over multiple lines.

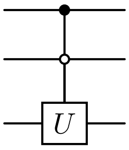


$$\sim \; U|x\rangle.$$

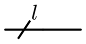
A controlled gate influences a particular Qubit only when there is an explicit control or target mark on the Qubit's line. For example, the operation



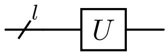
does not affect the middle Qubit, while



affects all three Qubits. For large systems, we also use the notation

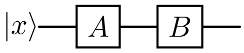


$$\sim \; |x_{l-1} \cdots x_1 x_0\rangle,$$



$$\sim \; U|x_{l-1} \cdots x_1 x_0\rangle.$$

3. **Diagrams are read left-to-right**. Some care is needed as this the opposite of how we write the action of linear operators,



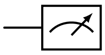
$$\sim \; BA|x\rangle.$$

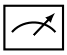
Measurement

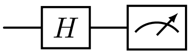
Given a n-Qubit system, we may wish to perform a measurement on one of the Qubits. To describe this, given any $j \in \{0, 1, \ldots, n-1\}$ and $\bullet \in \{x, y, z\}$, it is helpful to introduce the notation

$$\sigma_\bullet^{(j)} = \mathbf{1}_2^{\otimes(n-1-j)} \otimes \sigma_\bullet \otimes \mathbf{1}_2^{\otimes j}, \qquad \sigma_\bullet^{(j)}|x_{n-1} \cdots x_1 x_0\rangle = |x_{n-1} \cdots x_{n-j-1}\rangle \, \sigma_\bullet|x_j\rangle \, |x_{j-1} \cdots x_1 x_0\rangle.$$

That is, $\sigma_\bullet^{(j)}$ acts by $\sigma_\bullet$ on the jth Qubit and leaves the others alone. For a fixed $\bullet \in \{x, y, z\}$, we have that $\sigma_\bullet^{(j)}\sigma_\bullet^{(k)} = \sigma_\bullet^{(k)}\sigma_\bullet^{(j)}$ and so we can sharply and simultaneously measure any $\sigma_\bullet^{(j)}$ for any Qubit. We can therefore consider the observable $\sigma_z^{(j)} = P_{|0\rangle}^{(j)} - P_{|1\rangle}^{(j)}$, where a measurement of $\sigma_z^{(j)}$ will collapse the jth Qubit to $|0\rangle$ or $|1\rangle$ with a certain probability for each outcome. We write such a measurement in a circuit diagram as

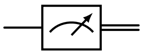


Like the input of a quantum circuit, unless otherwise stated, the symbol  corresponds to a measurement of $\sigma_z^{(j)}$. Measurements of other operators can be achieved by performing the relevant transformation prior to the final measurement. For example, because $H|x\rangle = |\pm\rangle$ for $x \in \{0, 1\}$,

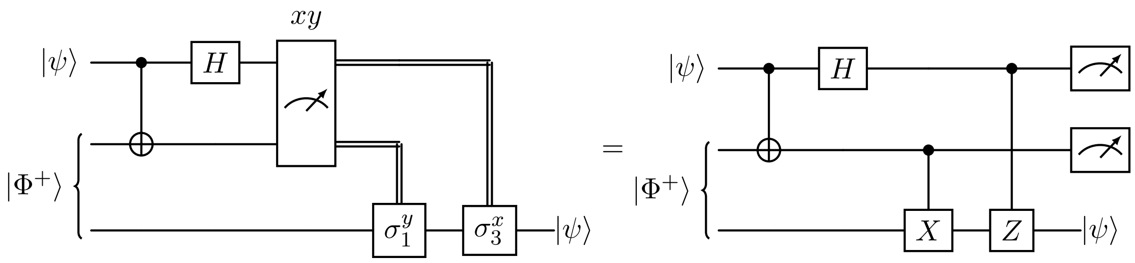


will produce the same result as a measurement of $\sigma_x^{(j)}$.

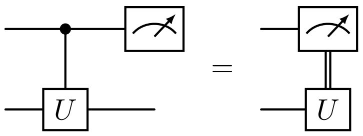
Generally speaking a measurement is a non-reversible operation that collapses a Qubit to one of at most two possibilities. Put another way, a measurement will turn quantum information, encoded by Qubits, into classical information, encoded by classical bits. Like the example of quantum teleportation, this classical information may still be used for quantum operations. In such cases, we use a double-line to denote a classical information channel.



Properties. 1. **Deferred measurement**. If a measurement is performed in the middle of a quantum circuit, this can be replaced with a controlled gate with measurement at the end of the circuit. For example, we can rephrase the circuit diagram for quantum teleportation as follows.



Writing the diagram without the classical information channel removes the physical interpretation of Alice communicating the result of her measurement to Bob, but mathematically the two diagrams represent the same procedure. Put more simply,



where the second diagram refers to the use of a measurement to apply/not apply a quantum gate.

2. **Implicit measurement**. At the end of a quantum circuit, we assume that any Qubits that have not been measured are then measured. Specifically, unless stated otherwise, we measure $\sigma_z^{(j)}$ at the end of all wires which have not been measured. Because all these operators commute, there is no uncertainty introduced in this operation. If the output Qubit is $|x_{l-1} \cdots x_1 x_0\rangle$, then this will output the classical l-bit $x_{l-1} \cdots x_1 x_0$.

5.5 Universal quantum gates

As previously stated, our aim is to try and understand any quantum circuit from a smaller set of concrete gates. We first make this notion precise.

Definition 5.13. Let $\{U_1, \dots, U_K\}$ be a set of quantum gates such that U_j is unitary on n_j-Qubit space. We denote by $\mathrm{Gate}(U_1, \dots, U_K)$ the set of quantum gates that are generated by $U_1, \dots, U_K$ in the following sense:

1. $U_1, \dots, U_K \in \mathrm{Gate}(U_1, \dots, U_K)$ and $\mathbf{1}^{\otimes n} \in \mathrm{Gate}(U_1, \dots, U_K)$ for all $n \in \mathbb{N}$.
2. If $V_1, V_2 \in \mathrm{Gate}(U_1, \dots, U_K)$, then V_1V_2, $V_1 \otimes V_2 \in \mathrm{Gate}(U_1, \dots, U_K)$.

We say that a set of quantum gates $\{U_1, \dots, U_K\}$ is *universal* if any unitary element $U \in \mathcal{L}\big((\mathbb{C}^2)^{\otimes n}\big)$ is such that $U \in \mathrm{Gate}(U_1, \dots, U_K)$.

Our task is therefore to find a simple but universal set of quantum gates. We have learned from Exercise 5.2 that for any single Qubit gate V,

$$V \in \mathrm{Gate}\big(\{e^{i\xi}\mathbf{1}_2, R_y(\alpha), R_z(\beta) \mid \xi, \alpha, \beta \in \mathbb{R}\}\big).$$

Let us therefore use the special notation,

$$\mathbf{M} = \{e^{i\xi}\mathbf{1}_2 \mid \xi \in \mathbb{R}\}, \quad \mathbf{R_y} = \{R_y(\alpha) \mid \alpha \in \mathbb{R}\}, \quad \mathbf{R_z} = \{R_z(\beta) \mid \beta \in \mathbb{R}\}.$$

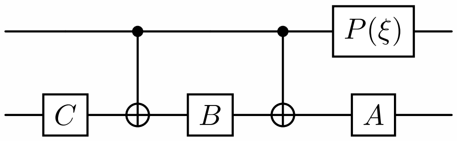
To extend this result to higher-order gates, we first note the following.

Lemma 5.14. *Let V be a quantum gate on 2-Qubits. Then for any $k \in \mathbb{N}$,*

$$C^{k,1}(V) \in \mathrm{Gate}\big(\mathbf{M}, \mathbf{R_y}, \mathbf{R_z}, C^1(X)\big).$$

Proof. Recall from Equation (5.1) that we can decompose $C(V) = C^1(V)$ as follows

$C(V)\ \sim$

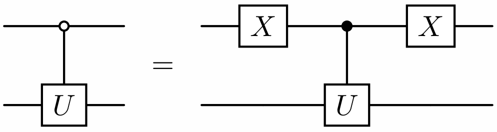


with $P(\xi), A, B, C \in \mathrm{Gate}\big(\mathbf{M}, \mathbf{R_y}, \mathbf{R_z}\big)$. This shows the result for $k = 1$. For the general case, we saw in the previous subsection that $C^{k,1}(V)$ can be decomposed into a combination of Q-TOF gates and $C^1(V)$. Furthermore, by Exercise 5.5, Q-TOF $\in \mathrm{Gate}\big(\mathbf{M}, \mathbf{R_y}, \mathbf{R_z}, C^1(X)\big)$ and so we are done. □

Corollary 5.15. *For any $k \in \mathbb{N}$ and quantum gate on 2-Qubits V,*

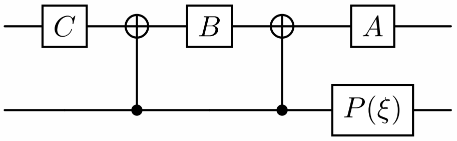
$$C_{k,1}(V),\ {}_0C^{k,1}(V) \in \mathrm{Gate}\big(\mathbf{M}, \mathbf{R_y}, \mathbf{R_z}, C^1(X)\big).$$

Proof. The identity



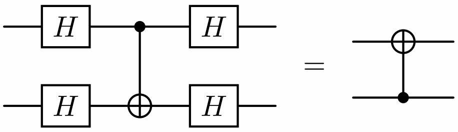
shows that ${}_0C^1(V) \in \text{Gate}\big(\mathbf{M}, \mathbf{R_y}, \mathbf{R_z}, C^1(X)\big)$. Analogous to the case of $C(V)$, we can write

$C_1(V) \sim$



,

which along with the identity



shows that $C_1(V) \in \text{Gate}\big(\mathbf{M}, \mathbf{R_y}, \mathbf{R_z}, C^1(X)\big)$. The higher-order case again follows from a combination of Q-TOF and controlled-V gates. □

We now consider the question of how to decompose general quantum gates on n-Qubits. To do this, we show that a unitary matrix can be decomposed into a product of so-called 2-level unitary matrices. Then, implementing 2-level unitary matrices can be decomposed as a combination of controlled gates and 2-Qubit gates. Both of these parts require several steps.

Decomposition into two-level gates

We consider U to be a generic $d \times d$ unitary matrix. Our aim is to decompose $U = U_1 \cdots U_k$ with $k = \mathcal{O}(d^2)$. and each U_j of the form below.

Definition 5.16. We say that a $d \times d$ matrix V is a two-level unitary matrix if it is unitary and acts non-trivially on at most two rows and columns.

The definition is best illustrated through examples.

Example 5.17.

$$V^{(1,3)} = \begin{pmatrix} v_{00} & 0 & v_{10} & 0 \\ 0 & 1 & 0 & 0 \\ v_{10} & 0 & v_{11} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} v_{00} & v_{01} \\ v_{10} & v_{11} \end{pmatrix} \text{ unitary.}$$

where we have two non-trivial entries in the *first* and *third* rows and columns.

$$U^{(2,5)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & u_{00} & 0 & 0 & u_{01} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & u_{10} & 0 & 0 & u_{11} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} u_{00} & u_{01} \\ u_{10} & u_{11} \end{pmatrix} \text{ unitary}$$

with non-trivial entries in the second and fifth rows and columns. Generally $V^{(p,q)}$ will have non-trivial entries in the p and qth rows and columns

Theorem 5.18. *Any $d \times d$ unitary matrix U with $d \geq 2$ can be factorised $U = U_1 \cdots U_k$ with each U_j a two-level unitary and $k = \mathcal{O}(d^2)$.*

We start with a lemma.

Lemma 5.19. *Let $v \in \mathbb{C}^d$ be a unit vector. There exist unitary two-level unitary matrics $U_1, \dots, U_m$ with $m \leq d$ such that*

$$U_m^* \cdots U_1^* v = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Proof. We prove the statement via induction on the number of non-zero entries in v. If v has one non-zero entry, say $v_j = e^{i\beta}$ (as $\|v\| = 1$), then take

$$U = \begin{pmatrix} 0 & \cdots & e^{-i\beta} & \cdots & 0 \\ \vdots & \ddots & & & \vdots \\ e^{i\beta} & & 0 & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}, \quad Uv = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

If v has ≥ 2 non-zero entries, say a and b, then we consider

$$\frac{1}{\alpha} \begin{pmatrix} \alpha & \cdots & & \cdots & & \cdots & 0 \\ \vdots & \ddots & & & & & \vdots \\ & & a^* & \cdots & b^* & & \\ & & \vdots & & \vdots & & \\ & & b & \cdots & a & & \\ \vdots & & & & & & \vdots \\ 0 & & & & & & \alpha \end{pmatrix} \begin{pmatrix} \vdots \\ \\ a \\ \vdots \\ b \\ \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \\ \alpha \\ \vdots \\ 0 \\ \\ \vdots \end{pmatrix}, \qquad \alpha = \sqrt{|a|^2 + |b|^2}.$$

The matrix on the left will act trivially on the other rows of v. So this operation reduces the amount of non-zero entries of v by one. We can continue this procedure up to $d-1$ times to obtain a v with one non-zero entry, which we can then transform to $(1, 0, \dots, 0)$. □

Proof (Proof of Theorem 5.18). We proceed via induction on d, where U is unitary $d \times d$ matrix. If $d = 2$, then there is nothing to do. If $d \geq 2$, we let v be the first column of U. Then by the previous lemma, there is a two-level unitary U_1 such that

$$U_1^* v = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad U_1^* U = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & U' & \\ 0 & & & \end{pmatrix},$$

where U' is a unitary $(d-1) \times (d-1)$ matrix. Applying the induction hypothesis, $U' = U_2 \cdots U_k$ with each U_j a two-level matrix. Therefore we have that

$$U_1^* U = U_2 \cdots U_k, \quad \Longrightarrow \quad U = U_1 \cdots U_k.$$ □

Example 5.20. The following unitary on 2-Qubit space

$$U = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$$

can be decomposed as

$$U = W_1^{(4,3)} W_2^{(4,2)} W_3^{(4,1)} W_4^{(3,2)} W_5^{(3,1)} W_6^{(2,1)},$$

where

$$W_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} -i & 1 \\ -1 & i \end{pmatrix}, \qquad W_2 = \frac{1}{\sqrt{3}}\begin{pmatrix} \sqrt{2} & -i \\ -i & \sqrt{2} \end{pmatrix}, \qquad W_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix},$$

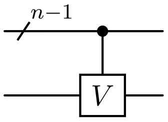
$$W_4 = \frac{1}{4}\begin{pmatrix} -(i+1)\sqrt{3} & 3-i \\ -(3+i) & (i-1)\sqrt{3} \end{pmatrix}, \qquad W_5 = \frac{1}{\sqrt{3}}\begin{pmatrix} \sqrt{2} & -1 \\ 1 & \sqrt{2} \end{pmatrix}, \qquad W_6 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}.$$

Implementing two-level unitaries

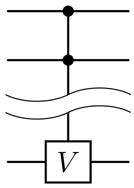
Thanks to Theorem 5.18, our task of decomposing a general quantum gate on n-Qubits has reduced to understanding how to implement two-level unitaries.

Exercise 5.6. Show that for $V = \begin{pmatrix} v_{00} & v_{01} \\ v_{10} & v_{11} \end{pmatrix}$ a unitary matrix,

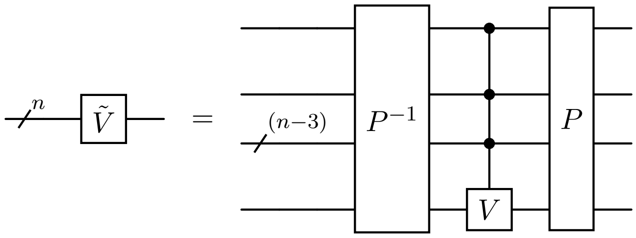
$$V^{(n-1,n)} = \begin{pmatrix} \mathbf{1}_{2^n-2} & & \\ & v_{00} & v_{01} \\ & v_{10} & v_{11} \end{pmatrix} = C^{n-1,1}(V) =$$



$=$



The exercise shows that certain two-level unitaries can be written as a controlled-V-gate, where V is the gate on 2-Qubits that enters the two-level unitary. Any two-level unitary can be put in this form via a conjugation of permutation matrices,



A permutation is any bijective map $\sigma : \{0, 1, \ldots, 2^n - 1\} \to \{0, 1, \ldots, 2^n - 1\}$, which then induces a unitary map $U_\sigma|x\rangle = |\sigma(x)\rangle$. Given a set of size 2^n, there are $2^n!$ possible permutations, an unfathomable number for n moderately large. To find the right permutation for a given two-level unitary, we use so-called gray codes. Given the binary representations $x = x_{n-1}\cdots x_1 x_0, y = y_{n-1}\cdots y_1 y_0 \in \{0, 1, \ldots, 2^n - 1\}$, a gray code is a sequence of binary numbers connecting x to y such that the any member of the sequence differs by the previous by exactly one bit.

Example 5.21. For $x = 011001$ and $y = 001110$, we can construct the gray code

$$\begin{aligned} x \;&= 0\,1\,1\,0\,0\,1 \\ &\; 0\,0\,1\,0\,0\,1 \\ &\; 0\,0\,1\,1\,0\,1 \\ &\; 0\,0\,1\,1\,1\,1 \\ &\; 0\,0\,1\,1\,1\,0 \;=\; y. \end{aligned}$$

For $x, y \in \{0, 1, \ldots, 2^n - 1\}$, then their binary representations differ by at most n (classical) bits. Hence the gray code is a sequence $\{g_m\}_{m=1}^{k}$ with $g_0 = x$, $g_k = y$ and $k \leq n + 1$. We then use this gray code to perform a permutation between the Qubits $|x\rangle$ and $|y\rangle$. We give a rough explanation of this procedure and skip some details. Suppose g_m and g_{m-1} differ in the ith slot. Then we can map $|g_m\rangle \mapsto |g_{m-1}\rangle$ by a controlled gate at the ith Qubit, which is conditional on all other Qubits being the same as those in g_{m-1} and g_m. We do this step by step, first swapping g_1 with g_2, g_2 with g_3 and so on until we reach g_{k-1},

$$|g_1\rangle \mapsto |g_{m-1}\rangle, \quad |g_2\rangle \mapsto |g_1\rangle, \; \ldots, \; |g_{k-1}\rangle \mapsto |g_{k-2}\rangle,$$

where other basis elements $|x\rangle$ are left unchanged if $x = x_{n-1} \cdots x_1 x_0 \notin \{g_m\}_{m=1}^{k}$. For the last step, we implement a controlled-V gate. If g_{k-1} and g_k differ on the jth Qubit, then we apply a controlled-V gate with the jth Qubit as the target, conditional on all other Qubits having the same values as those which appear in both g_k and g_{k-1}. Lastly, the permutations are undone and we have achieved an implementation of the two-level unitary. Importantly, the steps in the permuatation are done via a series of controlled gates, which we know can be handled with controlled-NOT and single Qubit gates.

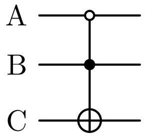
Example 5.22. Suppose we have a 3-Qubit system and wish to implement the two-level unitary,

$$\tilde{V} = \begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & v_{00} & & & & & v_{01} \\ & & & 1 & & & & \\ & & & & 1 & & & \\ & & & & & 1 & & \\ & & & & & & 1 & \\ & & v_{10} & & & & & v_{11} \end{pmatrix}, \qquad V = \begin{pmatrix} v_{00} & v_{01} \\ v_{10} & v_{11} \end{pmatrix},$$

which acts non-trivially on the basis elements $|2\rangle = |010\rangle$ and $|7\rangle = |111\rangle$. We construct a gray code, labeling each Qubit,

$$\begin{array}{ccc} A & B & C \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1. \end{array}$$

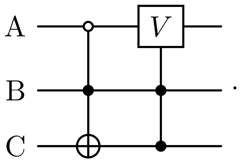
To implement $\tilde{V}$ as a controlled-V gate, we first swap C from 0 to 1 dependent on $A = 0$ and $B = 1$,



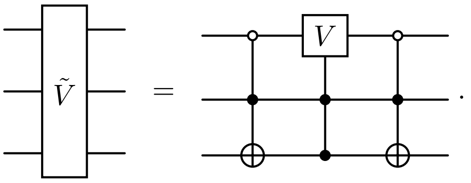
.

Because 011 and 111 differ at the first Qubit A, we then apply V to the Qubit A conditional on

$B = 1$ and $C = 1$,



Finally, we reverse the permutations to obtain an implementation of $\tilde{V}$,



The controlled-V and Q-TOF-like gates that appear in the implementation of $\tilde{V}$ can all be further decomposed in terms of single Qubit gates and controlled-NOT. Putting everything together, we have arrived at the following theorem.

Theorem 5.23. *The set of quantum gates*

$$\{\mathbf{M}, \mathbf{R_y}, \mathbf{R_z}, C^1(X)\} = \{e^{i\xi}\mathbf{1}_2, R_y(\alpha), R_z(\beta), C^1(X) \,:\, \xi, \alpha, \beta \in \mathbb{R}\}$$

are universal.

There are some severe limitations with the above theorem.

1. The set of universal gates is uncountably infinite. This is unavoidable as the set of unitary operators on $\mathbb{C}^d$ is also an uncountably infinite space, but is not so practical from the perspective of implementing a quantum circuit in a laboratory.
2. A decomposition of a generic unitary $U \in \mathcal{L}\big((\mathbb{C}^2)^{\otimes n}\big)$ takes several steps. Implementing a two-level unitary requires (at most) $2(n-1)$ controlled operations to permute the Qubits and then an application of controlled-V. Each controlled operation and controlled-V requires $\mathcal{O}(n)$ gates. So to implement $\tilde{V}$ requires $\mathcal{O}(n^2)$ gates. Similarly, decomposing a $2^n \times 2^n$ unitary matrix into two-level unitaries requires $\mathcal{O}((2^n)^2) = \mathcal{O}(4^2)$ two-level unitary operations. Putting this all together, a generic quantum gate on n qubits requires $\mathcal{O}(n^2 4^n)$ single-Qubit and controlled-NOT gates. As n increases, this becomes intractable very quickly. So the construction of a quantum algorithm is often quite different from this generic decomposition.

5.6 Approximating quantum gates by a finite generating set

The set $\{U \in \mathcal{L}(\mathbb{C}^m) \mid U \text{ unitary}\}$ is uncountably infinite even in the case $m = 1$. So any hope of *exactly* describing any unitary on n-Qubit space by a finite generating set is lost. Instead our aim is to find a finite set of operators $\{U_1, \ldots, U_r\}$ such that any unitary operator can be *approximated* by an element in $\text{Gate}(U_1, \ldots, U_r)$ up to arbitrary precision. We first make this notion precise.

Definition 5.24 (Approximating unitary operators)**.**

1. Let U and V be unitary operators on a Hilbert space $\mathcal{H}$. The error of implementing V instead of U is given by the quantity
$$E(U,V) = \sup_{\psi\in\mathcal{H},\, \|\psi\|=1} \big\|(U-V)|\psi\rangle\big\|.$$
2. We say that the unitary gates $\{U_1,\dots,U_r\}$, $U_j \in \mathcal{L}\big((\mathbb{C}^2)^{\otimes n_j}\big)$ are a finite universal set if for any unitary gate $U \in \mathcal{L}\big((\mathbb{C}^2)^{\otimes n}\big)$ and $\epsilon > 0$, there is some $V \in \text{Gate}\big(U_1,\dots,U_r\big)$ such that $E(U,V) < \epsilon$.

Exercise 5.7.

1. Let U, V be unitary operators on $\mathcal{H}$ and $|\psi\rangle$ a pure state. Show that for any orthogonal projection P on $\mathcal{H}$,
$$\big|\langle P\rangle_{U|\psi\rangle} - \langle P\rangle_{V|\psi\rangle}\big| \leq 2E(U,V).$$
2. Let $U_1,\dots,U_m, V_1,\dots,V_m$ be unitary operators on $\mathcal{H}$. Show that
$$E\big(U_mU_{m-1}\cdots U_1,\, V_mV_{m-1}\cdots V_1\big) \leq \sum_{j=1}^m E(U_j,V_j).$$

We saw in the previous subsection that the set of gates
$$\big\{\mathbf{M}, \mathbf{R_y}, \mathbf{R_z}, C^1(X)\big\} = \big\{e^{i\xi}\mathbf{1}_2, R_y(\alpha), R_z(\beta), C^1(X) \,:\, \xi,\alpha,\beta\in\mathbb{R}\big\}$$
is universal. In particular, this set is determined solely by single-Qubit gates and the controlled-NOT gate. So to find a finite universal set, it suffices to find a finite universal set for single-Qubit gates. We also recall Theorem 3.25, which says that for any $U \in \mathcal{L}(\mathbb{C}^2)$, there is some $\xi,\alpha\in\mathbb{R}$ and point $\hat{n}\in S_{\mathbb{R}^3}$, the Bloch sphere such that
$$U = e^{i\xi}D_{\hat{n}}(\alpha) = e^{i\xi}e^{-i\frac{\alpha}{2}(\hat{n}\cdot\sigma)}.$$
The first thing we will do is ignore the global phase $e^{i\xi}\mathbf{1}$ as we consider $|\psi\rangle$ and $e^{i\xi}|\psi\rangle$ as representing the same physical state. So our aim is to find a finite approximating set of unitaries for the general spin rotations $D_{\hat{n}}(\alpha)$.

Lemma 5.25. *Let $\hat{n}_1$ and $\hat{n}_2$ be fixed and non-parallel unit vectors in the Bloch sphere $S_{\mathbb{R}^3}$. Then for any $\hat{n}\in S_{\mathbb{R}^3}$ and $\alpha\in\mathbb{R}$, there is some $\beta,\gamma,\delta\in\mathbb{R}$ such that $D_{\hat{n}}(\alpha) = D_{\hat{n}_1}(\beta)D_{\hat{n}_2}(\gamma)D_{\hat{n}_1}(\delta)$.*

Proposition 5.26. *For any $\alpha\in\mathbb{R}$ and $\epsilon > 0$ there is some $V \in$ Gate$\big(P(\frac{\pi}{4}), H\big)$ such that $E(D_{\hat{n}}(\alpha), V) < \epsilon$.*

Proof (Proof sketch)*.* Up to a global phase rotation (which we ignore), we have that
$$\begin{aligned} P(\tfrac{\pi}{4})HP(\tfrac{\pi}{4})H &= \exp\big(-i\frac{\pi}{8}\sigma_z\big)\exp\big(-i\frac{\pi}{8}\sigma_x\big) \\ &= \cos^2\big(\frac{\pi}{8}\big)\mathbf{1}_2 - i\sin\big(\frac{\pi}{8}\big)\big(\cos\big(\frac{\pi}{8}\big)(\sigma_x+\sigma_z) + \sin\big(\frac{\pi}{8}\big)\sigma_y\big) \\ &= D_{\hat{n}_0}(\theta), \qquad \hat{n}_0 = \big(\cos(\tfrac{\pi}{8}),], \sin(\tfrac{\pi}{8}),\, \cos(\tfrac{\pi}{8})\big) \end{aligned}$$
and $\theta\in\mathbb{R}$ is such that $\cos(\frac{\theta}{2}) = \cos^2(\frac{\pi}{8})$. Such a θ is an irrational number. If we then consider the product $D_{\hat{n}_0}(\theta)^k = D_{\hat{n}_0}(k\theta)$, the irrationality of θ implies that sequence of numbers $\{k\theta \bmod 2\pi \mid k = 0,1,2,\dots\}$ will be dense in the interval $[0,2\pi)$. Hence $D_{\hat{n}_0}(\beta)$ can be approximated arbitrarily

well by $D_{\hat{n}_0}(\theta)^k$ for some $k \in \mathbb{N}$. In particular, for any $\epsilon > 0$, there is some $k \in \mathbb{N}$ such that $E\big(D_{\hat{n}_0}(\beta), D_{\hat{n}_0}(\theta)^k\big) < \epsilon/3$. We also note that for any $\alpha \in \mathbb{R}$.

$$HD_{\hat{n}_0}(\alpha)H = D_{\hat{n}_1}(\alpha), \qquad \hat{n}_1 = \big(\cos(\tfrac{\pi}{8}),\, -\sin(\tfrac{\pi}{8}),\, \cos(\tfrac{\pi}{8})\big).$$

We can similarly approximate $E\big(D_{\hat{n}_1}(\beta), D_{\hat{n}_1}(\theta)^k\big) < \epsilon/3$ for some $k \in \mathbb{N}$. Because $\hat{n}_0$ and $\hat{n}_1$ are not parallel, then we can apply Lemma 5.25 to say that for any $\hat{n} \in S_{\mathbb{R}^3}$ and $\alpha \in \mathbb{R}$,

$$D_{\hat{n}}(\alpha) = D_{\hat{n}_0}(\beta) D_{\hat{n}_1}(\gamma) D_{\hat{n}_0}(\delta),$$

where each term in the right-hand side can be approximated by a unitary in Gate $\big(P(\frac{\pi}{4}), H\big)$. and so there is some $k_1, k_2, k_3 \in \mathbb{N}$ such that

$$\begin{aligned} E\big(D_{\hat{n}}(\alpha),\, D_{\hat{n}_0}(\theta)^{k_1} D_{\hat{n}_1}(\theta)^{k_2} D_{\hat{n}_0}(\theta)^{k_3}\big) &= E\big(D_{\hat{n}_0}(\beta) D_{\hat{n}_1}(\gamma) D_{\hat{n}_0}(\delta),\, D_{\hat{n}_0}(\theta)^{k_1} D_{\hat{n}_1}(\theta)^{k_2} D_{\hat{n}_0}(\theta)^{k_3}\big) \\ &\le E\big(D_{\hat{n}_0}(\beta), D_{\hat{n}_0}(\theta)^{k_1}\big) + E\big(D_{\hat{n}_1}(\gamma), D_{\hat{n}_1}(\theta)^{k_2}\big) + E\big(D_{\hat{n}_0}(\delta), D_{\hat{n}_0}(\theta)^{k_3}\big) \\ &< \epsilon, \end{aligned}$$

where $D_{\hat{n}_0}(\theta)^{k_1} D_{\hat{n}_1}(\theta)^{k_2} D_{\hat{n}_0}(\theta)^{k_3} \in \text{Gate}\,\big(P(\frac{\pi}{4}), H\big)$. □

We therefore have the following.

Corollary 5.27. *The unitary gates $\big\{P(\frac{\pi}{4}), H, C^1(X)\big\}$ are a finite universal set.*

While it is certainly helpful that we can approximate single Qubit gates arbitrarily well by elements in Gate $\big(P(\frac{\pi}{4}), H\big)$, there is also an issue of scale. Suppose it takes $\mathcal{O}(2^{1/\epsilon})$ gates from $\big\{P(\frac{\pi}{4}), H\big\}$ to approximate $U \in \mathcal{L}(\mathbb{C}^2)$ up to an error of ϵ. Then if we want to approximate a unitary gate on m Qubits, we will require $\mathcal{O}(m2^{m/\epsilon})$ gates built from $\big\{P(\frac{\pi}{4}), H, C^1(X)\big\}$. This is far from ideal as the number of gates required scales exponentially as we increase the number of Qubits. So building a quantum algorithm on a system of, say, 10 Qubits would be very challenging.

Thankfully, the approximation of single gate unitaries by by elements in Gate $\big(P(\frac{\pi}{4}), H\big)$ is much more efficient. This improvement follows from the Solovay–Kitaev Theorem, which more generally concerns finite approximations of elements in the group $SU(d)$.

Theorem 5.28 (Solovay–Kitaev Theorem (special case), [3, Appendix 3])**.** *A single Qubit gate can be approximated up to an error ϵ using $\mathcal{O}\big(\log^c(1/\epsilon)\big)$ gates constructed from $\{P(\frac{\pi}{4}), H\}$, where $c \approx 2$ is a constant.*

The Solovay–Kitaev Theorem implies that a gate on m Qubits can be approximated up to error ϵ by $\mathcal{O}\big(m \log^c(m/\epsilon)\big)$ gates from $\{P(\frac{\pi}{4}), H, C^1(X)\}$, a substantial improvement over the exponential increase.

6 Quantum algorithms

In the previous section, we studied quantum gates and their decomposition into simpler parts. Equipped with this knowledge, we can now use these quantum gates to solve problems. Very loosely speaking, our aim is to find quantum circuits that transform an input pure state $|\psi_{\text{in}}\rangle \in (\mathbb{C}^2)^{\otimes n}$ to an output state $|\psi_{\text{out}}\rangle(\mathbb{C}^2)^{\otimes n}$, where a measurement of $|\psi_{\text{out}}\rangle$ (or the state itself) is telling us useful information.

6.1 Binary addition

The first operation we define is a basic addition of basis vectors. Given numbers $x, y \in \{0, 1, \ldots, 2^n - 1\}$, we define a binary addition by taking a mod 2 addition in every component of the binary decomposition

$$x \oplus y = x_{n-1} \cdots x_1 x_0 \oplus y_{n-1} \cdots y_1 y_0 = (x_{n-1} \oplus y_{n-1}) \cdots (x_1 \oplus y_1)(x_0 \oplus y_0).$$

That is, $x_j \oplus y_j \in \{0, 1\}$ is the usual binary addition.

Example 6.1. Suppose $n = 5$ and we consider the binary addition of

$$20 \oplus 25 = 10100 \oplus 11001 = 01101 = 13.$$

Note that $13 = (20 + 25) \bmod 32$. But generally binary addition will not implement addition modulo 2^n. Indeed $4 \oplus 4 = 00100 \oplus 00100 = 00000 = 0 \neq 8 \bmod 32$.

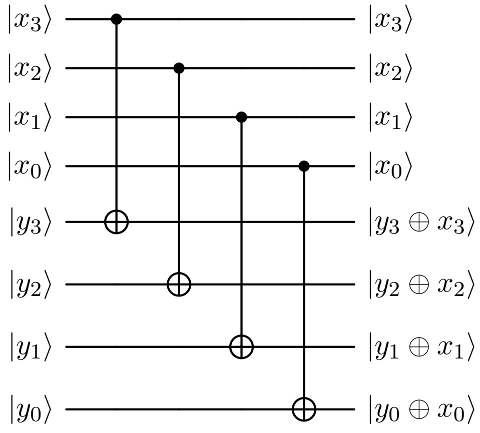
Definition 6.2. We define the map on the canonical basis,

$$(\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes n} \to (\mathbb{C}^2)^{\otimes n}, \qquad |x\rangle \otimes |y\rangle \mapsto |x \oplus y\rangle = \bigotimes_j |x_j \oplus y_j\rangle.$$

We also define the operator

$$U_\oplus : (\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes n} \to (\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes n}, \qquad U_\oplus(|x\rangle \otimes |y\rangle) = |x\rangle \otimes |y \oplus x\rangle.$$

The mod 2 addition is quite naturally defined. Furthermore, the linear operator $U_\oplus$ can be easily implemented by a composition of controlled-NOT gates, which also shows it is unitary. The circuit diagram for $n = 4$ is shown below.



A more challenging task is to construct a quantum circuit for addition modulo 2^n, $|x\rangle \otimes |y\rangle \mapsto |x\rangle \otimes |(x + y) \bmod 2^n\rangle$. We leave this as an exercise for the interested reader.

6.2 Logic gates and quantum parallelism

In what follows, we will freely pass between the following presentations

- The natural number $0 \leq x \leq 2^n - 1$ for some $n \in \mathbb{N}$,
- The binary decomposition $x = \sum_{j=0}^{n-1} x_j 2^j \sim x_{n-1} \cdots x_1 x_0$,
- An element in the cartesian product $\{0,1\}^n$, $x \sim (x_{n-1}, \ldots, x_1, x_0) \in \{0,1\}^n$.

Definition 6.3. Let $f : \{0,1\}^n \to \{0,1\}^m$ be a function with $n \geq m$. We define the operator U_f on the canonical basis such that

$$U_f : (\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes m} \to (\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes m}, \qquad U_f |x\rangle \otimes |y\rangle = |x\rangle \otimes |y \oplus f(x)\rangle.$$

We emphasise that the function $f : \{0,1\}^n \to \{0,1\}^m$ need not be a bijection. Despite this generality for f, we still have the following.

Lemma 6.4. *The operator U_f is unitary.*

Proof. The operator is defined by its action on an othonormal basis and so is linear. Because our Hilbert space is finite-dimensional, it suffices to show that $\langle U_f \Psi \mid U_f \Phi \rangle = \langle \Psi \mid \Phi \rangle$ for any $\Psi, \Phi \in (\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes m}$. We take a decompositon $\Psi = \sum_{j,k} \Psi_{jk} |j \otimes k\rangle$ and $\Phi = \sum_{j'k'} \Phi_{j'k'} |j' \otimes k'\rangle$ and compute

$$\begin{aligned}
\langle U_f \Psi \mid U_f \Phi \rangle &= \sum_{j,j'} \sum_{k,k'} \overline{\Psi_{jk}} \Phi_{j'k'} \langle U_f (j \otimes k) \mid U_f (j' \otimes k') \rangle \\
&= \sum_{j,j'} \sum_{k,k'} \overline{\Psi_{jk}} \Phi_{j'k'} \langle j \otimes k \oplus f(j) \mid j' \otimes k' \oplus f(j') \rangle \\
&= \sum_{j} \sum_{k,k'} \overline{\Psi_{jk}} \Phi_{jk'} \langle k \oplus f(j) \mid k' \oplus f(j) \rangle_{(\mathbb{C}^2)^{\otimes m}} \\
&= \sum_{j} \sum_{k,k'} \overline{\Psi_{jk}} \Phi_{jk'} \prod_{l=0}^{m-1} \langle k_l \oplus f(j)_l \mid k'_l \oplus f(j)_l \rangle_{\mathbb{C}^2} \\
&= \sum_{j} \sum_{k,k'} \overline{\Psi_{jk}} \Phi_{jk'} \prod_{l=0}^{m-1} \delta_{k_l, k'_l} \\
&= \sum_{j} \sum_{k} \overline{\Psi_{jk}} \Phi_{jk} = \langle \Psi \mid \Phi \rangle.
\end{aligned}$$

□

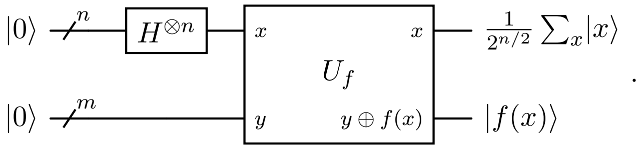
Let us now consider the question of evaluating a given function $f : \{0,1\}^n \to \{0,1\}^m$ via a quantum circuit. To do this, we use the Hadamard operator, where

$$H^{\otimes n} = \frac{1}{2^{n/2}} \sum_{x,y=0}^{2^n-1} (-1)^{x \cdot y} |x\rangle\langle y|, \quad \Longrightarrow \quad H^{\otimes n} |0\rangle^{\otimes n} = \sum_{x=0}^{2^n-1} |x\rangle.$$

We can therefore apply U_f to the sum of states $\sum_x |x\rangle \otimes |y\rangle$ to obtain $\sum_x |x\rangle \otimes |y \oplus f(x)\rangle$. In particular, we find that

$$U_f (H^{\otimes n} \otimes \mathbf{1}_{2^m}) |0\rangle^{\otimes n} \otimes |0\rangle^{\otimes m} = \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} U_f (|x\rangle \otimes |0\rangle) = \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} |x\rangle \otimes |f(x)\rangle.$$

Or written pictorially,



Therefore with an application of this quantum circuit, we obtain an output state which includes the information of *all* values of the classical logical gate $f : \{0,1\}^n \to \{0,1\}^m$. This simultaneous evaulation of a function is called quantum parallelism and is not something that can be easily reproduced on a classical computer. However, we can only learn information about the output state by taking a measurement. If we choose to measure the projections $P_x = |x\rangle\langle x|$ for all x, this output state will collapse into a single value $|x'\rangle \otimes |f(x')\rangle$ for some $x' \in \{0, 1, \ldots, 2^n - 1\}$. So we learn the value of f at a single value, certainly something that a classical computer is also capable of doing! Instead, the challenge of constructing quantum algorithms comes from how to use properties like quantum parallelism with the limitation that any measurement of an output state will collapse the Qubit into something much simpler.

6.3 The Deutsch–Jozsa algorithm revisited

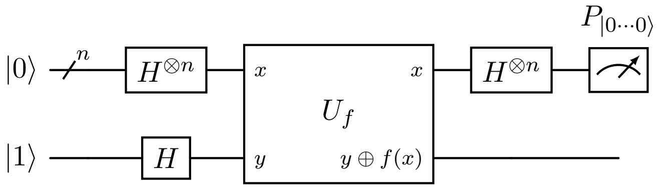
Let us return to the example used at the very start of these notes, the Deutsch–Jozsa algorithm. Though we now consider a slightly more complicated setting. Given $I \subset \mathbb{R}$ a function $f : I \to \mathbb{R}$, we define the *support* of f, $\operatorname{supp}(f) = \{x \in I : f(x) \neq 0\}$. We can use this notion to define a balanced function $f : \{0,1\}^n \to \{0,1\}$.

Definition 6.5. We say that a logic gate $f : \{0,1\}^n \to \{0,1\}$ is balanced if $|\operatorname{supp}(f)| = \frac{1}{2}2^n = 2^{n-1}$. That is, $\operatorname{supp}(f)$ is half the size of the domain.

If a function is balanced, then $f(x) = 1$ for half of all $x \in \{0,1\}^n$. Similarly $f(x') = 0$ for the remaining half of $x' \in \{0,1\}^n$.

Suppose we are given a function $f : \{0,1\}^n \to \{0,1\}$ that we know to be *either* constant, $f(x) = c \in \{0,1\}$ for all $x \in \{0,1\}^n$, or balanced. We would like to determine the nature of f (constant or balanced) in as few steps as possible. If we are given a classical computer and a constant function, we will need to evaluate $f(x)$ for $2^{n-1}+1$ different values of x to know with absolute certainty that it is constant. With a quantum computer, we can answer the question via a single quantum circuit, the Deutsch–Jozsa algorithm.

While we will explain the details of the procedure, we first provide the circuit diagram and encourage the reader to work out the details for themselves. Like the previous subsection, we will freely pass between a presentation of numbers $\{0, 1, \ldots, 2^n - 1\}$ as elements of the set $\{0,1\}^n$ and vice verca.



The Deutsch–Jozsa algorithm uses a slightly modified version of quantum parallelism. We take $\mathcal{H} = (\mathbb{C}^2)^{\otimes(n+1)}$ with initial state $|0\rangle^{\otimes n} \otimes |1\rangle$. Applying $H^{\otimes(n+1)}$ will transform this state to

$$\frac{1}{\sqrt{2^{n+1}}} \sum_{x=0}^{2^n-1} |x\rangle \otimes (|0\rangle - |1\rangle).$$

For a fixed $x \in \{0, 1, \dots, 2^n - 1\}$, we note that

$$U_f\big(|x\rangle \otimes (|0\rangle - |1\rangle)\big) = |x\rangle \otimes (|f(x)\rangle - |1 \oplus f(x)\rangle) = \begin{cases} |x\rangle \otimes (|0\rangle - |1\rangle), & f(x) = 0, \\ -|x\rangle \otimes (|0\rangle - |1\rangle), & f(x) = 1 \end{cases}$$
$$= (-1)^{f(x)}|x\rangle \otimes (|0\rangle - |1\rangle))$$

and therefore taking the sum over x,

$$U_f H^{\otimes(n+1)}\big(|0\rangle^{\otimes n} \otimes |1\rangle\big) = \frac{1}{\sqrt{2^{n+1}}} \sum_{x=0}^{2^n-1} (-1)^{f(x)}|x\rangle \otimes (|0\rangle - |1\rangle).$$

We now apply the gate $H^{\otimes n} \otimes \mathbf{1}_2$, where by the formulas

$$H^{\otimes n} = \frac{1}{2^{n/2}} \sum_{y,z=0}^{2^n-1} (-1)^{y \cdot z} |y\rangle\langle z|, \qquad H^{\otimes n}|x\rangle = \frac{1}{2^{n/2}} \sum_{y=0}^{2^n-1} (-1)^{y \cdot x}|y\rangle,$$

one finds that

$$(H^{\otimes n} \otimes \mathbf{1}_2) U_f H^{\otimes(n+1)}\big(|0\rangle^{\otimes n} \otimes |1\rangle\big) = \frac{1}{2^n} \sum_{x,y=0}^{2^n-1} (-1)^{f(x)}(-1)^{x \cdot y}|y\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

We now take a measurement of the first n-Qubits. In particular, we measure the projection onto the state $|0\rangle^{\otimes n}$. The probability the first n-Qubits are in this state is determined by the (modulus squared) of the coefficient of $|0 \cdots 0\rangle$. We find that

$$c_{0 \cdots 0} = \sum_{x=0}^{2^n-1} \frac{(-1)^{f(x)}}{2^n} = \begin{cases} \pm 1, & f \text{ constant}, \\ 0, & f \text{ balanced}, \end{cases}$$

Thus, by taking a measurement of the first n-Qubits, we are able to distinguish the cases of a balanced and constant function.

6.4 Quantum search algorithm

The quantum search algorithm, also called the Grover search algorithm, provides a method to find a specific object (a needle) in a large unordered set (a haystack) of size N. Our aim is to find a method such that the probability we have found the object/solution x^* after k steps is greater than $\frac{1}{2}$. Using classical methods, this can be done in $k = \mathcal{O}\big(\frac{N}{2}\big)$ steps. Using Grover's algorithm the probability to find x^* is greater than 50% after $k = \mathcal{O}(\sqrt{N})$ steps. For large systems, e.g. $N \sim 10^{10}$, this difference is substantial.

To formalise the problem, we take $N = 2^n$ for some n and consider the set $\{0, 1, \dots, 2^n - 1\}$.[2] Within this set, the objects we are looking for are given by a solution subset $S \subset \{0, 1, \dots, 2^n - 1\}$ of size $m < N/2$. (If $m \geq N/2$, then picking a random element will give $x \in S$ with probability $\geq 50\%$.) We can therefore define the orthogonal projections on n-Qubit space,

$$P_S = \sum_{x \in S} |x\rangle\langle x|, \qquad \mathbf{1} - P_S = P_{S^\perp} = \sum_{x \notin S} |x\rangle\langle x|.$$

We can similarly define a function that will indicate if we have found a solution

$$g : \{0, 1, \dots, 2^n - 1\} \to \{0, 1\}, \qquad g(x) = \begin{cases} 1, & x \in S, \\ 0, & x \notin S. \end{cases}$$

We can therefore implement the function via the unitary U_g on $(\mathbb{C}^2)^{\otimes n} \otimes \mathbb{C}^2$,

$$U_g|x\rangle \otimes |y\rangle = |x\rangle \otimes |y \oplus g(x)\rangle.$$

[2] If $N \neq 2^n$, we can always add extra elements $N+1, N+2, \dots$ until we reach $2^{n'}$ for some n'.

Exercise 6.1. Show that for any pure state $|\psi\rangle \in (\mathbb{C}^2)^{\otimes n}$,

$$U_g(|\psi\rangle \otimes |-\rangle) = (\mathbf{1} - 2P_S)|\psi\rangle \otimes |-\rangle, \qquad P_S = \sum_{x \in S} |x\rangle\langle x|,$$

where $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$

We can now describe the procedure to implement the quantum search. We set $N = 2^n$ take the following initial state

$$|\Psi_0\rangle = |\psi_0\rangle \otimes |-\rangle, \qquad |\psi_0\rangle = \frac{1}{2^{n/2}} \sum_{x=0}^{N} |x\rangle = H^{\otimes n}|0\rangle^{\otimes n}.$$

Exercise 6.2. Let $\theta \in [0, \frac{\pi}{2}]$ be such that $\sin(\theta) = \sqrt{\frac{m}{N}}$. Show that

$$|\psi_0\rangle = \cos(\theta)|\psi_{S^\perp}\rangle + \sin(\theta)|\psi_S\rangle,$$

where

$$|\psi_S\rangle = \frac{1}{\sqrt{m}} \sum_{x \in S} |x\rangle, \qquad |\psi_{S^\perp}\rangle = \frac{1}{\sqrt{N-m}} \sum_{x \notin S} |x\rangle.$$

Definition 6.6. Given the function $g : \{0, 1, \dots, N\}$ and initial state $|\Psi_0\rangle = |\psi_0\rangle \otimes |-\rangle$, the Grover iteration is the operator

$$G \in \mathcal{L}\big((\mathbb{C}^2)^{\otimes n} \otimes \mathbb{C}^2\big), \qquad G = \big((2|\psi_0\rangle\langle\psi_0| - \mathbf{1}) \otimes \mathbf{1}_2\big)U_g.$$

It is straightforward to verify that $(2|\psi_0\rangle\langle\psi_0| - \mathbf{1})$ and therefore G is a unitary operator.

Lemma 6.7. *For $j \in \mathbb{N}$, $G^j|\Psi_0\rangle = |\psi_j\rangle \otimes |-\rangle$, where*

$$|\psi_j\rangle = \cos((2j+1)\theta)|\psi_{S^\perp}\rangle + \sin((2j+1)\theta)|\psi_S\rangle.$$

Proof (Proof sketch). The proof is via induction with the case $j = 0$ proved in Exercise 6.2. We let $\theta_j = (2j+1)\theta$ and consider the inductive step, where

$$\begin{aligned} G^{j+1}|\Psi_0\rangle &= \big((2|\psi_0\rangle\langle\psi_0| - \mathbf{1}) \otimes \mathbf{1}_2\big)U_g\big((\cos(\theta_j)|\psi_{S^\perp}\rangle + \sin(\theta_j)|\psi_S\rangle) \otimes |-\rangle\big) \\ &= \big(2|\psi_0\rangle\langle\psi_0| - \mathbf{1}\big)\big(\cos(\theta_j)|\psi_{S^\perp}\rangle + \sin(\theta_j)|\psi_S\rangle\big) \otimes |-\rangle \\ &= \Big(\cos(\theta_j)\big(2|\psi_0\rangle\langle\psi_0 \mid \psi_{S^\perp}\rangle - |\psi_{S^\perp}\rangle\big) \\ &\qquad - \sin(\theta_j)\big(2|\psi_0\rangle\langle\psi_0 \mid \psi_S\rangle - |\psi_S\rangle\big)\Big) \otimes |-\rangle \\ &= |\psi_{j+1}\rangle \otimes |-\rangle. \end{aligned}$$

The identities $\langle\psi_0 \mid \psi_{S^\perp}\rangle = \cos(\theta)$ and $\langle\psi_0 \mid \psi_S\rangle = \sin(\theta)$ and some trigonometric formulas for sin and cos can be used to simplify

$$\begin{aligned} |\psi_{j+1}\rangle &= \big(\cos(\theta_j)\cos(2\theta) - \sin(\theta_j)\sin(2\theta)\big)|\psi_{S^\perp}\rangle + \big(\cos(\theta_j)\sin(2\theta) + \sin(\theta_j)\cos(2\theta)\big)|\psi_S\rangle \\ &= \cos(\theta_j + 2\theta)|\psi_{S^\perp}\rangle + \sin(\theta_j + 2\theta)|\psi_S\rangle \\ &= \cos(\theta_{j+1})|\psi_{S^\perp}\rangle + \sin(\theta_{j+1})|\psi_S\rangle \end{aligned}$$

as required. □

We can think of the Grover iteration $G^j|\Psi_0\rangle = \big(\cos((2j+1)\theta)|\psi_{S^\perp}\rangle + \sin((2j+1)\theta)|\psi_S\rangle\big) \otimes |-\rangle$ as rotating the initial state $|\psi_0\rangle$ between the two orthogonal subspaces $P_{S^\perp}\mathcal{H}$ and $P_S\mathcal{H}$. In particular, we can find a value j that maximises $\sin((2j+1)\theta)$, where

$$\mathbf{P}\big(\text{Measurement of } P_S \text{ with } |\psi_j\rangle \text{ returns value } 1\big) = \big\|P_S|\psi_j\rangle\big\|^2 = \sin^2((2j+1)\theta).$$

If a measurement of P_S in the state $G^j|\Psi_0\rangle$ returns 1, then $|\psi_j\rangle$ will collapse to the state $|x\rangle$ for some $x \in S$. So we have found the desired object. Note that $\sin^2(\theta) = \frac{m}{N}$ and by repeated actions of the Grover operator G, we can increase $\sin^2((2j+1)\theta)$ and take $|\psi_j\rangle$ such that the probability of a successful search is high.

Lemma 6.8 ([4, Lemma 6.29]). *For $a \in \mathbb{R}$, let $\lfloor a \rfloor$ denote the smallest integer $\leq a$. Then for*

$$j_N = \left\lfloor \frac{\pi}{4\sin^{-1}\left(\sqrt{\frac{m}{N}}\right)} \right\rfloor,$$

we have that

$$\mathbf{P}\big(\textit{Measurement of } P_S \textit{ with } |\psi_{j_N}\rangle \textit{ returns value } 1\big) \geq 1 - \frac{m}{N}.$$

We also note that $j_N = \mathcal{O}\big(\sqrt{\frac{N}{m}}\big)$ for $N \to \infty$.

Our precise value for j_N required both a knowledge of N and m the size of the solution set S. In the case the value m is unknown (but is still a non-zero number), the algorithm can be modified so that for $0 \leq j \leq \lfloor\sqrt{N}\rfloor + 1$,

$$\mathbf{P}\big(\text{Measurement of } P_S \text{ with } |\psi_j\rangle \text{ returns value } 1\big) \geq \frac{1}{4},$$

see [4, Theorem 6.32].

6.5 Quantum Fourier transform

Exercise 6.3. Write a report on the quantum Fourier transform and some of its applications. (See [3, Chapter 5] or [4, Chapter 5] for example.)

6.6 Quantum cryptography

Exercise 6.4. Write a report on the BB84 (or any other) model of quantum cryptography. (See [2, Section 7.5] or [4, Chapter 6] for example.)

7 Elements of quantum information theory

In the last section of these notes, we briefly look at some of the fundamental objects that appear in quantum information theory. Quantum information does not have a clean and unambiguous definition. It may refer to *any* study of the transmission of data or information via a quantum mechanical process. The study of quantum gates, quantum computation and quantum algorithms would therefore be considered a special case of quantum information. A more specific characterisation of quantum information is often in direct comparison with *classical* information theory. Namely what are the fundamental objects of information that can be transmitted, what is the process that sends quantum information, what physical or theoretical limits are there on such transmissions, what are the effects of noise/errors, etc.

We will not be so concerned with providing a precise definition of quantum information and will instead focus on introducing some common objects that appear in the study of quantum information. The content in this section represents a tiny segment of this much larger theory. Many more details can be found in [3, Part III] for example.

7.1 Quantum operations and channels

As we have studied, the basic element of a quantum system described by a Hilbert space $\mathcal{H}_A$ is a state $\rho_A \in \mathrm{Dens}(\mathcal{H}_A)$. We can change this state by taking a time evolution, $\rho_A \mapsto U\rho_A U^*$, or by taking a measurement, $\rho_A \mapsto \dfrac{P_\lambda \rho_A P_\lambda}{\mathrm{Tr}(\rho_A P_\lambda)}$ for some eigenspace projection P_λ of an observable $A = A^* \in \mathcal{L}(\mathcal{H}_A)$.

We now consider these processes but where our system $\mathcal{H}_A$ and state is in interaction with some environment or external system $\mathcal{H}_B$. Such an space $\mathcal{H}_B$ may be used to encode:

1. The interaction of our system with an external environment that may change the state. For example, the space $\mathcal{H}_B$ can be used to give a mathematical description of a noisy system which may introduce errors or change the state ρ_A.
2. The introduction of an extra system as a resource. For example, the introduction of work qubits to assist with implementation of a quantum gate.

Mathematically, we combine our state $\rho_A \in \mathrm{Dens}(\mathcal{H}_B)$ with the new system $\mathcal{H}_B$ via a state $\rho_B \in \mathrm{Dens}(\mathcal{H}_B)$ and the tensor product,

$$\rho_A \mapsto \rho_A \otimes \rho_B \in \mathrm{Dens}(\mathcal{H}_A \otimes \mathcal{H}_B).$$

We can now consider the evolution of the composite system $\rho_A \otimes \rho_B$ in $\mathcal{H}_A \otimes \mathcal{H}_B$. Namely, for unitary operator $U \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we have the evolution $U(\rho_A \otimes \rho_B)U^* \in \mathrm{Dens}(\mathcal{H}_A \otimes \mathcal{H}_B)$. If this is the only process we do on the composite state, we can then return to just the initial system $\mathcal{H}_A$ via the partial trace,

$$\mathrm{Tr}^B\big(U(\rho_A \otimes \rho_B)U^*\big) \in \mathrm{Dens}(\mathcal{H}_A).$$

Because $U(\rho_A \otimes \rho_B)U^* \in \mathrm{Dens}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we know that $\mathrm{Tr}^B\big(U(\rho_A \otimes \rho_B)U^*\big)$ is a density operator by Theorem 4.18. Hence, the map

$$\Phi : \mathrm{Dens}(\mathcal{H}_A) \to \mathrm{Dens}(\mathcal{H}_A), \qquad \Phi(\rho_A) = \mathrm{Tr}^B\big(U(\rho_A \otimes \rho_B)U^*\big)$$

is well-defined for any $\rho_B \in \mathrm{Dens}(\mathcal{H}_B)$ and unitary $U \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$. We can extend the above map Φ to a map on any linear operator.

Lemma 7.1. *Let $\rho_B \in \mathrm{Dens}(\mathcal{H}_B)$ and $U \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Define the map*

$$\Phi : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_A), \qquad \Phi(T) = \mathrm{Tr}^B\big(U(T \otimes \rho_B)U^*\big).$$

Then Φ is linear, trace-preserving and completely positive*, for any $k \in \mathbb{N}$ and positive operator*

$S \in \mathcal{L}(\mathbb{C}^k \otimes \mathcal{H})$, the operator $(\mathbf{1}_k \otimes \Phi)(S) \in \mathcal{L}(\mathbb{C}^k \otimes \mathcal{H})$ is positive.

Proof. The linearity of Φ is a simple check. Using the properties of the partial trace,

$$\begin{aligned}\mathrm{Tr}_A\big(\Phi(T)\big) &= \mathrm{Tr}_A\big(\mathrm{Tr}^B(U(T\otimes\rho_B)U^*)\big) = \mathrm{Tr}_{A\otimes B}\big(U(T\otimes\rho_B)U^*\big)\\ &= \mathrm{Tr}_{A\otimes B}\big(T\otimes\rho_B\big) = \mathrm{Tr}_A(T)\ \mathrm{Tr}_B(\rho_B) = \mathrm{Tr}_A(T),\end{aligned}$$

which shows that Φ is trace-preserving. To show Φ is completely positive, we first note that the map $\mathbf{1}_k \otimes \Phi$ is still given via the partial trace Tr^B,

$$\mathbf{1}_k\otimes\Phi(S) = \mathrm{Tr}^B\big(U(S\otimes\rho_B)U^*\big), \quad U\in\mathcal{L}(\mathbb{C}^k\otimes\mathcal{H}_A\otimes\mathcal{H}_B), \quad \mathrm{Tr}^B : \mathcal{L}(\mathbb{C}^k\otimes\mathcal{H}_A\otimes\mathcal{H}_B)\to\mathcal{L}(\mathbb{C}^k\otimes\mathcal{H}_A).$$

The result follows because the partial trace preserves positivity, $R \in \mathcal{L}(\mathcal{H}\otimes\mathcal{K})$ positive implies that $\mathrm{Tr}^{\mathcal{K}}(R)$ is positive. □

The map Φ provides an example of a quantum channel, a trace-preserving and completely positive map. More generally, we can consider maps between operators on different Hilbert spaces.

Definition 7.2. Let $\mathcal{H}$ and $\mathcal{K}$ be finite-dimensional Hilbert spaces. A quantum channel is a linear, trace-preserving and completely positive map

$$\Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K}), \qquad \mathrm{Tr}_{\mathcal{K}}\big(\Phi(T)\big) = \mathrm{Tr}_{\mathcal{H}}(T) \text{ for all } T\in\mathcal{L}(\mathcal{H}).$$

In particular, $\Phi : \mathrm{Dens}(\mathcal{H}) \to \mathrm{Dens}(\mathcal{K})$.

A quantum channel is one example of an operation that we can do a state $\rho_A \in \mathrm{Dens}(\mathcal{H}_A)$. Because all properties of a density operator are preserved, it is a particularly nice example. But there are other processes we may wish to consider for the composite state $\rho_A\otimes\rho_B \in \mathrm{Dens}(\mathcal{H}_A\otimes\mathcal{H}_B)$. For example, we may wish to take a measurement in the environment $\mathcal{H}_B$ before taking the partial trace to $\mathcal{H}_A$. Suppose we measure the observable $\mathbf{1}_{\mathcal{H}_A}\otimes B \in \mathcal{L}(\mathcal{H}_A\otimes\mathcal{H}_B)$ and return the eigenvalue $\lambda\in\sigma(B)=\sigma(\mathbf{1}_{\mathcal{H}_A}\otimes B)$. Then if P_B is the corresponding eigenspace projection, then the projection postulate says that the composite state will collapse to

$$\rho_A\otimes\rho_B \xrightarrow{\text{measurement}} \frac{(\mathbf{1}_A\otimes P_B)(\rho_A\otimes\rho_B)(\mathbf{1}_A\otimes P_B)}{\mathrm{Tr}((\mathbf{1}_A\otimes P_B)(\rho_A\otimes\rho_B))},$$

where we write $\mathbf{1}_A = \mathbf{1}_{\mathcal{H}_A}$ for simplicity. More generally still, we might take a measurement after having evolved the composite state $\rho_A\otimes\rho_B$,

$$\rho_A\otimes\rho_B \mapsto U(\rho_A\otimes\rho_B)U^* \mapsto \frac{(\mathbf{1}_A\otimes P_B)U(\rho_A\otimes\rho_B)U^*(\mathbf{1}_A\otimes P_B)}{\mathrm{Tr}((\mathbf{1}_A\otimes P_B)U(\rho_A\otimes\rho_B)U^*)}.$$

We can once again restrict to $\mathcal{H}_A$, our 'primary system' by taking the partial trace over $\mathcal{H}_B$. So in total we have a map

$$\rho_A \mapsto \frac{\mathrm{Tr}^B((\mathbf{1}_A\otimes P_B)U(\rho_A\otimes\rho_B)U^*(\mathbf{1}_A\otimes P_B))}{\mathrm{Tr}((\mathbf{1}_A\otimes P_B)U(\rho_A\otimes\rho_B)U^*)} \in \mathrm{Dens}(\mathcal{H}_A)$$

where we have used the fact Tr^B is linear. Let us briefly ignore the denominator term, which is used for normalisation and just consider the map $\rho_A \mapsto \mathrm{Tr}^B\big((\mathbf{1}_A\otimes P_B)U(\rho_A\otimes\rho_B)U^*(\mathbf{1}_A\otimes P_B)\big)$. Then

$$\begin{aligned}\mathrm{Tr}_A\big(\mathrm{Tr}^B\big((\mathbf{1}_A\otimes P_B)U(\rho_A\otimes\rho_B)U^*(\mathbf{1}_A\otimes P_B)\big)\big) &= \mathrm{Tr}_{A\otimes B}\big((\mathbf{1}_A\otimes P_B)U(\rho_A\otimes\rho_B)U^*\big)\\ &\le \mathrm{Tr}_{A\otimes B}\big(U(\rho_A\otimes\rho_B)U^*\big) = \mathrm{Tr}_{A\otimes B}(\rho_A\otimes\rho_B) = 1.\end{aligned}$$

So ignoring the denominator, we have a map that preserves positivity and is such that

$$0 \leq \mathrm{Tr}_A \left(\mathrm{Tr}^B \left((\mathbf{1}_A \otimes P_B) U (\rho_A \otimes \rho_B) U^* (\mathbf{1}_A \otimes P_B) \right) \right) \leq 1.$$

If we let $V = (\mathbf{1}_A \otimes P_B) U$, then we see that our map can be written as $\mathrm{Tr}^B \left(V(\rho_A \otimes \rho_B) V^* \right)$, where

$$V^* V = U^* (\mathbf{1} \otimes P_B) U = \mathbf{1} \otimes P' \leq \mathbf{1}_{A \otimes B}$$

and we recall that $T \leq S$ if $S - T$ is positive.

Definition 7.3. Let $\mathcal{H}$ and $\mathcal{K}$ be finite-dimensional Hilbert spaces. A quantum operation is a linear completely positive map $\Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$ such that $\mathrm{Tr}_{\mathcal{K}} \left(\Phi(T) \right) \leq \mathrm{Tr}_{\mathcal{H}}(T)$ for any positive operator $T \in \mathcal{L}(\mathcal{H})$. (In particular, $\mathrm{Tr}_{\mathcal{K}} \left(\Phi(\rho_A) \right) \leq 1$ for any $\rho \in \mathrm{Dens}(\mathcal{H})$.)

We see that quantum channels are a special case of quantum operations.

Exercise 7.1. Let $\mathcal{H}_B$ be a finite-dimensional Hilbert space and $\rho_B \in \mathrm{Dens}(\mathcal{H})$. Then for any operator $V \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $V^* V \leq \mathbf{1}_{A \otimes B}$, show that

$$\Phi(T) = \mathrm{Tr}^B \left(V(T \otimes \rho_B) V^* \right) \in \mathcal{L}(\mathcal{H}_A), \qquad T \in \mathcal{L}(\mathcal{H}_A)$$

defines a quantum operation.

Remark 7.4. The partial trace and other quantum operations are maps between *operators on Hilbert spaces*. This is in contrast to observables, which are operators between Hilbert spaces. However, we can reconcile these viewpoints. Recall that $\mathcal{L}(\mathcal{H})$ is also a Hilbert space with inner-product $\langle A \mid B \rangle = \mathrm{Tr}(A^* B)$. Then, we can understand a quantum channel as a positive linear map between the Hilbert spaces,

$$\Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K}), \qquad \Phi \geq 0, \qquad \|\Phi\| = \sup_{A \in \mathcal{L}(\mathcal{H}), \|A\| = 1} \big\| \Phi(A) \big\| \leq 1.$$

Consider a quantum channel $\Phi : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_A)$. We have seen that such channels are often constructed by embedding an operator in a larger tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ and then restricting via the partial trace. A striking and property of quantum operations is that we can understand this map using only information from the Hilbert space $\mathcal{H}_A$.

Theorem 7.5 (Kraus' Theorem, [3, Theorem 8.1, 8.3])**.** *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be Hilbert spaces of dimension n and m respectively. If $\Phi : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is a quantum operation, then there exist linear operators $\{K_j\}_{j=1}^M$, where $K_j : \mathcal{H}_A \to \mathcal{H}_B$ and $M \leq nm$ such that for all $T \in \mathcal{L}(\mathcal{H}_A)$,*

$$\Phi(T) = \sum_{j=1}^M K_j T K_j^*, \qquad \sum_{j=1}^M K_j^* K_j \leq \mathbf{1}_{\mathcal{H}_A}.$$

If Φ is a quantum channel, then $\sum_j K_j^* K_j = \mathbf{1}_{\mathcal{H}_A}$.

The equality $\Phi(T) = \sum_{j=1}^M K_j T K_j^*$ is sometimes called the operator-sum representation of Φ and the operators $\{K_j\}_{j=1}^M$ are sometimes called Kraus operators. We will not prove Kraus' Theorem, though will make a few remarks. To construct the operators K_j, one uses the Riesz Representation Theorem (Theorem 2.7) and the action of $\mathbf{1}_n \otimes \Phi$ on the density matrix $|\Psi\rangle\langle\Psi|$ built from the 'maximally entangled state' $|\Psi\rangle = \sum_j |e_j \otimes e_j\rangle \in \mathcal{H}_A \otimes \mathcal{H}_A$.

Let us also remark on the generality of Kraus' Theorem. Suppose that $\mathcal{H} = \mathbb{C}^n$ and consider a family of unitaries, $\{X_i\}_{i=1}^m$. Then for any set of probabilities $\{p_i\}_{i=1}^m$, we can define $K_i = \sqrt{p_i} X_i$ and the quantum channel

$$\Phi(T) = \sum_{i=1}^m p_i X_i T X_i^*, \qquad \sum_{i-1}^m K_i^* K_i = \sum_{i=1}^m p_i X_i^* X_i = \sum_{i=1}^m p_i \mathbf{1}_n = \mathbf{1}_n.$$

Recall Theorem 3.16, which characterises when different ensembles are represented by the same density operator. Similarly, the Kraus operators of a quantum operation are not unique and we can characterise this non-uniqueness.

Theorem 7.6 ([3, Theorem 8.2]). *Let $\Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$ be a quantum operation. Then*

$$\Phi(T) = \sum_{j=1}^M K_j T K_j^* = \sum_{j=1}^M L_j T L_j^* \quad \iff \quad K_j = \sum_{k=1}^M u_{jk} L_k$$

such that $\{u_{jk}\}_{j,k=1}^M$ forms a unitary matrix.

Let us return to our motivating example, the interaction of a state ρ_A with an environment where in the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$, an evolution or measurement may take place,

$$\rho_A \mapsto \mathrm{Tr}^B \big((\mathbf{1}_A \otimes P_B) U (\rho_A \otimes \rho_B) U^* (\mathbf{1}_A \otimes P_B) \big) = \sum_{j=1}^M K_j \rho_A K_j^*.$$

We obtain a quantum channel when $P_B = \mathbf{1}_B$, no information is lost from the interaction of ρ_A with the environment. When we do take a (non-trivial) measurement, then we generally have that $\mathrm{Tr}_A\big(\Phi(A)\big) < 1$ and the operation Φ is not sufficient to give a complete description on what has happened to the state ρ_A.

Let us try and obtain a more concrete understanding of the operator-sum representation. Suppose we have a quantum channel, so $\sum_j K_j^* K_j = \mathbf{1}_A$. By further supplementing our 'environment' Hilbert space $\mathcal{H}_B$ and taking a purification, we can assume that $\rho_B = |f_1\rangle\langle f_1|$ for some state $|f_1\rangle \in \mathcal{H}_B$. Using the Gram–Schmidt procedure, $|f_1\rangle$ can be completed to an orthonormal basis $\{|f_k\rangle\}_{k=1}^m \subset \mathcal{H}_B$. We can therefore consider the observables $\mathbf{1}_A \otimes |f_k\rangle\langle f_k|$ for any k So we can evolve our composite state and consider such a measurement,

$$\rho_A \otimes |f_1\rangle\langle f_1| \mapsto U(\rho_A \otimes |f_1\rangle\langle f_1|)U^* \mapsto (\mathbf{1}_A \otimes |f_k\rangle\langle f_k) U(\rho_A \otimes |f_1\rangle\langle f_1|)U^* (\mathbf{1}_A \otimes |f_k\rangle\langle f_k).$$

The probability that a measurement of $\mathbf{1}_A \otimes |f_k\rangle\langle f_k|$ returns 1 is given by the trace

$$\begin{aligned} \mathbf{P}_1(\mathbf{1}_A \otimes |f_k\rangle\langle f_k|) &= \mathrm{Tr}_{A\otimes B}\big((\mathbf{1}_A \otimes |f_k\rangle\langle f_k) U(\rho_A \otimes |f_1\rangle\langle f_1|)U^* \big) \\ &= \mathrm{Tr}_B \big(|f_k\rangle\langle f_k| \, \mathrm{Tr}^A (U(\rho_A \otimes |f_1\rangle\langle f_1|)U^*) \big) \\ &= \mathrm{Tr}_A \big(K_k \rho_A K_k \big). \end{aligned}$$

If we let $\rho_A^{(k)} = \dfrac{K_k \rho_A K_k}{\mathrm{Tr}_A\big(K_k \rho_A K_k\big)}$ be the normalised state, then

$$\Phi(\rho_A) = \sum_{j=1}^M K_j \rho_A K_j^* = \sum_{j=1}^M \mathbf{P}_1(\mathbf{1}_A \otimes |f_j\rangle\langle f_j|)\, \rho_k.$$

That is, the operator-sum decomposition gives a probabilistic expansion of the state ρ_A coming from its exposure to an environment where possible measurements can be taken.

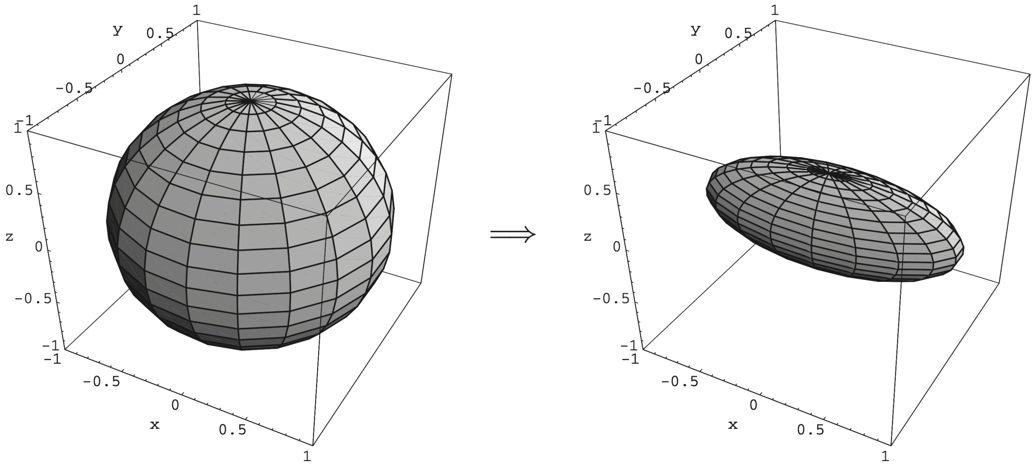


Figure 3: Bit-flip channel on the Bloch ball [3, Figure 8.8].

Example 7.7 (Bit-flip channel)**.** We consider $\mathcal{H} = \mathbb{C}^2$, where we recall that any state $\rho \in \operatorname{Dens}(\mathbb{C}^2)$,

$$\rho \sim \rho_{\mathbf{x}} = \frac{1}{2}\big(\mathbf{1}_2 + \mathbf{x}\cdot\sigma\big) = \frac{1}{2}\begin{pmatrix} 1+x_3 & x_1 - ix_2 \\ x_1 + ix_2 & 1 - x_3 \end{pmatrix} \in \operatorname{Dens}(\mathbb{C}^2),$$

where $\mathbf{x} \in \overline{B_{\mathbb{R}^3}(\mathbf{0},1)}$, the closed unit Ball in $\mathbb{R}^3$. Let us consider a quantum channel that encodes the introduction of a possible error. Namely, starting from a state $|x\rangle$, the channel will change $|x\rangle$ to $|1 \oplus x\rangle$ with probability $p \in [0,1]$ and leave it as is with probability $1-p$. We describe this channel via its operator-sum representation,

$$\Phi(\rho) = pX\rho X + (1-p)\rho.$$

Therefore, we can take Kraus operators $K_1 = \sqrt{p}X$ and $K_2 = (1-p)\mathbf{1}_2$. A simple computation will give that

$$\begin{aligned}\Phi(\rho_{\mathbf{x}}) &= \Phi\left(\frac{1}{2}\begin{pmatrix} 1+x_3 & x_1 - ix_2 \\ x_1 + ix_2 & 1 - x_3 \end{pmatrix}\right) \\ &= \frac{p}{2}\begin{pmatrix} 1-x_3 & x_1 + ix_2 \\ x_1 - ix_2 & 1 + x_3 \end{pmatrix} + \frac{1-p}{2}\begin{pmatrix} 1+x_3 & x_1 - ix_2 \\ x_1 + ix_2 & 1 - x_3 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 1+(1-2p)x_3 & x_1 - i(1-2p)x_2 \\ x_1 + i(1-2p)x_2 & 1-(1-2p)x_3 \end{pmatrix}.\end{aligned}$$

A plot of this bit-flip transformation on the unit ball for $p = 0.7$ is given In Figure 3.

Exercise 7.2. Explore the geometric effects of quantum operations on states in single Qubit space via the equivalence of $\rho \in \operatorname{Dens}(\mathbb{C}^2)$ with points in the Bloch ball $\{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\| \leq 1\}$.

7.2 Measures of quantum information

In the previous subsection we introduced quantum operations, the basic transformation that can be done to a quantum state $\rho \in \operatorname{Dens}(\mathcal{H})$ such that the quantum information it contains is all or partially preserved. Of course, if a quantum state undergoes a transformation $\rho \mapsto \Phi(\rho)$, then it would be useful to keep track of how far we have moved from our original state. This is particularly important for keeping track of possible errors and noise so that these issues can be corrected.

Classical trace distance and fidelity

It will be useful to briefly review the classical setting. While we have a good analogy with a Qubit $|x\rangle = |x_{n-1}\cdots x_1x_0\rangle$ with a classical bit $x_{n-1}\cdots x_1x_0$, what is the classical analogue of a density operator $\rho \in \mathrm{Dens}(\mathcal{H})$? Recall that any density operator has a canonical decomposition as an orthonormal basis of pure states, where there is an associated probability of being in any state of the ensemble,

$$\rho = \sum_{j=1}^{n} p_j|\psi_j\rangle\langle\psi_j|, \qquad p_j \in [0,1], \qquad \sum_{j=1}^{n} p_j = 1.$$

So any density operator gives rise to a probability distribution via its spectrum $\mathbf{P}_\rho = \{p_j\}_{j=1}^n = \sigma(\rho)$. These note are not an introduction to probability theory, an enormous topic in its own right. But in the context of finite-dimensional spaces and operators, it will suffice to consider a probability distribution as any collection of numbers $\{p_j\}_{j=1}^m$ such that $p_j \in [0,1]$ for all j and $\sum_{j=1}^m p_j = 1$.

We will therefore take the classical analogue of a density operator ρ to be given by a probability distribution $\mathbf{P} = \{p_j\}_{j=1}^n$. Classical information is contained in a process X that can produces outcomes $x_1, \dots, x_n$ with probability $p_1, \dots, p_n$.[3] We can therefore ask for various notions of distance between probability distributions as encoding something like the difference between packets of classical information.

Definition 7.8. Let $\mathbf{P} = \{p_j\}_{j=1}^m$ and $\mathbf{Q} = \{q_k\}_{k=1}^m$ be finite probability distributions.

1. The trace distance between $\mathbf{P}$ and $\mathbf{Q}$ is given by

$$D(\mathbf{P},\mathbf{Q}) = \frac{1}{2}\sum_{j=1}^{n} |p_j - q_j| \in [0,\infty)$$

2. The fidelity of the probability distributions is the number

$$F(\mathbf{P},\mathbf{Q}) = \sum_{j=1}^{n} \sqrt{p_jq_j}$$

To define the trace distance and fidelity, we require the probability $\mathbf{P}$ and $\mathbf{Q}$ to have the same number of elements. However, we can always put ourselves in this situation by adding 0s to the smaller of the two sets until they are the same size.

Exercise 7.3. Show that the trace distance is a metric, i.e. $D(\mathbf{P},\mathbf{Q}) = 0$ if and only if $\mathbf{P} = \mathbf{Q}$, $D(\mathbf{P},\mathbf{Q}) = D(\mathbf{Q},\mathbf{P})$ and

$$D(\mathbf{P},\mathbf{Q}) \leq D(\mathbf{P},\mathbf{R}) + D(\mathbf{R},\mathbf{Q})$$

for any probability distributions $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$.

The fidelity is *not* a metric, indeed $F(\mathbf{P},\mathbf{P}) = 1$. Instead we can think of the fidelity as an inner-product of probability vectors

$$F(\mathbf{P},\mathbf{Q}) = \langle\psi_\mathbf{P} \mid \psi_\mathbf{Q}\rangle, \qquad \psi_\mathbf{P} = \begin{pmatrix}\sqrt{p_1}\\ \vdots \\ \sqrt{p_n}\end{pmatrix}, \qquad \psi_\mathbf{Q} = \begin{pmatrix}\sqrt{q_1}\\ \vdots \\ \sqrt{q_n}\end{pmatrix}.$$

Note also that $\|\psi_\mathbf{P}\| = 1 = \|\psi_\mathbf{Q}\|$. So we can also think of $F(\mathbf{P},\mathbf{Q}) = \cos(\theta)$ with θ the angle between the unit vectors $\psi_\mathbf{P}, \psi_\mathbf{Q} \in \mathbb{R}^n$.

[3] A more comprehensive framework to approach classical information is via the theory of random variables. But our aim is to keep the necessary background in probability theory to a minimum.

Quantum trace distance

Let us now return to the quantum world and consider a fixed Hilbert space $\mathcal{H}$ and density operators $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$. Recall that for any operator $T \in \mathcal{L}(\mathcal{H})$, T^*T is self-adjoint and therefore diagonalisable, $T^*T = \sum_j \mu_j P_j$. We then define $|T| = \sqrt{T^*T} = \sum_j \sqrt{\mu_j} P_j$, which is used in the polar decomposition, $T = U|T|$. In particular, the operator $|\rho - \nu| = |\nu - \rho|$ is well-defined and positive.

Definition 7.9. The trace distance of quantum states is a map

$$D : \mathrm{Dens}(\mathcal{H})^{\oplus 2} \to [0, \infty), \qquad D(\rho, \nu) = \frac{1}{2} \operatorname{Tr}\big(|\rho - \nu|\big).$$

Example 7.10. Let's consider the case that ρ and ν are simultaneously diagonalisable (e.g. the commutator $[\rho, \nu] = 0$). Then

$$\rho = \sum_{j=1}^n p_j |\psi_j\rangle\langle\psi_j|, \qquad \nu = \sum_{j=1}^n q_j |\psi_j\rangle\langle\psi_j|, \qquad |\rho - \nu| = \sum_{j=1}^n |p_j - q_j| |\psi_j\rangle\langle\psi_j|$$

and so

$$D(\rho, \nu) = \frac{1}{2} \operatorname{Tr}\Big(\sum_{j=1}^n |p_j - q_j| |\psi_j\rangle\langle\psi_j|\Big) = \frac{1}{2} \sum_{j=1}^n |p_j - q_j| = D(\mathbf{P}, \mathbf{Q}).$$

Therefore we recover the classical trace distance in the special case where our density operators are simultaneously diagonalisable.

Example 7.11 (Trace distance on the Bloch ball)**.** Let us consider $\mathcal{H} = \mathbb{C}^2$, where we then take density operators

$$\rho_{\mathbf{x}} = \frac{1}{2}(\mathbf{1}_2 + \mathbf{x} \cdot \sigma), \qquad \rho_{\mathbf{y}} = \frac{1}{2}(\mathbf{1}_2 + \mathbf{y} \cdot \sigma)$$

for points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ with length ≤ 1. We note that

$$\rho_{\mathbf{x}} - \rho_{\mathbf{y}} = \frac{1}{2}(\mathbf{x} - \mathbf{y}) \cdot \sigma.$$

It is simple to show that $\sigma\big((\mathbf{x} - \mathbf{y}) \cdot \sigma\big) = \pm\|\mathbf{x} - \mathbf{y}\|$. In particular, $\big|(\mathbf{x} - \mathbf{y}) \cdot \sigma\big|$ will have $\|\mathbf{x} - \mathbf{y}\|$ as a double-eigenvalue and hence $\operatorname{Tr}\big(\big|(\mathbf{x} - \mathbf{y}) \cdot \sigma\big|\big) = 2\|\mathbf{x} - \mathbf{y}\|$. Thus

$$D(\rho_{\mathbf{x}}, \rho_{\mathbf{y}}) = \frac{1}{2} \operatorname{Tr}\big(|\rho_{\mathbf{x}} - \rho_{\mathbf{y}}|\big) = \frac{1}{4} \operatorname{Tr}\big(\big|(\mathbf{x} - \mathbf{y}) \cdot \sigma\big|\big) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|.$$

So in this example the trace distance of states is half the Euclidean distance of points in the Bloch ball in $\mathbb{R}^3$.

Exercise 7.4. Show that quantum trace distance is a metric (cf. Exercise 7.3).

To show properties of the (quantum) trace distance, we first recall a useful result from linear algebra/functional analysis. The proof is an exercise (hint: use the spectral decomposition).

Lemma 7.12. *Let $A = A^* \in \mathcal{L}(\mathcal{H})$. Then there are positive operators $T, S \in \mathcal{L}(\mathcal{H})$ such that*

$A = T - S$ and $TS = ST = 0$.

Corollary 7.13. *Let $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$, then there are positive operators $T, S \in \mathcal{L}(\mathcal{H})$ such that $|\rho - \nu| = T + S$ and $TS = ST = 0$.*

Proof. By the previous lemma, $\rho - \nu = T - S$ with $TS = 0$. Hence

$$(\rho - \nu)^2 = T^2 + S^2 = (T + S)^2 \quad \Longrightarrow \quad |\rho - \nu| = T + S. \qquad \square$$

A natural question to ask is whether the trace distance changes under the time evolution of states.

Lemma 7.14. *Let $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$.*

1. *For any unitary $U \in \mathcal{L}(\mathcal{H})$, $D(\rho, \nu) = D\big(U\rho U^*, U\nu U^*\big)$.*
2. $$D(\rho, \nu) = \max\big\{ \operatorname{Tr}\big(P(\sigma - \nu)\big) \,\big|\, P = P^* = P^2 \in \mathcal{L}(\mathcal{H})\big\}$$

Proof. It is simple to verify that $(U\rho U^* - U\nu U^*)^2 = U(\rho - \nu)^2 U^*$ and so $|U\rho U^* - U\nu U^*| = U|\rho - \nu|U^*$. Hence $\operatorname{Tr}(|\rho - \nu|) = \operatorname{Tr}(|U\rho U^* - U\nu U^*|)$, which shows part (1).
For part (2), we have that $0 = \operatorname{Tr}(\rho) - \operatorname{Tr}(\nu) = \operatorname{Tr}(\rho - \nu) = \operatorname{Tr}(T - S)$ and so $\operatorname{Tr}(T) = \operatorname{Tr}(S)$, where T and S come from the decomposition in Corollary 7.13. Then

$$D(\rho, \nu) = \frac{1}{2} \operatorname{Tr}\big(|\rho - \nu|\big) = \frac{1}{2} \operatorname{Tr}(T + S) = \operatorname{Tr}(T).$$

For any projection $P \in \mathcal{L}(\mathcal{H})$,

$$\operatorname{Tr}\big(P(\rho - \nu)\big) \leq \operatorname{Tr}\big(P(T - S)\big) \leq \operatorname{Tr}(PT) \leq \operatorname{Tr}(T) = D(\rho, \nu).$$

We now let P_T be the projection onto $\mathrm{Ker}(T)^\perp$, where $SP_T = 0$ and $TP_T = T$. Then

$$\operatorname{Tr}\big(P_T(\rho - \nu)\big) = \operatorname{Tr}\big(P_T(T - S)\big) = \operatorname{Tr}(P_T T) - \operatorname{Tr}(P_T S) = \operatorname{Tr}(T) = D(\rho, \nu)$$

and therefore $D(\rho, \nu) = \max\big\{ \operatorname{Tr}(P(\sigma - \nu)) \,\big|\, P = P^* = P^2 \in \mathcal{L}(\mathcal{H})\big\}$. $\square$

Hence a unitary evolution of states will not affect the trace distance between states. Another natural question is the effect on the trace distance when a quantum operation is applied. Perhaps surprisingly, quantum operations are *contractive* under the trace distance.

Theorem 7.15. *Let $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$ and $\Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$ be a quantum operation. Then*

$$D\big(\Phi(\rho),\, \Phi(\nu)\big) \leq D(\rho,\, \nu).$$

Proof. We again use the description $|\rho - \nu| = T + S$ from Corollary 7.13. Then by properties of the trace distance and quantum operations,

$$\begin{aligned} D(\rho, \nu) &= \operatorname{Tr}(T) \geq \operatorname{Tr}\big(\Phi(T)\big) = \operatorname{Tr}\big(P_{\Phi(T)}\Phi(T)\big) \\ &\geq \operatorname{Tr}\big(P_{\Phi(T)}(\Phi(T) - \Phi(S))\big) = \operatorname{Tr}\big(P_{\Phi(T)}\Phi(T - S)\big) \\ &= \operatorname{Tr}\big(P_{\Phi(T)}(\Phi(\rho) - \Phi(\nu))\big) = D\big(\Phi(\rho), \Phi(\nu)\big), \end{aligned}$$

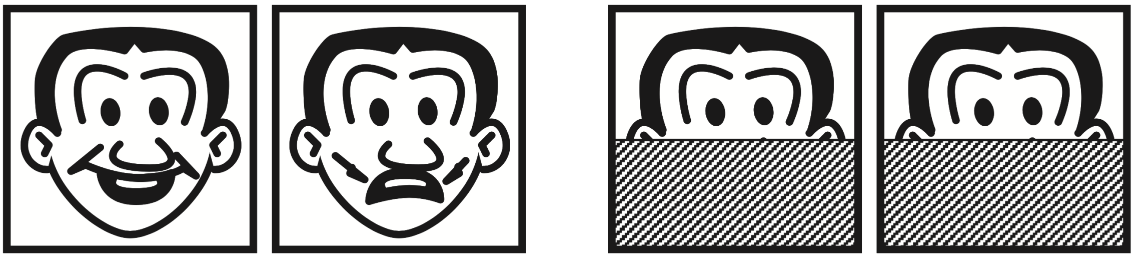


Figure 4: Objects become less distinguishable when only partial information is available. [3, Figure 9.5].

where $P_{\Phi(T)}$ is the projection onto the $\mathrm{Ker}(\Phi(T))^{\perp}$. □

The theorem says that any physical process that transmits some quantum information does not increase the distance between states. Applying a quantum operation, we may lose information about the states. So states cannot become *more* distinguishable when we only have partial information, see Figure 4.

As an example, consider $\rho_A, \nu_A \in \mathrm{Dens}(\mathcal{H}_A)$ and $\rho_B \otimes \nu_B \in \mathrm{Dens}(\mathcal{H}_B)$. Then

$$D(\rho_A, \nu_A) = D\big(\,\mathrm{Tr}^B(\rho_A \otimes \rho_B),\, \mathrm{Tr}^B(\nu_A \otimes \nu_B)\big) \le D\big(\rho_A \otimes \rho_B,\, \nu_A \otimes \nu_B\big)$$

as Tr^B is an example of a quantum operation.

Quantum fidelity

We now consider the quantum analogue of fidelity of a probability distribution.

Definition 7.16. Let $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$. Then the fidelity is given by

$$F(\rho,\nu) = \mathrm{Tr}\left(\sqrt{\rho^{1/2}\nu\rho^{1/2}}\right)$$

The definition needs a few comments. First, notice that $\rho^{1/2}\nu\rho^{1/2}$ is a positve operator as for any $|\psi\rangle \in \mathcal{H}$,

$$\langle \psi \mid \rho^{1/2}\nu\rho^{1/2}\psi\rangle = \langle \rho^{1/2}\psi \mid \nu^{1/2}\nu^{1/2}\rho^{1/2}\psi\rangle = \langle \nu^{1/2}\rho^{1/2}\psi \mid \nu^{1/2}\rho^{1/2}\psi\rangle \ge 0.$$

This means that $\sqrt{\rho^{1/2}\nu\rho^{1/2}}$ is well-defined and positive. Using the fact that $\sqrt{UAU^*} = U\sqrt{A}U^*$ for any unitary $U \in \mathcal{L}(\mathcal{H})$ and $A \ge 0$, we also have the property that

$$F(\rho,\nu) = F\big(U\rho U^*, U\nu U^*\big)$$

for any unitary $U \in \mathcal{L}(\mathcal{H})$.

Example 7.17. We consider the case that $[\rho,\nu] = 0$ and so are simultaneously diagonalisable. Then

$$\sqrt{\rho^{1/2}\nu\rho^{1/2}} = \sqrt{\rho\nu} = \sum_{j=1}^{m} \sqrt{p_j q_j}|\psi_j\rangle\langle\psi_j|,$$

where $\sigma(\rho) = \{p_j\}_{j=1}^m$ and $\sigma(\nu) = \{q_j\}_{j=1}^m$. Hence we have that

$$D(\rho,\nu) = \mathrm{Tr}\left(\sum_{j=1}^{m} \sqrt{p_j q_j}|\psi_j\rangle\langle\psi_j|\right) = \sum_{j=1}^{m} \sqrt{p_j q_j} = D(\mathbf{P}_\rho, \mathbf{Q}_\nu),$$

where $\mathbf{P}_\rho$ and $\mathbf{Q}_\nu$ are the probability distributions associated to ρ and ν. So in this special case, we recover the classical fidelity.

Example 7.18. Suppose that $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle \in \mathcal{H}$. Then $\rho^{1/2} = |\psi\rangle\langle\psi|$ and for any $\nu \in \mathrm{Dens}(\mathcal{H})$,

$$F(|\psi\rangle\langle\psi|, \nu) = \mathrm{Tr}\left(\sqrt{\langle\psi \mid \nu\psi\rangle\, |\psi\rangle\langle\psi|}\right) = \sqrt{\langle\psi \mid \rho\psi\rangle} = \sqrt{\langle\rho\rangle_\psi}.$$

When ν is also pure, $\nu = |\phi\rangle\langle\phi|$, then

$$F\big(|\psi\rangle\langle\psi|,\, |\phi\rangle\langle\phi|\big) = \sqrt{\langle\psi \mid \phi\rangle\langle\phi \mid \psi\rangle} = \big|\langle\psi \mid \phi\rangle\big|.$$

We see that for pure states $F(\rho_\psi, \nu_\phi) = F(\nu_\phi, \rho_\psi)$. It is not obvious at all that $F(\rho, \nu) = F(\nu, \rho)$ for more general density operators. But this will follow as an application of Uhlmann's Theorem, which we will not prove.

Theorem 7.19 (Uhlmann's Theorem, [3, Theorem 9.4]). *Let $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$. Then*

$$F(\rho,\nu) = \max\Big\{\big|\langle\Psi_\rho \mid \Phi_\nu\rangle\big| \;:\; \Psi_\rho, \Phi_\nu \in \mathcal{H}\otimes\mathcal{K} \textit{ are purifications of } \rho \textit{ and } \nu \textit{ respectively}\Big\}.$$

Uhlmann's Theorem shows us that $F(\rho, \nu) = F(\nu, \rho)$ as well as that $F(\rho, \nu) \in [0, 1]$ for any $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$. In particular, like the classical case, we can think of $F(\rho, \nu) = \cos(\theta)$ for some angle θ between the quantum states ρ, ν. That is, we define $\theta(\rho, \nu) = \cos^{-1}\big(F(\rho, \nu)\big)$.

The following should be compared to Theorem 7.15.

Theorem 7.20 ([3, Theorem 9.6, Exercise 9.18]). *Let $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$ and $\Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$ a quantum channel. Then*

$$F\big(\Phi(\rho),\, \Phi(\nu)\big) \geq F(\rho, \nu), \qquad \theta\big(\Phi(\rho),\, \Phi(\nu)\big) \leq \theta(\rho, \nu).$$

We have introduced two notions of distance between quantum states, a metric D and $0 \leq F \leq 1$ such that $F(\rho, \rho) = 1$. If $D(\rho, \nu)$ is small, then we would also expect $|1 - F(\rho, \nu)|$ to also be small.

Exercise 7.5. Let $|\psi\rangle$ and $|\phi\rangle$ be pure states in $\mathcal{H}$ with density operators ρ_ψ and ρ_ϕ. Show that

$$D\big(\rho_\psi, \nu_\phi\big) = \sqrt{1 - F(\rho_\psi, \nu_\phi)^2}.$$

For more general density operators ρ and ν, we can take a purification $|\Psi_\rho\rangle$ and $|\Phi_\nu\rangle$ in $\mathcal{H}\otimes\mathcal{K}$ such that $F(\rho, \nu) = \big|\langle\Psi_\rho \mid \Phi_\nu\rangle\big|$. Then using the previous exercise,

$$\begin{aligned} D(\rho,\nu) &= D\big(\,\mathrm{Tr}^{\mathcal{K}}(|\Psi_\rho\rangle\langle\Psi_\rho|),\, \mathrm{Tr}^{\mathcal{K}}(|\Phi_\nu\rangle\langle\Phi_\nu|)\big) \leq D\big(|\Psi_\rho\rangle\langle\Psi_\rho|,\, |\Phi_\nu\rangle\langle\Phi_\nu|\big) \\ &= \sqrt{1 - F\big(|\Psi_\rho\rangle\langle\Psi_\rho|,\, |\Phi_\nu\rangle\langle\Phi_\nu|\big)^2} = \sqrt{1 - F(\rho,\nu)^2}. \end{aligned}$$

So we can say that $F(\rho, \nu) \to 1$ implies that $D(\rho, \nu) \to 0$.

Example 7.21. An obvious application of the trace distance and fidelity is to see to what extent a quantum operation changes a state, namely we consider $D\big(\rho, \Phi(\rho)\big)$ and $F\big(\rho, \Phi(\rho)\big)$. As a simple example, we consider the depolarising quantum channel. We fix $\mathcal{H} = \mathbb{C}^n$ and consider an operation that sends the state ρ to $\frac{1}{n}\mathbf{1}_n$ with probability $p \in [0, 1]$ and does nothing otherwise,

$$\Phi(\rho) = \frac{p}{n}\mathbf{1}_n + (1 - p)\rho.$$

It is clear that the states ρ and $\Phi(\rho)$ commute and so the trace distance and fidelity can be computed classically via the probability distributions

$$\rho \sim \{q_j\}_{j=1}^n, \qquad \frac{p}{n}\mathbf{1}_n + (1-p)\rho \sim \{\frac{p}{n} + (1-p)q_j\}_{j=1}^n,$$

where we have used the fact that

$$\frac{p}{n}\mathbf{1}_n + (1-p)\rho = \sum_{j=1}^n \big(\frac{p}{n} + (1-p)q_j\big)|\psi_j\rangle\langle\psi_j|.$$

Hence

$$D\big(\rho, \Phi(\rho)\big) = \frac{1}{2}\sum_{j=1}^n \big|\frac{p}{n} + (1-p)q_j - q_j\big| = p\sum_{j=1}^n \big|\frac{1}{n} - q_j\big|.$$

In particular we see that $D\big(\rho, \Phi(\rho)\big)$ increases as the depolarisation probability p increases. Similarly

$$F\big(\rho, \Phi(\rho)\big) = \sum_{j=1}^n \sqrt{\big(\frac{p}{n} + (1-p)q_j\big)q_j}.$$

In the case that $\rho = |\psi\rangle\langle\psi|$, then we can further simplify

$$\begin{aligned} F\big(\rho_\psi, \Phi(\rho_\psi)\big) &= \sqrt{\langle\psi \mid \Phi(\rho_\psi)\psi\rangle} = \sqrt{\langle\psi \mid \big(\frac{p}{n}\mathbf{1}_n + (1-p)|\psi\rangle\langle\psi|\big)\psi\rangle} \\ &= \sqrt{\frac{p}{n} + (1-p)} = \sqrt{1 - \frac{n-1}{n}p}. \end{aligned}$$

When we are in the space of single Qubits, $n = 2$ the fidelity reduces to $\sqrt{1 - \frac{p}{2}}$.

7.3 Entropy of quantum states

Entropy is a word used in various fields, from thermodynamics and statistical mechanics to ecology and computer science. Here we will be interested in understanding entropy in the context of (quantum) information theory. To begin, we first consider the classical setting and Shannon entropy.

Shannon entropy

Recall that our classical analogue of a quantum state $\rho \in \mathrm{Dens}(\mathcal{H})$ is a probability distribution $\{p_j\}_{j=1}^n$, $p_j \in [0,1]$ and $\sum_j p_j = 1$. This probability distribution distributes information in the following sense: we have some quantity X, which may take certain values $x_1, \ldots, x_n$ with probability $p_1, \ldots, p_n$. By learning the outcome of X, we have gained some information. We would like to quantify the amount of information we obtain (on average) by learning the value of X.

Suppose that $p_1 = 1$ and $p_2 = \ldots = p_n = 0$. Then we know that X will take value x_1 and we gain no meaningful information by learning this quantity. On the other hand. Suppose that x_j describes a very unlikely event, such as the numbers of a winning lottery ticket. Then learning this value would be very useful! Hence we expect the information gained from learning the value x_j to be related to $\frac{1}{p_j}$. We want this quantity to be zero when $p_j = 1$, so we therefore consider $p_j \log\big(\frac{1}{p_j}\big) = -p_j \log(p_j)$ with the convention that $0\log(0) = 0$. By taking the probabilistic expectation over such quantities, we learn the average information gained by learning an outcome of X.

Definition 7.22. The Shannon entropy of a finite probability distribution $\mathbf{P} = \{p_j\}_{j=1}^n$ is given by

$$H(\mathbf{P}) = -\sum_{j=1}^n p_j \log(p_j),$$

where we use the convention $0\log(0) := 0$.

We should clarify that in the context of information theory, one typically takes $\log(x) = \log_2(x)$. The base of the logarithm does not play a crucial role in the basic properties of the Shannon entropy, so we will not emphasise this point. The convention $0\log(0) := 0$ is partially justified by noting that $\lim\limits_{x\to 0^+} x\log(x) = 0$.

Example 7.23. Suppose that we have fair coin, so a coin flip will give heads/tails with probability $\frac{1}{2}$. The associated probability distribution is $\{\frac{1}{2}, \frac{1}{2}\}$ and

$$H(\mathbf{P}) = -2\frac{1}{2}\log_2\big(\frac{1}{2}\big) = \log_2(2) = 1.$$

On the other hand, if we have an unfair coin, say the probability of heads is $\frac{3}{4}$, then

$$\begin{aligned} H(\tilde{\mathbf{P}}) &= -\frac{3}{4}\log\big(\frac{3}{4}\big) - \frac{1}{4}\log\big(\frac{1}{4}\big) = -\frac{3}{4}\big(\log(3) - \log(4)\big) + \frac{1}{4}\log(4) \\ &= \log(4) - \frac{3}{4}\log(3) = 2 - \frac{3}{4}\log(3) \approx 0.811. \end{aligned}$$

Taking a less fair coin means that we are more likely to be able to predict the outcome. So we gain less information by learning the value of an outcome. This is reflected in the fact that $H(\tilde{\mathbf{P}}) < H(\mathbf{P})$. Indeed, for a binary probability distribution $\{p, 1-p\}$, the Shannon entropy is maximised when $p = \frac{1}{2}$.

Given probability distributions $\mathbf{P}$ and $\mathbf{Q}$, we can compare $H(\mathbf{P})$ and $H(\mathbf{Q})$. A more direct comparison can be done via the *relative* entropy.

Definition 7.24. Let $\mathbf{P} = \{p_j\}_{j=1}^n$ and $\mathbf{Q} = \{q_j\}_{j=1}^n$ be probability distributions. The relative entropy is given by

$$H(\mathbf{P} \mid \mathbf{Q}) = \sum_{j=1}^n p_j \log\big(\frac{p_j}{q_j}\big) = -H(\mathbf{P}) - \sum_{j=1}^n p_j \log(q_j),$$

where we use the convention $0\log(0) = 0$ and $-p_j\log(0) = +\infty$ if $p_j > 0$.

Example 7.25. Let's consider two binary probability distributions $\mathbf{P} = \{p, 1-p\}$ and $\mathbf{Q} = \{q, 1-q\}$ with $p, q \in (0,1)$. Then

$$\begin{aligned} H(\mathbf{P} \mid \mathbf{Q}) &= p\log\big(\frac{p}{q}\big) + (1-p)\log\big(\frac{1-p}{1-q}\big) \\ &= p\big(\log(p) - \log(q)\big) + (1-p)\big(\log(1-p) - \log(1-q)\big). \end{aligned}$$

We see that $H(\mathbf{P} \mid \mathbf{Q}) = 0$ when $p = q$, i.e. $\mathbf{P} = \mathbf{Q}$.

Theorem 7.26. *The relative entropy is non-negative,* $H(\mathbf{P} \mid \mathbf{Q}) \in [0, +\infty]$*, and* $H(\mathbf{P} \mid \mathbf{Q}) = 0$ *if and only if* $\mathbf{P} = \mathbf{Q}$.

Proof. We will make use of the inequality $\log_2(x) \geq \frac{1}{\ln(2)}(1-x)$ with equality if and only if $x = 1$. We then compute

$$\begin{aligned} H(\mathbf{P} \mid \mathbf{Q}) &= \sum_{j=1}^{n} p_j \log\big(\frac{p_j}{q_j}\big) = -\sum_{j=1}^{n} p_j \log\big(\frac{q_j}{p_j}\big) \\ &\geq \frac{1}{\ln(2)} \sum_{j=1}^{n} p_j \big(1 - \frac{q_j}{p_j}\big) = \frac{1}{\ln(2)} \sum_{j} (p_j - q_j) \\ &= \frac{1-1}{\ln(2)} = 0. \end{aligned}$$

We have the equality $H(\mathbf{P} \mid \mathbf{Q}) = 0$ if and only if $\frac{q_j}{p_j} = 1$ for all j, i.e. $\mathbf{P} = \mathbf{Q}$. □

Von Neumann entropy

We now consider a notion of entropy for quantum states $\rho \in \mathrm{Dens}(\mathcal{H})$. This should reproduce the Shannon entropy when we consider the probability distribution $\mathbf{P}_\rho = \sigma(\rho)$. Hence

$$H(\mathbf{P}_\rho) = -\sum_{\lambda \in \sigma(\rho)} \lambda \log(\lambda) = -\operatorname{Tr}\big(\rho \log \rho\big).$$

So we take the following as our definition.

Definition 7.27. The von Neumann entropy of $\rho \in \mathrm{Dens}(\mathcal{H})$ is given by

$$S(\rho) = -\operatorname{Tr}\big(\rho \log(\rho)\big) = -\sum_{j=1}^{n} p_j \log(p_j), \qquad \rho = \sum_{j=1}^{n} p_j |\psi_j\rangle\langle\psi_j|,$$

where we use the convention $0 \log(0) = 0$.

Example 7.28. If $\rho = |\psi\rangle\langle\psi|$ is a pure state, then $S(|\psi\rangle\langle\psi|) = -\log(1) = 0$. Conversely, if $\mathcal{H} = \mathbb{C}^n$ and $\rho = \frac{1}{n}\mathbf{1}_n$ is the maximally mixed state, then

$$S\big(\tfrac{1}{n}\mathbf{1}_n\big) = -\sum_{j=1}^{n} \frac{1}{n} \log\big(\frac{1}{n}\big) = \log(n).$$

For other states $\rho \in \mathrm{Dens}(\mathbb{C}^n)$, $0 \leq S(\rho) \leq \log(n)$.

Example 7.29 (Quantum channels can increase or decrease entropy)**.** Let $\nu \in \mathrm{Dens}(\mathcal{K})$ and define the map

$$\Phi_\nu : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K}), \qquad \Phi_\nu(T) = \operatorname{Tr}(T)\, \nu.$$

It is straightforward to show that ν is a quantum channel. For an arbitrary $\rho \in \mathrm{Dens}(\mathcal{H})$, $S(\rho)$ and $S(\nu) = S\big(\Phi_\nu(\rho)\big)$ are independent of each other. So in general the entropy can increase or decrease under a quantum channel.

Exercise 7.6. Show that the depolarising channel, $\Phi(\rho) = \frac{p}{n}\mathbf{1}_n + (1-p)\rho$ for $\rho \in \mathrm{Dens}(\mathbb{C}^n)$ and $p \in [0,1]$, is such that $S\big(\Phi(\rho)\big) \geq S(\rho)$.

Exercise 7.7. Let $\mathbf{P} = \{p_j\}_{j=1}^n$ be a probability distribution and $\{\rho_j\}_{j=1}^n \subset \mathrm{Dens}(\mathcal{H})$. Show that

$$\sum_{j=1}^n p_j S(\rho_j) \leq S\Big(\sum_{j=1}^n p_j \rho_j\Big) \leq \sum_{j=1}^n p_j S(\rho_j) + H(\mathbf{P}).$$

Definition 7.30. For $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$, the relative entropy

$$S(\rho \mid \nu) = \mathrm{Tr}\,\big(\rho \log(\rho)\big) - \mathrm{Tr}\,\big(\rho \log(\nu)\big) = -S(\rho) - \mathrm{Tr}\,\big(\rho\log(\nu)\big)$$

and where $S(\rho \mid \nu) = +\infty$ if $\mathrm{Ker}(\nu) \cap \mathrm{Ker}(\rho)^\perp \neq \{0\}$.

We remark that $S(\rho \mid \nu) \neq S(\nu \mid \rho)$ in general.

Theorem 7.31 (Klein's inequality)**.** *The relative entropy is non-negative, $S(\rho \mid \nu) \in [0, \infty]$ and $S(\rho \mid \nu) = 0$ if and only if $\rho = \nu$.*

Proof. We take the canonical decomposition of the density operators, $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ and $\nu = \sum_k q_k |\phi_k\rangle\langle\phi_k|$. Then by definition

$$\begin{aligned}
S(\rho \mid \nu) &= \sum_{j=1}^n p_j \log(p_j) - \mathrm{Tr}\,\big(\rho\log(\nu)\big) = \sum_{j=1}^n p_i \log(p_i) - \sum_{j=1}^n \langle \psi_j \mid \rho \log(\nu)\psi_j\rangle \\
&= \sum_{j=1}^n p_j \log(p_j) - \sum_{j=1}^n \langle \rho\psi_j \mid \log(\nu)\psi_j\rangle = \sum_{j=1}^n p_j \log(p_j) - \sum_{j=1}^n p_j \langle \psi_j \mid \log(\nu)\psi_j\rangle \\
&= \sum_{j=1}^n p_j \log(p_j) - \sum_{j,k=1}^n p_j \langle \psi_j \mid \big(\log(q_k)|\phi_k\rangle\langle\phi_k|\big)\psi_j\rangle \\
&= \sum_{j=1}^n p_j \Big(\log(p_j) - \sum_{k=1}^n P_{jk}\log(q_k)\Big),
\end{aligned}$$

where $P_{jk} = |\langle \psi_j \mid \phi_k\rangle|^2 \geq 0$. We also note that

$$\sum_j P_{jk} = \sum_{j=1}^n \langle \phi_k \mid \psi_j\rangle\langle\psi_j \mid \phi_k\rangle = \langle \phi_k \mid \phi_k\rangle = 1$$

and similarly $\sum_k P_{jk} = 1$. By the concavity of the logarithm

$$\sum_{k=1}^n P_{jk}\log(q_k) \leq \log\Big(\sum_k P_{jk} q_j\Big) = \log(r_j), \qquad r_j = \sum_k P_{jk} q_j$$

with equality if and only if there is some k such that $P_{jk} = 1$. Then

$$\begin{aligned} S(\rho \mid \nu) &= \sum_{j=1}^{n} p_j \Big(\log(p_j) - \sum_{k=1}^{n} P_{jk} \log(q_k) \Big) \\ &\geq \sum_{j=1}^{n} p_j \big(\log(p_j) - \log(r_j) \big) = \sum_{j=1}^{n} p_j \log \big(\frac{p_i}{r_i} \big). \end{aligned}$$

Using the identity

$$\sum_{j=1}^{n} r_j = \sum_{j,k=1}^{n} \langle \phi_k \mid \psi_j \rangle \langle \psi_j \mid \phi_k \rangle = 1,$$

we see that $\mathbf{R} = \{r_j\}_{j=1}^{n}$ is also a probability distribution and

$$S(\rho \mid \nu) \geq \sum_{j=1}^{n} p_j \log \big(\frac{p_i}{r_i} \big) = H(\mathbf{P} \mid \mathbf{R}) \geq 0,$$

where we have used the non-negativity of the classical relative entropy (Theorem 7.26). We also obtain equality $S(\rho \mid \nu) = 0$ if and only if for each $j \in \{1, \dots, n\}$ there is a $k \in \{1, \dots, n\}$ such that $P_{jk} = 1$ and $\mathbf{P} = \mathbf{R}$. Considering $\{P_{jk}\}_{j,k=1}^{n}$ as a matrix, the condition on P is equivalent to the matrix P having a 1 in every column. This implies that P is a permutation matrix and is just a relabeling of the eigenstates of ρ and ν. Rearranging the labelling as necessary, we can assume that $P = \mathbf{1}_n$, which also implies $r_j = q_j$. Therefore $S(\rho \mid \nu)$ if and only if $p_j = q_j$ for all j, which means that the eigenvalues of ρ and ν are same and, hence, $\rho = \nu$. □

Example 7.32 (Subadditivity of the entropy in composite systems). Take $\rho_A \in \mathrm{Dens}(\mathcal{H}_A)$ and $\rho_B \in \mathrm{Dens}(\mathcal{H}_B)$ and a state in the composite system $\rho \in \mathrm{Dens}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $\mathrm{Tr}^B(\rho) = \rho_A$ and $\mathrm{Tr}^A(\rho) = \rho_B$. The product state $\rho^A \otimes \rho^B$ is an example of such a $\rho \in \mathrm{Dens}(\mathcal{H}_A \otimes \mathcal{H}_B)$, but there may be other examples in general. Let us therefore compare these systems by computing

$$S\big(\rho \,\big|\, \rho_A \otimes \rho_B\big) = -S(\rho) - \mathrm{Tr}_{A\otimes B}\big(\rho \log(\rho_A \otimes \rho_B)\big).$$

We first note that

$$\begin{aligned} \log(\rho_A \otimes \rho_B) &= \sum_{j,k} \log(p_j^A q_k^B)(|\psi_j^A\rangle\langle\psi_j^A| \otimes |\phi_k^B\rangle\langle\phi_k^B|) \\ &= \sum_{j,k} \big(\log(p_j^A) + \log(q_k^B)\big)(|\psi_j^A\rangle\langle\psi_j^A| \otimes |\phi_k^B\rangle\langle\phi_k^B|) \\ &= \Big(\sum_j \log(p_j^A)|\psi_j^A\rangle\langle\psi_j^A|\Big) \otimes \Big(\sum_k |\phi_k^B\rangle\langle\phi_k^B|\Big) + \Big(\sum_j |\psi_j^A\rangle\langle\psi_j^A|\Big) \otimes \Big(\sum_k \log(q_k^B)|\phi_k^B\rangle\langle\phi_k^B|\Big) \\ &= \log(\rho_A) \otimes \mathbf{1}_B + \mathbf{1}_A \otimes \log(\rho_B), \end{aligned}$$

which we then use to compute

$$\begin{aligned} -\mathrm{Tr}_{A\otimes B}\big(\rho \log(\rho_A \otimes \rho_B)\big) &= -\mathrm{Tr}_{A\otimes B}\big(\rho(\log(\rho_A) \otimes \mathbf{1}_B + \mathbf{1}_A \otimes \log(\rho_B))\big) \\ &= -\mathrm{Tr}_{A\otimes B}\big(\rho(\log(\rho_A) \otimes \mathbf{1}_B)\big) - \mathrm{Tr}_{A\otimes B}\big(\rho(\mathbf{1}_A \otimes \log(\rho_B))\big) \\ &= -\mathrm{Tr}_A\big(\mathrm{Tr}^B(\rho)\log(\rho_A)\big) - \mathrm{Tr}_B\big(\mathrm{Tr}^A(\rho)\log(\rho_B)\big) \\ &= -\mathrm{Tr}_A\big(\rho_A \log(\rho_A)\big) - \mathrm{Tr}_B\big(\rho_B \log(\rho_B)\big) \\ &= S(\rho_A) + S(\rho_B). \end{aligned}$$

And therefore

$$S\big(\rho\,\big|\,\rho_A\otimes\rho_B\big) = -S(\rho) - \mathrm{Tr}_{A\otimes B}\left(\rho\log(\rho_A\otimes\rho_B)\right) = -S(\rho) + S(\rho_A) + S(\rho_B).$$

Applying Klein's inequality,

$$S(\rho) \leq S(\rho_A) + S(\rho_B)$$

with equality if and only if $\rho = \rho_A\otimes\rho_B$.

Example 7.33 (Entropy on the Bloch ball)**.** Let us consider two states $\rho_{\mathbf{x}}, \rho_{\mathbf{y}} \in \mathrm{Dens}(\mathbb{C}^2)$, where

$$\rho_{\mathbf{x}} = \frac{1}{2}(\mathbf{1}_2 + \mathbf{x}\cdot\sigma), \qquad \rho_{\mathbf{y}} = \frac{1}{2}(\mathbf{1}_2 + \mathbf{y}\cdot\sigma), \qquad \mathbf{x},\mathbf{y}\in\mathbb{R}^3,\ \|\mathbf{x}\|\leq 1,\ \|\mathbf{y}\|\leq 1.$$

For the entropy of a single state, we use that $\sigma(\rho_{\mathbf{x}}) = \frac{1}{2}(1\pm\|\mathbf{x}\|)$ and so

$$\begin{aligned} S(\rho_{\mathbf{x}}) &= -\operatorname{Tr}\left(\rho_{\mathbf{x}}\log(\rho_{\mathbf{x}})\right) \\ &= -\frac{1+\|\mathbf{x}\|}{2}\log\Big(\frac{1+\|\mathbf{x}\|}{2}\Big) - \frac{1-\|\mathbf{x}\|}{2}\log\Big(\frac{1-\|\mathbf{x}\|}{2}\Big). \end{aligned}$$

So the entropy is purely determined by the length $\|\mathbf{x}\|$. We see that $S(\rho_{\mathbf{x}})$ is maximal when $\|\mathbf{x}\| = 0$, vanishes for $\|\mathbf{x}\| = 1$ and is decreasing for $0 < \|\mathbf{x}\| < 1$.

For the relative entropy, a very long and somewhat tedious computation gives that

$$S(\rho_{\mathbf{x}}\mid\rho_{\mathbf{y}}) = \frac{1}{2}\log(1-\|\mathbf{x}\|^2) + \frac{\|\mathbf{x}\|}{2}\log\Big(\frac{1+\|\mathbf{x}\|}{1-\|\mathbf{x}\|}\Big) - \frac{1}{2}\log(1-\|\mathbf{y}\|^2) - \frac{\mathbf{x}\cdot\mathbf{y}}{2\|\mathbf{y}\|}\log\Big(\frac{1+\|\mathbf{y}\|}{1-\|\mathbf{y}\|}\Big).$$

See [1, Appendix A], for example.

Lastly, we wish to examine how the entropy is effected by measurements or quantum channels.

Theorem 7.34 (Projective measurements increase entropy)**.** *Let $\{P_j\}_{j=1}^m \in \mathcal{L}(\mathcal{H})$ be a set of projections such that $P_jP_k = \delta_{j,k}P_j$ and $\sum_j P_j = \mathbf{1}_{\mathcal{H}}$. Given a state $\rho\in\mathrm{Dens}(\mathcal{H})$, we let $\rho' = \sum_j P_j\rho P_j$. Then $S(\rho') \geq \rho(S)$ with equality if and only if $\rho = \rho'$.*

Proof. Using the the fact that $\{P_j\}_{j=1}^n$ are an orthogonal and complete set of projections, we use the properties of the trace to find that

$$\begin{aligned} -\operatorname{Tr}(\rho\log(\rho')) &= -\operatorname{Tr}\Big(\sum_j P_j\rho\log(\rho')\Big) = -\operatorname{Tr}\Big(\sum_j P_j\rho\log(\rho')P_j\Big) \\ &= -\operatorname{Tr}\Big(\sum_j P_j\rho P_j\log(\rho')\Big) = -\operatorname{Tr}\left(\rho'\log(\rho')\right) \\ &= S(\rho'), \end{aligned}$$

where we have used that $[P_j,\rho'] = \sum_k [P_j, P_k\rho_k P_k] = 0$. We can then use Klein's inequality, where

$$0 \leq S(\rho\mid\rho') = -S(\rho) - \operatorname{Tr}(\rho\log(\rho')) = -S(\rho) + S(\rho')$$

and hence $S(\rho')\geq S(\rho)$. □

In contrast to projective measurements increasing entropy, the relative entropy is monotonic under quantum channels.

Theorem 7.35 (Monotonicity of the relative entropy, [5])**.** *Let* $\rho, \nu \in \mathrm{Dens}(\mathcal{H})$ *and* $\Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$ *be a quantum channel. Then*

$$S\big(\Phi(\rho) \mid \Phi(\nu)\big) \leq S(\rho \mid \phi).$$

Exercise 7.8. Write a report on *conditional entropy* and *mutual information* in the classical and quantum setting.

A The complex number system

Remark A.1. The following notes is taken from the lecture notes for G30 Complex Analysis course. More detail is given than what is needed for the special mathematics lecture, where we just need basic properties (conjugate, norm, polar form and Euler's formula).

Not every real polynomial has real solutions, for example $x^2 + 1 = 0$. One motivation for defining complex numbers is that it precisely solves this problem: every degree n polynomial $a_n x^n + \cdots a_1 x + a_0 = 0$ has exactly n *complex* solutions in general. However, it will take us some time before we can prove this statement.

A.1 Basic definitions and properties

Definition A.2. A field is a set $\mathbb{F}$ and operations $(+, \cdot)$ such that for any $a, b, c \in \mathbb{F}$,

$$a + (b + c) = (a + b) + c, \qquad a \cdot (b \cdot c) = (a \cdot b) \cdot c, \qquad a + b = b + a, \qquad a \cdot b = b \cdot a,$$
$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

Furthermore, there are unique elements $0, 1 \in \mathbb{F}$ such that

$$a + 0 = a, \qquad\qquad a \cdot 1 = a.$$

Finally, for every $a \in \mathbb{F}$ and $b \in \mathbb{F}$ with $b \neq 0$, there are unique elements $(-a)$ and b^{-1} such that

$$a + (-a) = 0, \qquad\qquad b \cdot b^{-1} = 1.$$

Examples of fields include the real numbers $\mathbb{R}$ and the rational numbers $\mathbb{Q} = \left\{ \frac{m}{n} \,|\, m, n \in \mathbb{Z},\, n \neq 0 \right\}$. If $p \in \mathbb{N}$ is a prime number, then $\mathbb{Z}_p = \{0, 1, \ldots, p - 1\}$ is a field with addition and multiplication given modulo p.

Definition/Theorem A.3. *There exists a field $\mathbb{C} = \mathbb{R} \oplus i\mathbb{R}$, with $\pm i := \pm\sqrt{-1}$ the solutions to the polynomial $x^2 + 1 = 0$.*

We typically write a generic complex number $z = x + iy \in \mathbb{C}$ with $x, y \in \mathbb{R}$. In particular, we define $x = \mathrm{Re}(z)$, $y = \mathrm{Im}(z) \in \mathbb{R}$.

Addition and multiplication are defined as

$$\begin{aligned}
&(a + ib) + (c + id) = (a + c) + i(b + d), \\
&(a + ib)(c + id) = ac + i(ad + bc) + i^2 bd = (ac - bd) + i(ad + bc),
\end{aligned}$$

where we have used that $i^2 = -1$, which we take as a definition. Note that $(-i)^2 = (-1)^2 i^2 = -1$ as well.

The additive identity for $\mathbb{C}$ is $0 = 0 + i0$ and the multiplicative identity is $1 = 1 + 0i$.

We see that a complex number $z = x + iy = 0$ if and only if $x^2 + y^2 = 0$. If $z = x + iy \neq 0$, then a simple check will show that

$$\frac{1}{x^2 + y^2}(x + iy)(x - iy) = \frac{x^2 + y^2}{x^2 + y^2} = 1, \quad \Longrightarrow \quad z^{-1} = \frac{1}{x^2 + y^2}(x - iy) = \frac{x}{x^2 + y^2} + i\frac{-y}{x^2 + y^2}.$$

Checking the other properties to show that $\mathbb{C}$ is a field are straightforward computations.

Complex numbers can be written as an ordered pair of real numbers $x + iy \sim (x, y)$, where one can define

$$(a, b) + (c, d) = (a + b, c + d), \qquad (a, b) \cdot (c, d) = (ac - bd, ad + bc),$$

though generally we use the convention $z = x + iy \in \mathbb{C}$.

Example A.4. The inverse $i^{-1} = -i$, indeed $i(-i) = -i^2 = -(-1) = 1$.

Remark A.5. We can also think of $\mathbb{C}$ as a vector space over the field $\mathbb{C}$. Namely, we define scalar multiplication $\alpha \cdot z = \alpha z$ for $\alpha, z \in \mathbb{C}$. As a complex vector space, $\mathbb{C}$ is one-dimensional and any non-zero complex number is a basis for $\mathbb{C}$.

Analogously to real numbers, we can take powers z^n for $n \in \mathbb{Z}$ as follows,

$$z^n = \begin{cases} \overbrace{z \cdot z \cdot \ldots \cdot z}^{n \text{ times}}, & n > 0, \\ 1, & n = 0, \\ (z^{-1})^{-n}, & n < 0 \text{ and } z \neq 0. \end{cases}$$

Like real numbers, we can also use the binomial formula/expansion to take powers of sums: if $z, w \in \mathbb{C}$ and $n > 0$ then

$$(z+w)^n = \sum_{j=0}^{n} \binom{n}{j} z^j w^{n-j}, \qquad \binom{n}{0} = 1, \quad \binom{n}{j} = \frac{n(n-1)\cdots(n-j+1)}{j!}. \tag{A.1}$$

Definition A.6 (Complex conjugate)**.** *The complex conjugate is a map $\mathbb{C} \ni z \mapsto \overline{z} \in \mathbb{C}$, $\overline{(x+iy)} = x - iy$.*

We collect some basic properties of the complex conjugate, which are all simple checks to prove.

Lemma A.7. *For any $z, z_1, z_2 \in \mathbb{C}$,*

$$\overline{\overline{z}} = z, \qquad \overline{z_1 \pm z_2} = \overline{z_1} \pm \overline{z_2}, \qquad \overline{z_1 z_2} = \overline{z_1}\,\overline{z_2},$$
$$\mathrm{Re}(z) = \frac{1}{2}(z + \overline{z}), \qquad \mathrm{Im}(z) = \frac{1}{2i}(z - \overline{z}), \qquad z\overline{z} \in \mathbb{R}.$$

Because $\mathbb{C} = \mathbb{R} \oplus i\mathbb{R}$, we can think of $\mathbb{R}$ as a subfield of $\mathbb{C}$ with the embedding $x \mapsto x + i0$. For any $z \in \mathbb{C}$, we have that

$$z \in \mathbb{R} \iff z = \overline{z}, \qquad z \in i\mathbb{R} \iff z = -\overline{z}$$

We also equip $\mathbb{C}$ with a norm/modulus/absolute value,

$$|z| = \sqrt{z\overline{z}} = \sqrt{x^2 + y^2}, \qquad z = 0 \iff |z| = 0.$$

One can check, for example, that $|z_1 + z_2| \leq |z_1| + |z_2|$. The complex conjugate and norm also give us the nice formula

$$z^{-1} = \frac{\overline{z}}{|z|^2}, \quad z \in \mathbb{C} \setminus \{0\}.$$

Examples A.8. The number $z = \dfrac{4-3i}{2+i}$ is a complex number, though it is not immediately obvious how to write this as $z = x + iy$. We can simplify z by some algebraic manipulation,

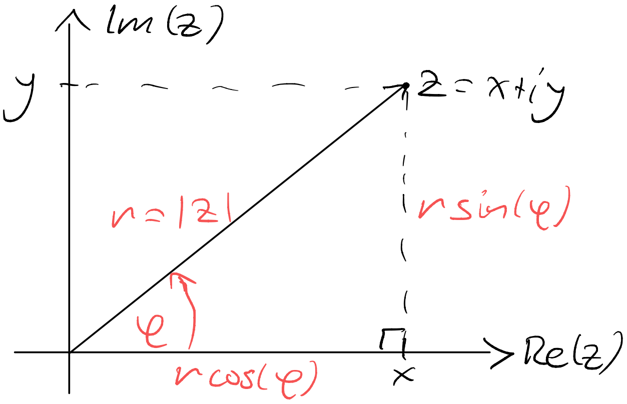
$$\frac{4-3i}{2+i}\frac{2-i}{2-i} = \frac{8+4i-6i+3}{4+1} = \frac{11-2i}{5} = \frac{11}{5} - i\frac{2}{5}.$$

Using what we know about i, we can compute higher powers easily,

$$(-i)^{637} = (-i)(-i)^{636} = (-i)(-1)^{318} = -i.$$

A.2 Polar form and the complex plane

Complex numbers $z = x + iy \in \mathbb{C}$ are built from a pair of real numbers, i.e. an element $(x, y) \in \mathbb{R}^2$. So it is natural to consider them as geometrically on a two-dimensional plane.



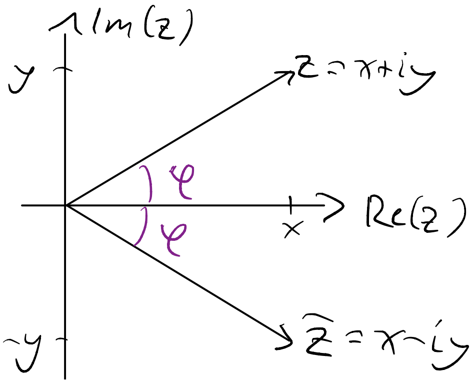
Definition A.9. Given $z \in \mathbb{C}$ and $z \neq 0$, z has the polar form,

$$z = r\big(\cos(\varphi) + i\sin(\varphi)\big), \qquad r = |z|, \qquad \varphi = \text{angle with real axis} = \arg(z).$$

Examples A.10.

$$\begin{aligned} 1 + i &= \sqrt{2}\big(\cos(\tfrac{\pi}{4}) + i\sin(\tfrac{\pi}{4})\big), \\ 6 - 3i &= \sqrt{45}\big(\cos(-\tfrac{\pi}{6}) + i\sin(-\tfrac{\pi}{6})\big) = \sqrt{45}\big(\cos(\tfrac{\pi}{6}) - i\sin(\tfrac{\pi}{6})\big) \end{aligned}$$

More generally, the complex conjugate $z \mapsto \overline{z}$ will give a reflection of z along the real axis,



The polar decomposition of z is *not* unique. Indeed,

$$r(\cos(\varphi) + i\sin(\varphi)) = r'(\cos(\varphi') + i\sin(\varphi')) \iff r = r' \text{ and } \varphi - \varphi' = 2\pi k,\ k \in \mathbb{Z}.$$

This means the argument $\arg(z)$ is not well-defined in general and takes multiple values. We define the Principal Value of the argument as the angle $\operatorname{Arg}(z) \in (-\pi, \pi]$ (note the capital 'A'), which *is* unique.

Proposition A.11. *Given* $z_1 = r_1(\cos(\theta_1) + i\sin(\theta_1))$, $z_2 = r_2(\cos(\theta_2) + i\sin(\theta_2))$,

$$z_1 z_2 = r_1 r_2\big(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\big).$$

> *That is, the absolute values are multiplied and the arguments are added.*

The statement is proved by a computation that uses the addition formulas for sin and cos. We leave the details as an exercise. The key point is that, when taking products, the absolute values multiply and the arguments add. This is particularly useful for taking powers, where

$$\big(r(\cos(\theta) + i\sin(\theta))\big)^n = r^n(\cos(n\theta) + i\sin(n\theta)).$$

Definition A.12. We call $\zeta \in \mathbb{C}$ an nth root of z if $\zeta^n = z$. In particular, if $\zeta^n = 1$, we call ζ an nth root of unity.

Theorem A.13. *For each $n \in \mathbb{N}$ there are exactly n distinct nth roots of unity.*
More generally, if $z \in \mathbb{C} \setminus \{0\}$ and $n \in \mathbb{N}$, there are exactly n different nth roots of z, $\{\zeta_j\}_{j=0}^{n-1}$, $\zeta_j^n = z$ for all $j \in \{0, 1, \ldots, n-1\}$.

Proof (Proof sketch). We will outline the proof of the first statement. Suppose that $\zeta^n = 1$, i.e. ζ is a root of unity. Using Proposition A.11, we know that

$$1 = |1| = |\zeta^n| = |\zeta|^n, \quad \Longrightarrow \quad |\zeta| = 1^{1/n} = 1.$$

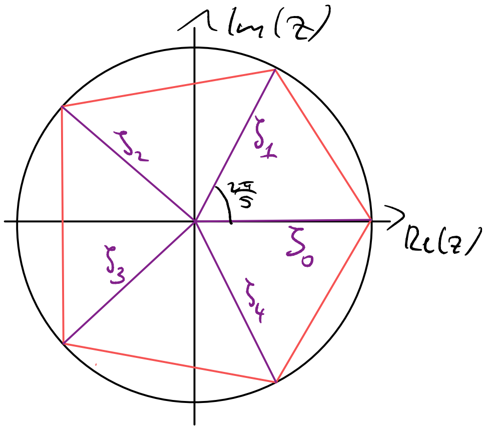
Let $\varphi = \arg(\zeta)$. We again use Proposition A.11 and note that

$$\arg(1) = \arg(\zeta^n) = n\arg(\zeta) = n\varphi, \quad \arg(1) = 2\pi k,\ k \in \mathbb{Z}, \quad \Longrightarrow \quad \varphi = \frac{2\pi k}{n}$$

By taking $k = 0, 1, \ldots, n-1$, we obtain n distinct solutions for φ. Putting this together gives us all roots of unity

$$\{\zeta_j\}_{j=0}^{n-1}, \qquad \zeta_j = \cos\left(\frac{2\pi j}{n}\right) + i\sin\left(\frac{2\pi j}{n}\right).$$ □

We remark that we can write $\zeta_j = (\zeta_1)^j$ with ζ_1 the first non-trivial root of unity. The roots of unity are spaced evenly around the unit circle. For example, the solutions to $z^5 = 1$ are sketched below.



A.3 Sequences and series

To talk about analysis and continuity on $\mathbb{C}$, we need to talk about sequences and what it means to converge. Because $\mathbb{C} = \mathbb{R} \oplus i\mathbb{R}$, we can use results that we already know about $\mathbb{R}$ and $\mathbb{R}^2$ to understand properties of $\mathbb{C}$.

Definition A.14. 1. We say that a sequence $\{z_n\}_{n\geq 0} \subset \mathbb{C}$ converges to an element $z \in \mathbb{C}$ if for all $\epsilon > 0$ there is some $N \in \mathbb{N}$ such that

$$|z_n - z| < \epsilon, \quad \text{for all } n \geq N.$$

2. We say that a sequence $\{z_n\}_{n\geq 0} \subset \mathbb{C}$ is a Cauchy sequence if for all $\epsilon > 0$ there is some $N \in \mathbb{N}$ such that

$$|z_n - z_m| < \epsilon, \quad \text{for all } n,\, m \geq N.$$

Definition A.15. Let V be a vector space over a field $\mathbb{F}$ with a norm $\|\cdot\|$. We say that V is complete if any Cauchy sequence $\{v_n\}_{n\geq 0}$ converges to an element $v \in V$.

The real numbers $\mathbb{R}$ and rational numbers $\mathbb{Q}$ are vector spaces over the fields $\mathbb{R}$ and $\mathbb{Q}$ respectively. As vector space $\mathbb{R}$ is complete, the rational numbers $\mathbb{Q}$ are not. Indeed, we can take sequences of rational numbers that converge to e or π.

The notion of completeness can be extended to metric spaces, a set X with a notion of distance $d(x, y) \geq 0$ for all $x, y \in X$.[4] An open interval $(a, b) \subset \mathbb{R}$ with $d(x, y) = |x - y|$ is a metric space, but it is not complete. For example, consider the sequence $a_n = \dfrac{1}{n}$ in the set $(0, 1)$. The sequence a_n is Cauchy, but it converges to 0, which is not an element of the set $(0, 1)$.

Theorem A.16. *The complex numbers $\mathbb{C}$ are complete.*

To show the above theorem, we use the following result as well as the fact that $\mathbb{R}$ is complete.

Lemma A.17. *A sequence $z_n \to z \in \mathbb{C}$ if and only if $\mathrm{Re}(z_n) \to \mathrm{Re}(z) \in \mathbb{R}$ and $\mathrm{Im}(z_n) \to \mathrm{Im}(z) \in \mathbb{R}$.*

We also note that if $z_n \to z \in \mathbb{C}$, then $|z_n| \to |z| \in \mathbb{R}$ and $\overline{z_n} \to \overline{z} \in \mathbb{C}$.
Furthermore, $\mathbb{C}$ is complete in this norm (every Cauchy sequence converges).

Definition A.18. The (infinite) series is defined

$$\sum_{j=0}^{\infty} z_j = \lim_{n\to\infty} S_n, \qquad S_n = \sum_{j=0}^{n} z_j.$$

We say that a series is absolutely convergent if

$$\sum_{j=0}^{\infty} |z_j| < \infty.$$

As in the case of real analysis,

$$\sum_{j=0}^{\infty} |z_j| < \infty \quad \Longrightarrow \quad \sum_{j=0}^{\infty} z_j \ \text{ well defined in } \mathbb{C}.$$

[4] A more precise definition of metric space is beyond the scope of this course.

Example A.19. The *geometric series* converges for all $|z| < 1$,

$$\sum_{n=0}^{\infty} z^n = 1 + z + z^2 + \cdots = \frac{1}{1-z}, \quad |z| < 1.$$

Lemma A.20 (Cauchy's multiplication Theorem). *Suppose that*

$$\sum_{n=0}^{\infty} a_n, \quad \text{and} \quad \sum_{n=0}^{\infty} b_n$$

are absolutely convergent. Then

$$\Big(\sum_{n=0}^{\infty} a_n\Big)\Big(\sum_{n=0}^{\infty} b_n\Big) = \sum_{j=0}^{\infty}\Big(\sum_{k=0}^{j} a_k b_{j-k}\Big)$$

and the series on the right-hand side is absolutely convergent.

Definition/Theorem A.21. *The series*

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}, \qquad \sin(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1}, \qquad \cos(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n}$$

are absolutely convergent for all $z \in \mathbb{C}$. Furthermore for all $z, w \in \mathbb{C}$,

$$\begin{aligned} &\exp(z+w) = \exp(z)\exp(w), && \exp(iz) = \cos(z) + i\sin(z) \\ &\cos(z) = \frac{1}{2}\big(\exp(iz) + \exp(-iz)\big), && \sin(z) = \frac{1}{2i}\big(\exp(iz) - \exp(-iz)\big). \end{aligned}$$

A comparison of the series formula for $\exp(z)$ with $\exp(x)$ with $x \in \mathbb{R}$ shows that $\exp(z)$ extends the real series expansion of $\exp(x)$ to the complex plane. This is similarly true for sin and cos.

We will omit the proof that the series for exp, sin and cos converge. Though let us check a few of the properties and leave the rest as an exercise.

We first show $\exp(z+w) = \exp(z)\exp(w)$, where we start with the right-hand side and apply Lemma A.20.

$$\begin{aligned} \exp(z)\exp(w) &= \sum_{n=0}^{\infty}\Big(\sum_{j=0}^{n} \frac{z^j w^{n-j}}{j!(n-j)!}\Big) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}\Big(\sum_{j=0}^{n} \binom{n}{j} z^j w^{n-j}\Big) \qquad \text{as } \frac{1}{j!(n-j)!} = \frac{1}{n!}\binom{n}{j} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}(z+w)^n \qquad \text{using the binomial expansion, Eq. (A.1)} \\ &= \exp(z+w). \end{aligned}$$

We similarly show the identity $\sin(z) = \frac{1}{2i}\big(\exp(iz) - \exp(-iz)\big)$,

$$\begin{aligned}\frac{1}{2i}\big(\exp(iz) - \exp(-iz)\big) &= \frac{1}{2i}\sum_{n=0}^{\infty}\Big(\frac{(iz)^n}{n!} - \frac{(-iz)^n}{n!}\Big)\\ &= \sum_{n=0}^{\infty}\frac{z^n i^{n-1}}{n!}\Big(\frac{1-(-1)^n}{2}\Big).\end{aligned}$$

The quantity $\big(\frac{1-(-1)^n}{2}\big)$ will vanish unless n is odd. We therefore let $n = 2m+1$ and note that $i^{n-1} = i^{2m+1-1} = i^{2m} = (-1)^m$. Therefore we can simplify the series to

$$\frac{1}{2i}\big(\exp(iz) - \exp(-iz)\big) = \sum_{m=0}^{\infty}\frac{(-1)^m z^{2m+1}}{(2m+1)!} = \sin(z).$$

A.4 The complex exponential and the logarithm

The expression $\exp(iz) = \cos(z) + i\sin(z)$ is particularly useful as it allows us to simplify the polar form of complex numbers,

$$z = |z|\big(\cos(\theta) + i\sin(\theta)\big) = |z|\exp(i\theta) = |z|\exp(i\theta + 2\pi ik), \quad k \in \mathbb{Z}.$$

For example, the unit circle in the complex plane $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$ can be neatly characterised as $\exp(i\theta)$ for $\theta \in (-\pi, \pi]$ (or $\theta \in \mathbb{R}$).

We will also write $e^z = \exp(z)$, though unless otherwise stated e^z will refer to the series expansion and not the complex power of the real number e (more on this below). As a bonus, we get the aesthetically pleasing equation

$$e^{i\pi} + 1 = 0.$$

Much like the polar form is not unique, the complex exponential is *not* injective,

$$e^z = e^w \quad \Longleftrightarrow \quad z - w = 2\pi ik,\ k \in \mathbb{Z}.$$

We see that the exponential is a periodic function, like sin and cos on the real axis. This is not suprising given that $e^{iz} = \cos(z) + i\sin(z)$. Note also that $e^z \neq 0$ for any $z \in \mathbb{C}$. So the exponential has range $\mathbb{C} \setminus \{0\}$.

Because $\mathbb{C} \ni z \mapsto e^z \in \mathbb{C}$ is not an injective function, the inverse (the logarithm) will not be uniquely specified. We can, however, restrict to a subregion.

Lemma A.22. *Let $S \subset \mathbb{C}$ denote the infinite strip*

$$S = \mathbb{R} \times i(-\pi, \pi] = \big\{x + iy \mid x \in \mathbb{R},\ y \in (-\pi, \pi]\big\}.$$

Then the map $S \ni w \mapsto e^w \in \mathbb{C} \setminus \{0\}$ is a bijection.

Proof. For $w = x + iy \in S$, $e^w = e^{x+iy} = e^x e^{iy}$. If $z \in \mathbb{C} \setminus \{0\}$, we can write z in polar form $z = |z|e^{i\theta}$ with $\theta \in (-\pi, \pi]$. We can therefore map w to z by taking $x = \ln(|z|)$ (the real natural logarithm) and $y = \theta$,

$$e^{\ln(|z|)+i\theta} = e^{\ln(|z|)}e^{i\theta} = |z|e^{i\theta} = z.$$

Therefore the map is surjective. Now suppose $w_1 = x_1 + iy_1$ and $w_2 = x_2 + iy_2$ are such that $e^{w_1} = e^{w_2}$, then $w_1 - w_2 = 0 + 2\pi ik$ and so $x_1 = x_2$ and $y_1 - y_2 = 2\pi k$ for some $k \in \mathbb{Z}$. But if $y_1, y_2 \in (-\pi, \pi]$, then this is only possible if $k = 0$. Thus $w_1 = w_2$ and exp is injective. □

Given some $z \in \mathbb{C} \setminus \{0\}$, the previous lemma tells us there is a unique element $w \in S \subset \mathbb{C}$ such that $e^w = z$. This allows us to unambiguously define the inverse/logarithm.

Definition A.23. Given $z \in \mathbb{C} \setminus \{0\}$ and $w \in S$ such that $\exp(w) = z$. We define the Principal branch of the logarithm $\operatorname{Log}(z) = w \in S$. The Principal branch of the logarithm is the unique function characterised by

$$\exp(\operatorname{Log}(z)) = z \quad \textit{and} \quad \operatorname{Im}(\operatorname{Log}(z)) \in (-\pi, \pi] \quad \text{for all } z \neq 0.$$

We first note that $\operatorname{Log}(x + i0) = \ln(x)$ the natural logarithm in $\mathbb{R}$. So Log extends our definition of the logarithm. Suppose $z = re^{i\varphi}$ with $\varphi \in (-\pi, \pi]$ and $r > 0$. We claim that $\operatorname{Log}\left(re^{i\varphi}\right) = \ln(r) + i\varphi \in S$. Because the Principal branch of the logarithm is unique, we just need to check that $\exp\left(\operatorname{Log}(re^{i\varphi})\right) = re^{i\varphi}$. Indeed,

$$\exp\left(\ln(r) + i\varphi\right) = e^{\ln(r)+i\varphi} = e^{\ln(r)}e^{i\varphi} = re^{i\varphi}.$$

Rephrasing the above argument, we have shown the following.

Theorem A.24. *For any $z \in \mathbb{C} \setminus \{0\}$,*

$$\operatorname{Log}(z) = \ln\left(|z|\right) + i \operatorname{Arg}(z)$$

with ln *the natural logarithm of a positive real number.*

Examples A.25.

$$\begin{aligned}\operatorname{Log}(-3i) &= \operatorname{Log}\left(3e^{-i\frac{\pi}{2}}\right) = \ln(3) - i\frac{\pi}{2}\\ \operatorname{Log}(4i - 4) &= \operatorname{Log}\left(\sqrt{32}e^{i\frac{3\pi}{4}}\right) = \ln(\sqrt{32}) + i\frac{3\pi}{4} = \ln(2^{5/2}) + i\frac{3\pi}{4} = \frac{5}{2}\ln(2) + i\frac{3\pi}{4}\end{aligned}$$

Exercise A.1. Show that for all $z_1, z_2 \in \mathbb{C} \setminus \{0\}$, $\operatorname{Log}(z_1 z_2) = \operatorname{Log}(z_1) + \operatorname{Log}(z_2) + 2\pi i k$ for some $k \in \mathbb{Z}$.

We would like to use the logarithm to define complex powers, $z^w = \exp(w \log(z))$, $z, w \in \mathbb{C}$, but this will take multiple values in general. Namely, for all $k \in \mathbb{Z}$,

$$\begin{aligned}&z = r\exp\left(i\theta + 2\pi i k\right) = \exp\left(\ln(r) + i(\theta + 2\pi k)\right)\\ &\quad \Longrightarrow \quad z^w = \exp\left(w(\ln(r) + i(\theta + 2\pi k))\right) = \exp\left(w(\ln(r) + i\theta)\right)\exp\left(2\pi i k w\right).\end{aligned}$$

Example A.26.

$$\begin{aligned}i^i &= \left(\exp(i\frac{\pi}{2} + 2\pi i k)\right)^i = \exp\left(i(i\frac{\pi}{2} + 2\pi i k)\right) = \exp\left(-\frac{\pi}{2} - 2\pi k\right)\\ &= \left\{e^{-\frac{\pi}{2}}, e^{-\frac{5\pi}{2}}, e^{\frac{3\pi}{2}}, e^{-\frac{9\pi}{2}}, e^{\frac{7\pi}{2}}, \ldots\right\}.\end{aligned}$$

It is perhaps surprising that i^i takes infinitely many values and that all of the possible values are real.

Let $z = re^{i\theta}$ with $\theta \in (-\pi, \pi]$. Let us look closer at the expression,

$$z^w = \exp\left(w(\ln(r) + i\theta)\right)\exp\left(2\pi i k w\right) = \exp\left(w \operatorname{Log}(z)\right)\exp\left(2\pi i k w\right).$$

We see the following behaviour:

(i) If $w \in \mathbb{Z}$, then $\exp(2\pi ikw) = \exp(2\pi ik') = 1$ and so z^w is unique,

(ii) If $w = \frac{m}{n} \in \mathbb{Q}$, then $\exp(2\pi ikw) = \exp\left(\frac{2\pi ikm}{n}\right)$, which will take finitely many values as we change $k \in \mathbb{Z}$. Hence z^w has finitely many values,

(iii) If $w \notin \mathbb{Q}$, then $\exp(2\pi ikw)$ has infinitely many values for $k \in \mathbb{Z}$.

References

[1] J. Cortese. *Relative entropy and single qubit Holevo-Schumacher-Westmoreland channel capacity.* arXiv:quant-ph/0207128 (2002).

[2] D. Johnston and B. J. Schroers. *Quantum Mechanics and Quantum Computing: an Introduction.* Lecture notes.

[3] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information, 10th Anniversary Edition.* Cambridge University Press, 2010.

[4] W. Scherer. *Mathematics of Quantum Computing, An Introduction.* Springer Nature, 2019.

[5] A. Uhlmann. *Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory.* Comm. Math. Phys., **54**, no. 1, 21–32, 1977.

[6] J. Yin, Y. Cao, Y.-H. Li, et al., *Satellite-based entanglement distribution over 1200 kilometers.* Science, **356** (2017), 1140–1144. Available at `http://science.sciencemag.org/content/356/6343/1140.full` and `arXiv:1707.01339`.