Course notes to accompany QuSTEAM Course 3: Mathematical methods for quantum information science

Kyle Kawagoe, Chaobin Liu, and David Penneys

August 26, 2026

Abstract

The purpose of *QuSTEAM Course 3: Mathematical methods for quantum information science* is to simultaneously teach quantum information and linear algebra, aimed at university students in year 2. This course was developed through by the QuSTEAM Course 3 Development Team 2022-23, consisting of Bennett Brown, Russell Ceballos, Eric Chitambar, Larry Hurtubise, Kyle Kawagoe (course co-lead), Ariel King, Eric Kuo, Chaobin Liu, David Penneys (course co-lead), Peter Plourde, Sean Sanford, Yudan Wang, and Lee Wayand. This manuscript consists of course notes to accompany QuSTEAM Course 3 produced in 2022-23 during the first 3 iterations of this course at The Ohio State University, taught by David Penneys (Au22), Sean Sanford (Sp23), and Kyle Kawagoe (Au23). These notes are divided into 3 modules, mirroring the 3 modules of the course: (1) Superposition: 1 qubit, (2) Entanglement: 2+ qubits, and (3) Applications.

Preface

The purpose of *QuSTEAM Course 3: Mathematical methods for quantum information science* is to simultaneously teach quantum information and linear algebra, aimed at university students in year 2. This course was developed through by the QuSTEAM Course 3 Development Team 2022-23, consisting of Bennett Brown, Russell Ceballos, Eric Chitambar, Larry Hurtubise, Kyle Kawagoe (course co-lead), Ariel King, Eric Kuo, Chaobin Liu, David Penneys (course co-lead), Peter Plourde, Sean Sanford, Yudan Wang, and Lee Wayand.

This manuscript consists of course notes produced in 2022-23 during the first 3 iterations of this course at The Ohio State University, taught by David Penneys (Au22), Sean Sanford (Sp23), and Kyle Kawagoe (Au23). They were heavily influenced by materials produced by and the weekly meetings of the QuSTEAM Course 3 Development Team 2022-23. We would like to thank Bennett Brown and Sean Sanford for helping fix typos in this manuscript.

These notes are divided into 3 modules, mirroring the 3 modules of the course:

(1) Superposition: 1 qubit

(2) Entanglement: 2+ qubits

(3) Applications.

The first module is centered around the Stern-Gerlach experiment, the second module is centered around the EPR paradox, the CHSH inequality, and the GHSZ paradox, and the third module is centered around applications of the graphical calculus to quantum information protocols. The idea to divide the course into these 3 modules was synthesized at the 2022 QuSTEAM Conference at The Ohio State University. We would especially like to thank Russell Ceballos, Dan Gauthier, and Ezekiel Johnston-Halperin for contributing to this division. We would also like to thank Giovanni Ferrer for the formatting style using the `mdframed` package.

This manuscript also includes a short appendix on supplemental linear algebra needed for some material in the second module. We try to present results in this section as algorithms whenever possible, which can be implemented on computer algebra software.

Acknowledgements

QuSTEAM was made possible via the NSF QuSTEAM C-ACCEL grant 2040581. The authors would like to thank the QuSTEAM Course 3 Development Team 2022-23 along with Dan Gauthier and Ezekiel Johnston-Halperin.

Contents

1 Superposition: 1 qubit

1.1 Introduction: the Stern-Gerlach experiment

The Stern-Gerlach experiment took place in 1922 [GS22]; in this experiment, silver atoms were sent through an inhomogeneous magnetic field. Behind the magnetic field, a measurement device was set up to record the position of the silver atom. Rather than obtaining some kind of Gaussian distribution (bell shaped curve), the atoms landed at two discrete bands, which corresponds to whether the single unpaired electron the 5s orbital of the silver atom is *spin up* or *spin down*. This was the first experiment to establish that electrons have a property called *spin* which is *quantized*, i.e., can take one of two values, not a continuum of values. (Electrons are *fermions* whose value for spin is a half-integer, but this is more information than we need to know right now.)

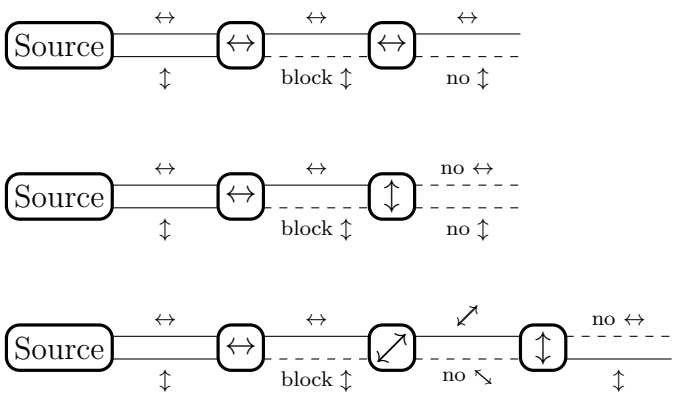
The spin of an electron can be measured in any of the x, y, or z axes, leading to 2 discrete answers each time. The truly remarkable property of spin is that measuring in the z-direction *affects* a subsequent measurement in the x-direction. If we measure in the z-direction getting 'spin up,' then there is a 50% chance that the measurement in the x-direction is 'spin up' and a 50% chance that it is 'spin down.' If we measure in the z-direction again, we may not get the same measurement that we obtained the first time! If we measure the z-direction again, we observe that we also have a 50% chance of obtaining 'spin up' and a 50% chance of obtaining 'spin down'!

This is truly remarkable, because it contradicts our typical world view. For example, let's consider a thought experiment where we have a drawer full of socks which are either red or blue, and each sock is a left sock or right sock. (This latter condition is not as ridiculous as it sounds if you are a runner; some high end running socks are designed independently for left and right feet!) You choose a sock at random from the drawer, and you measure the color: red. Then you measure the chirality: left. You would then know that the sock is a red left sock, correct? Surely looking at whether a sock is a left or right sock could not change the color of the sock, right? Well, not if color is the spin measurement in the z-direction and chirality is spin measurement in the x-direction!

We do not need silver atoms nor a magnetic field to verify this quantum phenomenon. A similar experiment can be carried out with light polarization filters. Light is made of *photons* which also have spin. (The spin values of photons are integers rather than half-integers as they are *bosons* rather than *fermions*, but this is beyond what we need to know right now.) The state of light polarized in the z-direction can be $|\updownarrow\rangle$ or $|\leftrightarrow\rangle$. If we pass $|\leftrightarrow\rangle$-polarized light through a $\leftrightarrow$-polarization filter, it will pass through unchanged, but if we pass $|\leftrightarrow\rangle$-polarized light through a $\updownarrow$-polarization filter, it will be completely blocked. Counter-intuitively, if we insert an x-direction '$\nearrow$-polarization' filter between two $\leftrightarrow$ and $\updownarrow$-polarization filters, we surprisingly see that some $\updownarrow$-polarized light passes through! This is truly remarkable, as all polarization filters do is block light (which has particle properties) from passing through. How can it be that *adding* a filter *increases* the total light passing through?

These outcomes are summarized in the cartoon on the next page. The goal of this first

module is to understand the mathematics behind this phenomenon. (We will not at all understand *why* the universe works this way – this will remain a mystery!)



1.2 The Hilbert space $\mathbb{R}^2$

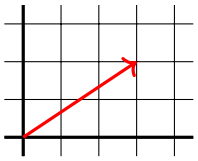
The set $\mathbb{R}^2$ is the collection of all ordered pairs of real numbers:

$$\mathbb{R}^2 = \left\{ \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\text{candidates}} \middle| \underbrace{x, y \in \mathbb{R}}_{\text{requirements}} \right\}.$$

Above, we have used *set notation*; the part on the left is the collection of all *candidates* that belong in this set, in this case, pairs $\begin{bmatrix} x \\ y \end{bmatrix}$. The part on the right is the *requirements*, which stipulate which things on the left actually belong in the set; in this case, we requre that both x and y are *elements of* (denoted $\in$) the *real numbers* (denoted $\mathbb{R}$).

You can think of an element $(x, y) \in \mathbb{R}^2$ as an ordered pair, but here we will think of it as a *vector* in the plane. The x coordinate tells us the x-position on the x-axis of the head of the vector, and the y-coordinate tells us the y-position on the y-axis of the head of the vector:

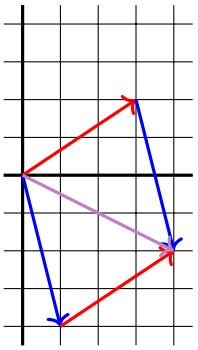
the vector $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$:



We add vectors by stacking them *head to tail*, which is the same as adding them *componen-*

twise:

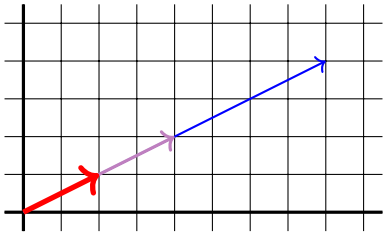
$$\begin{bmatrix}3\\2\end{bmatrix} + \begin{bmatrix}1\\-4\end{bmatrix} = \begin{bmatrix}4\\-2\end{bmatrix}$$



Since addition is *commutative*, you can stack the vectors in either order, and the sum is the point on the other side of the parallelogram. Addition is also *associative*, meaning you can add three vectors in different orders and obtain the same result. There is a *zero vector* $\mathbf{0} := \begin{bmatrix}0\\0\end{bmatrix}$ which leaves every other vector unchanged under addition. Also, given any vector, there is another vector you can add to it to get back to $\mathbf{0}$.

You can also *scale* vectors by multiplying by a constant, for example:

$$2 \cdot \begin{bmatrix}4\\2\end{bmatrix} = \begin{bmatrix}8\\4\end{bmatrix} \qquad \frac{1}{2} \cdot \begin{bmatrix}4\\2\end{bmatrix} = \begin{bmatrix}2\\1\end{bmatrix}$$



Scaling by 1 does nothing to the vector, and scaling by r and s is the same as scaling by rs. Finally, scalar multiplication and addition are compatible and satisfy *distributive properties*.

All this structure is what is needed to say that $\mathbb{R}^2$ is a *vector space*.

Definition 1.2.1 — A *vector space* over $\mathbb{R}$ is a set V together with an addition $+ : V \times V \to V$ and a scalar multiplication $\cdot : \mathbb{R} \times V \to V$ satisfying:

(V1) $u + v = v + u$ for all $u, v \in V$,

(V2) $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$,

(V3) there is an element $0 \in V$ such that $0 + v = v = v + 0$ for all $v \in V$,

(V4) for each $v \in V$, there is a $-v \in V$ such that $v + (-v) = 0 = -v + v$,

(V5) $(rs)v = r(sv)$ for all $r, s \in \mathbb{R}$ and $v \in V$,

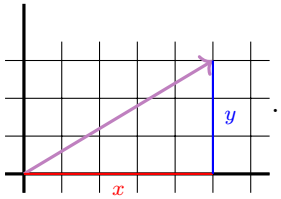
(V6) $1v = v$ for all $v \in V$,

(V7) $r(u + v) = ru + rv$ for all $r \in \mathbb{R}$ and $u, v \in V$, and

(V8) $(r+s)v = rv + sv$ for all $r, s \in \mathbb{R}$ and $v \in V$.

Vectors in $\mathbb{R}^2$ have a *magnitude/length* given by the Pythagorean Theorem:

$$\left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| = \sqrt{x^2 + y^2}$$



Given two vectors, we can also compute the *angle* between them. To do so, we use the *inner product*, which encodes all information about lengths of vectors and angles between vectors.

Definition 1.2.2 — The *inner product* on $\mathbb{R}^2$ is given by $\left\langle \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \middle| \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \right\rangle := x_1x_2 + y_1y_2$.

Notation 1.2.3 (Dirac) **—** When possible, we will use the notation $|v\rangle$ for a vector in $\mathbb{R}^2$. This allows us to write $\langle u|$ for the function $\mathbb{R}^2 \to \mathbb{R}$ given by $|v\rangle \mapsto \langle u|v\rangle$.

The vectors $|v\rangle$ are called *kets* and the maps $\langle u|$ are called *bras*. The action of a bra $\langle u|$ on a ket $|v\rangle$ produces the *bracket* $\langle u|v\rangle$.

Axioms 1.2.4 — Using the notation above, we can give the formal axioms of an abstract inner product $\langle \cdot | \cdot \rangle$ on a real vector space V:

(IP1) $\langle u|v+w\rangle = \langle u|v\rangle + \langle u|w\rangle$ for all $|u\rangle, |v\rangle, |w\rangle \in V$,

(IP2) $\langle u+v|w\rangle = \langle u|w\rangle + \langle v|w\rangle$ for all $|u\rangle, |v\rangle, |w\rangle \in V$,

(IP3) $\langle u|v\rangle = \langle v|u\rangle$ for all $|u\rangle, |v\rangle \in V$,

(IP4) $\langle v|v\rangle \geq 0$, with equality if and only if $|v\rangle = 0$.

Remark 1.2.5 — When we discuss inner products over *complex* vector spaces in §1.5 below, (IP3) is replaced with:

(IP3) $\langle u|v\rangle = \overline{\langle v|u\rangle}$ for all $|u\rangle, |v\rangle \in V$, where $\overline{\cdot}$ denotes the complex conjugate operation in $\mathbb{C}$.

We saw in Notation 1.2.3 above that we have bras $\langle u| : \mathbb{R}^2 \to \mathbb{R}$ given by $|v\rangle \mapsto \langle u|v\rangle$. The space of bras is called the *dual space* of $\mathbb{R}^2$, and sometimes denoted $(\mathbb{R}^2)^*$. Given a

vector/ket $|u\rangle = \begin{bmatrix} w \\ x \end{bmatrix} \in \mathbb{R}^2$, we define the bra $\langle u| := \begin{bmatrix} w & x \end{bmatrix}$, which is the *transpose*, that is, we switch the rows and columns. We then get the bracket $\langle u|v\rangle$ by *matrix multiplication*:

$$|v\rangle = \begin{bmatrix} y \\ z \end{bmatrix} \xmapsto{|u\rangle} \langle u|v\rangle = \begin{bmatrix} w & x \end{bmatrix} \cdot \begin{bmatrix} y \\ z \end{bmatrix} := wy + xz.$$

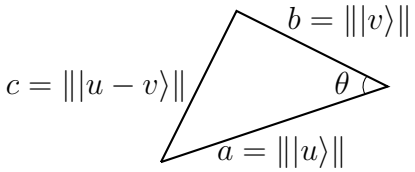
That is, we multiply the 11-entry of $\langle u|$ with the 11 entry of $|v\rangle$ and we add this to the product of the 12-entry of $\langle u|$ with the 21-entry of $|v\rangle$. Here, the term 'ij-entry' means the number in the i-th row and j-th column.

Exercise 1.2.6 — Show that the length of $|v\rangle \in \mathbb{R}^2$ is exactly $\sqrt{\langle v|v\rangle}$.

Proposition 1.2.7 — The angle between two non-zero vectors is determined by the formula

$$\cos(\theta) = \frac{\langle u|v\rangle}{\||u\rangle\| \cdot \||v\rangle\|}.$$

Proof. This follows from the *Law of cosines* which states that given a triangle with side lengths a, b, c and angle θ as pictured below,



we have the formula

$$c = \sqrt{a^2 + b^2 - 2ab\cos(\theta)}. \tag{1.2.8}$$

Now if the line segment of length a is $|u\rangle$ and the line segment of length b is $|v\rangle$, then the line segment of length c is $|u\rangle - |v\rangle$, whose length is given by

$$c = \||u\rangle - |v\rangle\| = \sqrt{\langle u|u\rangle + \langle v|v\rangle - 2\langle u|v\rangle}. \tag{1.2.9}$$

Now equating the right hand sides of (1.2.8) and (1.2.9), squaring, and canceling like terms, we have

$$2\||u\rangle\| \cdot \||v\rangle\| \cos(\theta) = 2\langle u|v\rangle.$$

The result follows after rearranging. ∎

There are several particular vectors that get special names in quantum information:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad |+\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad |-\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Observe that $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$; this is where the name $+/-$ for the ket comes from.

Definition 1.2.10 — An *orthonormal basis* (ONB) for $\mathbb{R}^2$ is a pair of vectors $\{|e_1\rangle, |e_2\rangle\}$ such that

$$\langle e_i|e_j\rangle = \delta_{i=j} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

Exercise 1.2.11 — Prove that $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$ are two ONBs.

ONBs have a particular significance: given an ONB $\{|e_1\rangle, |e_2\rangle\}$, every vector $|v\rangle \in \mathbb{R}^2$ can be written *uniquely* as

$$|v\rangle = a|e_1\rangle + b|e_2\rangle \tag{1.2.12}$$

for some $a, b \in \mathbb{R}$.

Exercise 1.2.13 — Show that $a = \langle e_1|v\rangle$ and $b = \langle e_2|v\rangle$.
Hint: apply $\langle e_1|$ to the left and right hand sides of (1.2.12) above.

We end this section with the *Cauchy-Schwarz inequality*, the most important inequality in the theory of inner product spaces.

Theorem 1.2.14 (Cauchy-Schwarz) **—** For every $|u\rangle, |v\rangle \in \mathbb{R}^2$,

$$|\langle u|v\rangle|^2 \leq \langle u|u\rangle \cdot \langle v|v\rangle.$$

Proof. We may assume $|u\rangle, |v\rangle$ are both non-zero. For $t \in \mathbb{R}$, consider the vector $|u - tv\rangle = |u\rangle - t|v\rangle$. Since its length is non-negative, we have

$$0 \leq \langle u - tv|u - tv\rangle = \langle u|u\rangle - 2t\langle u|v\rangle + t^2\langle v|v\rangle.$$

This is a quadratic in t which achieves its minimum at $t = \frac{2\langle u|v\rangle}{2\langle v|v\rangle}$. Plugging this value of t into the above quadratic yields

$$0 \leq \langle u|u\rangle - 2\frac{|\langle u|v\rangle|^2}{\langle v|v\rangle} + \frac{|\langle u|v\rangle|^2}{\langle v|v\rangle^2}\langle v|v\rangle = \langle u|u\rangle - \frac{|\langle u|v\rangle|^2}{\langle v|v\rangle}.$$

Rearranging yields the result. ∎

Remark 1.2.15 — There was nothing special about $\mathbb{R}^2$ in this theorem; the same proof works for $\mathbb{R}^n$ as well. We will see in §1.5 below that there is a complex version that works in $\mathbb{C}^n$.

1.3 $M_2(\mathbb{R})$: Operators on $\mathbb{R}^2$

A *matrix* is an array of numbers. An $m \times n$ matrix has m rows and n columns, and the ij-th entry is the number in the i-th row and j-th column. This section is all about 2×2 matrices, i.e., those of the form

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The collection of such matrices is called $M_2(\mathbb{R})$.

The 2×2 matrices with real entries act as *linear operators* on $\mathbb{R}^2$. We can multiply a 2×2 matrix A times a vector $|v\rangle$ by

$$A|v\rangle = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} := \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} \qquad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad |v\rangle = \begin{bmatrix} x \\ y \end{bmatrix}.$$

We can think of multiplying A times $|v\rangle$ another way. Let $\langle A_1|$ be the first row of A and $\langle A_2|$ be the second row of A. Then

$$A|v\rangle = \begin{bmatrix} \langle A_1| \\ \langle A_2| \end{bmatrix} |v\rangle = \begin{bmatrix} \langle A_1|v\rangle \\ \langle A_2|v\rangle \end{bmatrix} \qquad A = \begin{bmatrix} \langle A_1| \\ \langle A_2| \end{bmatrix}.$$

We can also multiply two 2×2 matrices by multiplying each column of the second matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} |u\rangle & |v\rangle \end{bmatrix} := \begin{bmatrix} A|u\rangle & A|v\rangle \end{bmatrix} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}.$$

Exercise 1.3.1 — Show that for 2×2 matrices A, B and $|v\rangle \in \mathbb{R}^2$, $A(B|v\rangle) = (AB)|v\rangle$.

Trick 1.3.2 — Suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then

$$A|0\rangle = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ c \end{bmatrix} \qquad \text{and} \qquad A|1\rangle = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}.$$

That is, $A|0\rangle$ is the first column of A, and $A|1\rangle$ is the second column of A.

Exercise 1.3.3 — Compute $\langle 0|A$ and $\langle 1|A$ in terms of the entries of A.

Example 1.3.4 — The most basic example of an operator in $M_2(\mathbb{R})$ is the *zero matrix*

$$0_2 := \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

For every $|v\rangle \in \mathbb{R}^2$, $0_2|v\rangle = \mathbf{0}$, the zero vector.

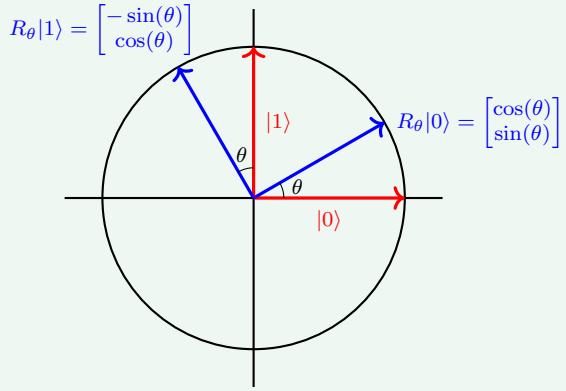
Warning — The zero vector $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ should never be confused with the 0 ket $|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Example 1.3.5 — The most basic example of an operator is the *identity matrix*

$$I_2 := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Observe that $I_2|v\rangle = |v\rangle$ for all $|v\rangle \in \mathbb{R}^2$. Using Trick 1.3.2, we can calculate that $AI_2 = A$ for every 2×2 matrix A, and using Exercise 1.3.3, we see $I_2 A = A$ for every 2×2 matrix A.

Example 1.3.6 — Given $\theta \in [0, 2\pi)$, the matrix $R_\theta := \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ rotates $\mathbb{R}^2$ counterclockwise by θ.



Example 1.3.7 — Given two vectors $|u\rangle, |v\rangle \in \mathbb{R}^2$, we get the *rank one* operator or

ket-bra $|u\rangle\langle v|$ which acts on a vector $|w\rangle \in \mathbb{R}^2$ by

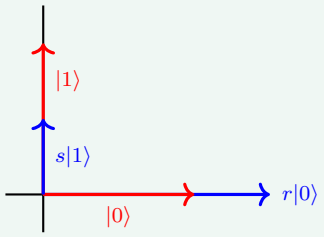
$$(|u\rangle\langle v|)|w\rangle := \langle v|w\rangle|u\rangle.$$

These ket-bras form the basic building blocks of $M_2(\mathbb{R})$. Indeed, observe that

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

Example 1.3.8 — Pick two non-zero scalars $r, s \in \mathbb{R} \setminus \{0\}$. The matrix $\begin{bmatrix}r & 0\\0 & s\end{bmatrix}$ scales vectors on the x-axis by r and vectors on the y-axis by s:

$$\begin{bmatrix}r & 0\\0 & s\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}rx\\sy\end{bmatrix}$$



Observe that this scaling matrix is *diagonal*: the ij-entry is zero unless $i = j$.

Exercise 1.3.9 — Plot $\begin{bmatrix}1\\2\end{bmatrix}$. Then plot $\begin{bmatrix}2 & 0\\0 & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}$. Does your answer make sense?

Example 1.3.10 — In Example 1.3.8, if $r = 1$ and $s = 0$, then

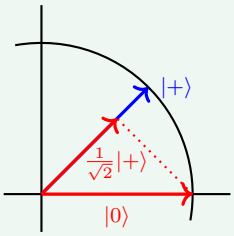
$$\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}x\\0\end{bmatrix} \in \mathbb{C}|0\rangle = \{z|0\rangle | z \in \mathbb{C}\},$$

the *span* of $|0\rangle$. This operator is called a *projection*, as it projects any vector to its x-component.

We could also project onto other lines passing through the origin. For example, we

project onto $\mathbb{C}|+\rangle$ via the operator $|+\rangle\langle+|$. Observe that

$$|+\rangle\langle+|v\rangle = \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \frac{1}{2}\begin{bmatrix}x+y\\ x+y\end{bmatrix}$$



Observe that in the diagram on the right, $\frac{1}{2}\begin{bmatrix}1\\ 0\end{bmatrix} = \frac{1}{\sqrt{2}}|+\rangle = |+\rangle\langle+|0\rangle$.

Algorithm 1.3.11 (2×2 Gram-Schmidt) — Using projections, we can complete any unit vector $|e_1\rangle$ to an ONB. First, choose any vector $|v\rangle$ which is not a scalar multiple of $|e_1\rangle$ (one of $|0\rangle$ or $|1\rangle$ must work.) Then form the vector

$$|v'\rangle := (I_2 - |e_1\rangle\langle e_1|)|v\rangle = |v\rangle - \langle e_1|v\rangle|e_1\rangle \qquad \text{and} \qquad |e_2\rangle := \frac{1}{\||v'\rangle\|}\cdot|v'\rangle.$$

Above, $|e_2\rangle$ is the *normalization* of $|v'\rangle$, which is obtained from dividing a non-zero vector by its norm to get a vector with unit length.

First, $|v'\rangle$ is not the zero vector, since $|v\rangle$ is not a multiple of $|e_1\rangle$. Second, $|v'\rangle$ is orthogonal to $|e_1\rangle$, since

$$\langle e_1|v'\rangle = \langle e_1|v\rangle - \langle e_1|v\rangle\langle e_1|e_1\rangle = \langle e_1|v\rangle - \langle e_1|v\rangle = 0.$$

Thus the normalization $|e_2\rangle$ of $|v'\rangle$ is still orthogonal to $|e_1\rangle$, so $\{|e_1\rangle, |e_2\rangle\}$ is an ONB.

We finish this section be defining some things one can do with a matrix that will be very important later on.

- The *determinant* of a 2×2 matrix is given by

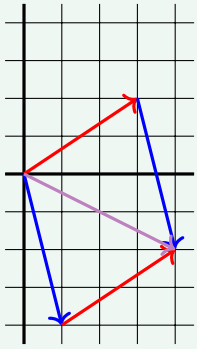
$$\det(A) = \det\begin{bmatrix}a & b\\ c & d\end{bmatrix} := ad - bc, \qquad A = \begin{bmatrix}a & b\\ c & d\end{bmatrix}$$

  which is the *signed area* of the parallelogram spanned by the two columns of A.

  Example 1.3.12 — Here are two examples of determinants as signed areas. Con-

sider the parallelogram spanned by the following vectors:

$$\begin{bmatrix}3\\2\end{bmatrix}, \begin{bmatrix}1\\-4\end{bmatrix}$$



Depending on the ordering of these two vectors as columns of a matrix, the determinant is either positive or negative:

$$\det\begin{bmatrix}3 & 1\\2 & -4\end{bmatrix} = -14 \qquad \det\begin{bmatrix}1 & 3\\-4 & 2\end{bmatrix} = 14.$$

- The *trace* of A is the sum of the diagonal entries, i.e., $\mathrm{Tr}(A) := a + d$.

Exercise 1.3.13 — Show that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ for all $A, B \in M_2(\mathbb{R})$.

- The *transpose* of a 2×2 matrix A^T is the matrix obtained from A by switching the rows and columns. That is, we reflect about the diagonal line from the top left to the bottom right of the matrix:

$$\begin{bmatrix}a & b\\c & d\end{bmatrix}^T := \begin{bmatrix}a & c\\b & d\end{bmatrix}.$$

Exercise 1.3.14 — Show that the transpose satisfies the following important identity:

$$\langle u|Av\rangle = \langle A^T u|v\rangle \qquad \forall\, |u\rangle, |v\rangle \in \mathbb{R}^2.$$

Theorem 1.3.15 (2×2 Square Matrix Theorem) — Suppose A is a 2×2 matrix. The following are equivalent.

(Sq1) $\det(A) \neq 0$.

(Sq2) A is *invertible*, i.e., there is a 2×2 matrix A^{-1} such that $A^{-1}A = I_2 = AA^{-1}$.

(Sq3) There is a 2×2 matrix B such that $BA = I_2$.

(Sq4) There is a 2×2 matrix B such that $AB = I_2$.

(Sq5) If $A|v\rangle = \mathbf{0}$, then $|v\rangle = \mathbf{0}$, the zero vector.

Proof.

(Sq1)⇒(Sq2): Define $A^{-1} := \dfrac{1}{ad-bc}\begin{bmatrix} a & -c \\ -b & d \end{bmatrix}$ and check that $AA^{-1} = A^{-1}A = I_2$.

(Sq2)⇒(Sq3),(Sq4): Set $B = A^{-1}$.

(Sq3)⇒(Sq5): Suppose $A|v\rangle = \mathbf{0}$. Multiply both sides by B to see that

$$|v\rangle = I_2|v\rangle = (BA)|v\rangle = B(A|v\rangle) = B\mathbf{0} = \mathbf{0}.$$

(Sq5)⇒(Sq1): We prove the *contrapositive*; that is, we assume that $\det(A) = 0$ and we construct a non-zero $|v\rangle \in \mathbb{R}^2$ such that $A|v\rangle = \mathbf{0}$.

Suppose $\det(A) = ad - bc = 0$. If A is the zero matrix, then any $|v\rangle$ will do, so we may assume A is not the zero matrix.

Case 1: If $a = 0$, then either $b = 0$ or $c = 0$. If $b = 0$, then the first row of A is zero, so $|v\rangle = \begin{bmatrix} -d \\ c \end{bmatrix}$ works. (Remember, we assumed A is not the zero matrix, so either $c \neq 0$ or $d \neq 0$.) If $c = 0$, then the first column of A is zero, so $|v\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ works by Trick 1.3.2.

Case 2: If $a \neq 0$, then $d = \dfrac{bc}{a}$. We then check that $|v\rangle = \begin{bmatrix} -b \\ a \end{bmatrix}$ works.

(Sq4)⇒(Sq2): We have already established that (Sq2) is equivalent to (Sq3). Observe that if $AB = I_2$, then by (Sq3)⇒(Sq2), B is invertible. This means we can multiply both sides by B^{-1} to see

$$A = AI_2 = A(BB^{-1}) = (AB)B^{-1} = I_2B^{-1} = B^{-1}.$$

But clearly B^{-1} is invertible with inverse equal to B, so A is invertible. ∎

Exercise 1.3.16 — Suppose $\det(A) \neq 0$. Verify that $AA^{-1} = I_2 = A^{-1}A$.

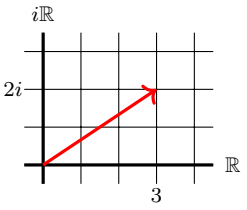
Exercise 1.3.17 — Suppose A, B, C are 2×2 matrices such that $AB = I_2$ and $BC = I_2$. Adapt the proof of (Sq4)⇒(Sq2) above to show that $A = C$. Deduce that B^{-1} is unique when it exists.

1.4 Complex numbers

In the study of quantum information, we will see that it is necessary to work with *complex numbers*:

$$\mathbb{C} = \left\{ x + iy \middle| x, y \in \mathbb{R} \text{ and } i^2 = -1 \right\}.$$

Just as a vector in $\mathbb{R}^2$ was described by two real numbers, so is a single complex number $z \in \mathbb{C}$. We can picture the complex numbers exactly like $\mathbb{R}^2$, with the real numbers $\mathbb{R}$ on the x-axis and the purely imaginary numbers $i\mathbb{R}$ on the y-axis.

$3 + 2i$ can be visualized as 

.

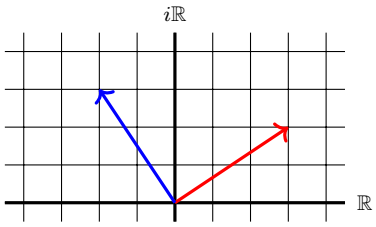
The *real part* of $z = x + iy$ is $\mathrm{Re}(z) = x$, and the *imaginary part* is $\mathrm{Im}(z) = y$.

We add complex numbers just like we add vectors in $\mathbb{R}^2$, and we can multiply complex numbers by real scalar exactly like in $\mathbb{R}^2$. However, $\mathbb{C}$ also has a *multiplication*:

$$(x_1 + iy_1)(x_2 + iy_2) = x_1x_2 - y_1y_2 + i(x_1y_2 + x_2y_1).$$

Exercise 1.4.1 — Prove that multiplication in $\mathbb{C}$ is commutative ($z_1z_2 = z_2z_1$) and associative ($z_1(z_2z_3) = (z_1z_2)z_3$).

How can we visualize complex multiplication? First, multiplication by i is *rotation* by $\pi/2$ in the counter-clockwise direction. Indeed, $i(x + iy) = -y + ix$.



.

In matrix form, the equation $i(x + iy) = -y + ix$ reads:

$$\begin{bmatrix} -y \\ x \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \cos\left(\frac{\pi}{2}\right) & -\sin\left(\frac{\pi}{2}\right) \\ \sin\left(\frac{\pi}{2}\right) & \cos\left(\frac{\pi}{2}\right) \end{bmatrix}.$$

Similarly, multiplying by $-i$ is rotation by $\pi/2$ in the clockwise direction.

It will take a little work to geometrically describe multiplication by an arbitrary complex number. To do so, we make the following definitions.

Definition 1.4.2 — The *complex conjugate* of $z = x + iy \in \mathbb{C}$ is $\overline{z} := x - iy$.
The *modulus* of $z = x + iy$ is $|z| := \sqrt{x^2 + y^2}$.

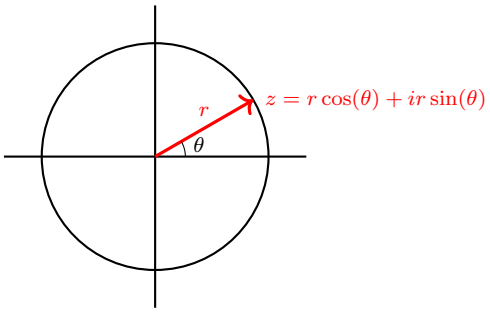
Exercise 1.4.3 — Prove that $\operatorname{Re}(z) = \dfrac{z + \overline{z}}{2}$ and $\operatorname{Im}(z) = \dfrac{z - \overline{z}}{2i}$.

Exercise 1.4.4 — Show that $|z|^2 = \overline{z}z$.

Exercise 1.4.5 — Prove the following inequalities for $w, z \in \mathbb{C}$:

$$|\operatorname{Re}(z)| \leq |z|, \qquad |\operatorname{Im}(z)| \leq |z|, \qquad \text{and} \qquad |w + z| \leq |w| + |z|.$$

Just as we can define polar coordinates in $\mathbb{R}^2$, we can define the *polar form* of a complex number z as $r\cos(\theta) + ir\sin(\theta)$ where $r = |z| \geq 0$ is the distance to zero, and θ is the angle from the x-axis measured in the counter-clockwise direction. Indeed, observe that $\cos(\theta) + i\sin(\theta)$ is the point on the unit circle whose angle from the x-axis is θ in the counter-clockwise direction, and multiplying by r scales the unit circle to have radius r.



We end this section with some notation that is often used for *phases*, i.e., complex numbers z with modulus $|z| = 1$, i.e., complex numbers that lie on the unit circle. The *exponential function* is defined by the infinite series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3 \cdot 2} + \frac{x^4}{4 \cdot 3 \cdot 2} + \cdots$$

which converges absolutely for all $x \in \mathbb{R}$. Now the above series also converges for all complex numbers, so we can plug in $x = i\theta$ for $\theta \in \mathbb{R}$. The first 8 terms (starting at zero) are:

$$1 + i\theta + \frac{i^2\theta^2}{2!} + \frac{i^3\theta^3}{3!} + \frac{i^4\theta^4}{4!} + \frac{i^5\theta^5}{5!} + \frac{i^6\theta^6}{6!} + \frac{i^7\theta^7}{7!} + \frac{i^8\theta^8}{8!} + \cdots$$

Now $i^2 = -1$, $i^3 = -i$ and $i^4 = (-1)^2 = 1$, so as the exponent of i grows, we cycle around the unit circle as $1, i, -1, -i$ and repeat in a period of 4. This allows us to separate the above sum as follows:

$$\begin{aligned}
&1 + i\theta + \frac{i^2\theta^2}{2!} + \frac{i^3\theta^3}{3!} + \frac{i^4\theta^4}{4!} + \frac{i^5\theta^5}{5!} + \frac{i^6\theta^6}{6!} + \frac{i^7\theta^7}{7!} + \frac{i^8\theta^8}{8!} + \cdots \\
&= 1 + i\theta - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + i\frac{\theta^5}{5!} - \frac{\theta^6}{6!} - i\frac{\theta^7}{7!} + \frac{\theta^8}{8!} + \cdots \\
&= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \frac{\theta^8}{8!} - \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots\right)
\end{aligned}$$

It turns out that the infinite sum in the first set of parentheses is exactly the *Taylor expansion* of the cosine function $\cos(\theta)$, and the infinite sum in the second set of parentheses is the Taylor expansion for the sine function $\sin(\theta)$. We have just proven the following theorem.

Theorem 1.4.6 (Euler's Formula) — For all $\theta \in \mathbb{R}$, $e^{i\theta} = \cos(\theta) + i\sin(\theta)$.

Exercise 1.4.7 — Show that under the identification of $\mathbb{R}^2$ with $\mathbb{C}$ sending $\begin{bmatrix} x \\ y \end{bmatrix}$ to $x+iy$, applying the rotation operator R_θ from Example 1.3.6 corresponds to multiplying by $e^{i\theta}$.

Returning to our discussion of polar form of a complex number, we see that polar form allows us to easily multiply complex numbers. Writing $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$, we have

$$\begin{aligned}
r_1 e^{i\theta_1} \cdot r_2 e^{i\theta_2} &= r_1(\cos(\theta_1) + i\sin(\theta_1)) \cdot r_2(\cos(\theta_2) + i\sin(\theta_2)) \\
&= r_1 r_2(\cos(\theta_1) + i\sin(\theta_1))(\cos(\theta_2) + i\sin(\theta_2)) \\
&= r_1 r_2\Big((\cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)) + i(\sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2)\Big) \\
&= r_1 r_2(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)) \\
&= r_1 r_2 e^{i(\theta_1 + \theta_2)}.
\end{aligned}$$

Here, we made use of the trigonometric identities for $\cos(\theta_1 + \theta_2)$ and $\sin(\theta_1 + \theta_2)$.

1.5 The qubit Hilbert space $\mathbb{C}^2$

In quantum information, a *qubit* is a state in the vector space $\mathbb{C}^2$, which is like $\mathbb{R}^2$, except that the ordered pair of numbers comes from $\mathbb{C}$ instead of $\mathbb{R}$:

$$\mathbb{C}^2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| x, y \in \mathbb{C} \right\}.$$

We can think of vectors in $\mathbb{C}^2$ like vectors in $\mathbb{R}^2$, with the exception that since numbers in $\mathbb{C}$ require 2 real numbers to specify, it requires 4 real numbers to specify a vector $|\psi\rangle \in \mathbb{C}^2$! This is not really possible to visualize in our 3-dimensional world.

The space $\mathbb{C}^2$ still has an addition and a scalar multiplication. We still have an inner product, which can be defined using matrix multiplication:

$$\left\langle \begin{bmatrix} w \\ x \end{bmatrix} \middle| \begin{bmatrix} y \\ z \end{bmatrix} \right\rangle := \begin{bmatrix} \overline{w} & \overline{x} \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \overline{w}y + \overline{x}z.$$

Exercise 1.5.1 — Compute the inner product $\langle v|w\rangle$ between the following three pairs of vectors in $\mathbb{C}^2$:

1. $|v\rangle = \begin{bmatrix} 1 \\ i \end{bmatrix}$, $|w\rangle = \begin{bmatrix} i \\ 1 \end{bmatrix}$
2. $|v\rangle = \begin{bmatrix} 1+i \\ i \end{bmatrix}$, $|w\rangle = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$
3. $|v\rangle = \begin{bmatrix} 1 \\ i \end{bmatrix}$, $|w\rangle = \begin{bmatrix} 1 \\ i \end{bmatrix}$

Remark 1.5.2 — In Remark 1.2.5, we saw that the axioms for a complex inner product are similar to those for a reall inner product from Axioms 1.2.4, but we replace (IP3) with the complex version.

We can still discuss kets $|\psi\rangle$ and bras $\langle\psi|$, with the added caveat that $\langle\psi|$ is the *conjugate transpose* of $|\psi\rangle$:

$$|\psi\rangle = \begin{bmatrix} x \\ y \end{bmatrix} \qquad \Longrightarrow \qquad \langle\psi| = \begin{bmatrix} \overline{x} & \overline{y} \end{bmatrix}.$$

The conjugate transpose is obtained by taking the transpose, and then taking the complex conjugate of every entry. Even though complex vectors have complex entries, we still can define their norms as before:

$$\||\psi\rangle\| = \sqrt{\langle\psi|\psi\rangle} \qquad \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| = \sqrt{|x|^2 + |y|^2}.$$

Exercise 1.5.3 — Compute the norms of the following three vectors in $\mathbb{C}^2$:

1. $|v\rangle = \begin{bmatrix} 1+i \\ 2 \end{bmatrix}$

2. $|v\rangle = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

3. $|v\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix}$

Similarly, we can discuss the angle between two vectors, as any two vectors live in a single plane which looks like $\mathbb{R}^2$. The angle formula is similar to before, with a real part added:

$$\cos(\theta) = \frac{\operatorname{Re}(\langle\psi|\eta\rangle)}{\||\psi\rangle\| \cdot \||\eta\rangle\|} \qquad |\psi\rangle, |\eta\rangle \in \mathbb{C}^2. \tag{1.5.4}$$

This real part appears because in the proof of Proposition 1.2.7 using the Law of Cosines, (1.2.9) is replaced in the complex setting by

$$\||\psi\rangle - |\eta\rangle\| = \sqrt{\langle\psi|\psi\rangle + \langle\eta|\eta\rangle - \langle\psi|\eta\rangle - \langle\eta|\psi\rangle} = \sqrt{\langle\psi|\psi\rangle + \langle\eta|\eta\rangle - 2\operatorname{Re}(\langle\psi|\eta\rangle)}.$$

Exercise 1.5.5 — Compute the angle between the vectors $|v\rangle = \begin{bmatrix} 2 \\ 1+i \end{bmatrix}$ and $|w\rangle = \begin{bmatrix} 2i \\ 3 \end{bmatrix}$ in $\mathbb{C}^2$.

In $\mathbb{C}^2$, as opposed to $\mathbb{R}^2$, there is a subtle difference between $|\psi\rangle, |\eta\rangle$ being *orthogonal*, i.e., $\langle\psi|\eta\rangle = 0$, and the angle between them being $\pi/2$ (90°).

Exercise 1.5.6 — Find 3 vectors in $\mathbb{C}^2$ such that the angle between any two of them is $\pi/2$. Note that this task is not possible in $\mathbb{R}^2$. If two vectors in $\mathbb{C}^2$ have an angle of $\pi/2$, are they necessarily orthogonal?

Again, $\mathbb{C}^2$ has *orthonormal bases* (ONBs) which play the same role as in $\mathbb{R}^2$: every vector can be written uniquely as a *linear combination* of the vectors in an ONB. Indeed, given an ONB $\{|e_1\rangle, |e_2\rangle\}$ for $\mathbb{C}^2$ and an arbitrary $|\psi\rangle \in \mathbb{C}^2$, we have

$$|\psi\rangle = \langle e_1|\psi\rangle|e_1\rangle + \langle e_2|\psi\rangle|e_2\rangle. \tag{1.5.7}$$

This is proven by taking the inner product of $|\psi\rangle$ with $|e_1\rangle$ and $|e_2\rangle$.

Finally, the Cauchy-Schwarz Inequality (Theorem 1.2.14) also holds for $\mathbb{C}^2$, and more generally for $\mathbb{C}^n$.

Exercise 1.5.8 — Write the vector $|\psi\rangle = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ as a linear combination of the elements of the orthonormal basis $\left\{|e_1\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix}, |e_2\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 1 \end{bmatrix}\right\}$.

1.6 $M_2(\mathbb{C})$: Operators on $\mathbb{C}^2$

Just as 2×2 real matrices in $M_2(\mathbb{R})$ act as linear operators on $\mathbb{R}^2$, 2×2 complex matrices in $M_2(\mathbb{C})$ act as linear operators on $\mathbb{C}^2$. We can still have the same types of operators as before, like the identity, rotations, diagonal scalings, and projections. We can still define the determinant and trace as before, and if $\det(A) \neq 0$, we still have an inverse matrix. Moreover, Theorem 1.3.15 still holds for 2×2 complex matrices. The transpose is also defined similarly, but we can also take complex conjugate to define the conjugate transpose.

Definition 1.6.1 — The *conjugate transpose* A^* (or $A^\dagger$) of A is given by taking the complex conjugate entries of the transpose matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^* := \begin{bmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{bmatrix}.$$

Exercise 1.6.2 — Show that the conjugate transpose satisfies the following important identity:

$$\langle \psi | A\eta \rangle = \langle A^*\psi | \eta \rangle \qquad \forall\, |\psi\rangle, |\eta\rangle \in \mathbb{C}^2.$$

Exercise 1.6.3 — Compute the conjugate transpose of the ket-bra $|\psi\rangle\langle\eta|$ for $|\psi\rangle, |\eta\rangle \in \mathbb{C}^2$.

Using the notion of the conjugate transpose, we can now define projection abstractly, and we can define a new kind of operator called a unitary.

Definition 1.6.4 — A 2×2 matrix P is called a *projection* if $P = P^* = P^2$.
A 2×2 matrix U is called a *unitary* if U is invertible and $U^* = U^{-1}$.

Exercise 1.6.5 — Suppose $|\psi\rangle \in \mathbb{C}^2$ is a vector with length 1. Show that $|\psi\rangle\langle\psi|$ is a projection.

Exercise 1.6.6 — Suppose $\{|e_1\rangle, |e_2\rangle\}$ is an ONB. Define $P_1 := |e_1\rangle\langle e_1|$ and $P_2 := |e_2\rangle\langle e_2|$. Prove that $P_1P_2 = 0 = P_2P_1$ and $P_1 + P_2 = I_2$.
Hint: To show $P_1 + P_2 = I_2$, show that $(P_1 + P_2)|\psi\rangle = |\psi\rangle$ for all $|\psi\rangle \in \mathbb{C}^2$ using (1.5.7).

Remark 1.6.7 — By (Sq3) in Theorem 1.3.15, to check U is unitary, it suffices to check either $U^*U = I_2$ or $UU^* = I_2$; you don't have to check both.

Exercise 1.6.8 — Show that $U = \begin{bmatrix} |e_1\rangle & |e_2\rangle \end{bmatrix}$ is unitary if and only if $\{|e_1\rangle, |e_2\rangle\}$ is an ONB.

The significance of the above exercise is that unitary matrices are exactly the matrices which map ONBs to ONBs.

Definition 1.6.9 — A 2×2 matrix A is called:

- *normal* if $A^*A = AA^*$ and
- *self-adjoint* if $A = A^*$.

Example 1.6.10 — The 2×2 *Pauli matrices* are given by

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Observe that they are all self-adjoint unitaries.

Exercise 1.6.11 — Verify the following relations:

- $X^2 = Y^2 = Z^2 = I$,
- $XY = -YX$, $XZ = -ZX$, and $YZ = -ZY$, and
- $-iXYZ = I_2$.

Then show that the relation $XY = -YX$ implies that $\mathrm{Tr}(XY) = 0$. Deduce that $\mathrm{Tr}(YZ) = \mathrm{Tr}(ZX) = 0$.

1.7 State vectors and vector states

Definition 1.7.1 — A *state vector* in $\mathbb{C}^2$ is a vector $|\psi\rangle \in \mathbb{C}^2$ with length one. Each state vector gives a corresponding *vector state* $|\psi\rangle\langle\psi| \in M_2(\mathbb{C})$, which is a projection.

This projection has *rank one*, meaning that if we apply $|\psi\rangle\langle\psi|$ to any vector $|\eta\rangle \in \mathbb{C}^2$, the result is a complex scaling of $|\psi\rangle$, i.e., it can be described by a single complex number. Rank is formally defined as the dimension of the image, which we will get to later on.

Observe that if $\alpha \in \mathbb{C}$ is a phase ($|\alpha| = 1$), then $|\alpha\psi\rangle$ is another state vector producing the *same vector state* as $|\psi\rangle$:

$$|\alpha\psi\rangle\langle\alpha\psi| = \overline{\alpha}\alpha|\psi\rangle\langle\psi| = |\alpha|^2|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|.$$

Since different state vectors can produce the same vector state, we can ask what the space of vector states looks like.

Proposition 1.7.2 — Given two state vectors $|\eta\rangle, |\psi\rangle$, we get equal vector states $|\eta\rangle\langle\eta| = |\psi\rangle\langle\psi|$ if and only if $|\eta\rangle = \alpha|\psi\rangle$ for some phase α.

Proof. We just saw that if $|\eta\rangle = \alpha|\psi\rangle$, then $|\eta\rangle\langle\eta| = |\psi\rangle\langle\psi|$. Suppose now $|\eta\rangle\langle\eta| = |\psi\rangle\langle\psi|$. Then

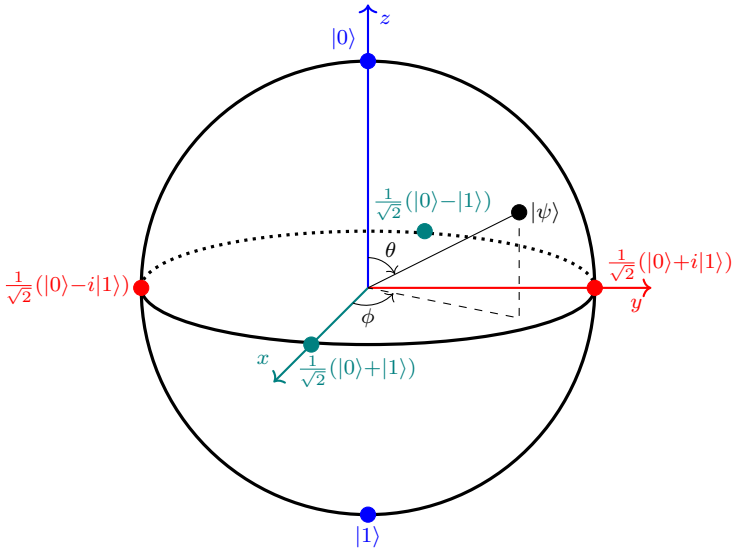
$$|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|\eta\rangle\langle\eta|\psi\rangle\langle\psi| = |\langle\eta|\psi\rangle|^2|\psi\rangle\langle\psi|,$$

so we can conclude that $|\langle\eta|\psi\rangle| = 1$. This means there is a phase $\alpha \in \mathbb{C}$ such that $\langle\eta|\alpha\psi\rangle = 1$. Let θ denote the angle between the two length one vectors $|\eta\rangle$ and $|\alpha\psi\rangle$. By (1.5.4),

$$\cos(\theta) = \frac{\mathrm{Re}(\langle\eta|\alpha\psi\rangle)}{\||\eta\rangle\| \cdot \||\alpha\psi\rangle\|} = \langle\eta|\alpha\psi\rangle = 1 \qquad \Longrightarrow \qquad \theta = 0.$$

This means that both $|\eta\rangle$ and $|\alpha\psi\rangle$ are length one vectors that point in the same direction, so they must be equal. ∎

This proposition tells us that the set of vector states in $M_2(\mathbb{C})$ is the same as the set of state vectors in $\mathbb{C}^2$ up to phase. We can now visualize this space as follows. We can write any state vector as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. multiplying by a phase, we may assume $\alpha \geq 0$. We have now used up our phase degree of freedom, so the parametrization we obtain will be one-to-one with vector states. Since $|\alpha|^2 + |\beta|^2 = \||\psi\rangle\|^2 = 1$, we may write $\alpha = \cos(\theta/2)$ for some and $\beta = e^{i\phi}\sin(\theta/2)$ for some $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. This is exactly the surface of the 2-sphere in $\mathbb{R}^3$ in polar coordinates! This representation is referred to as the *Bloch sphere*.



(1.7.3)

The angle θ represents the angle from the z-axis, and the angle ϕ represents the angle from the x-axis in the xy-plane.

Warning — In quantum information/mechanics, many people will just write $|\psi\rangle$ for the state instead of $|\psi\rangle\langle\psi|$, which is OK if you remember that $|\psi\rangle$ is the same state as $\alpha|\psi\rangle$ for a phase α.

One consequence of this is that the north and south poles of the Bloch sphere are labelled by $|0\rangle$ and $|1\rangle$ respectively, which can be quite confusing, as $|0\rangle$ and $|1\rangle$ are perpendicular vectors. Remember that the Bloch sphere is not the subset of state vectors in $\mathbb{C}^2$; the latter is the 3-sphere in $\mathbb{R}^4$. Rather, the Bloch sphere is the subset of *vector states* in $M_2(\mathbb{C})$. So we we should really label the north pole and south poles respectively by

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

Remark 1.7.4 — The map from the 3-dimensional sphere of state vectors to the corresponding vector state in the 2-dimensional sphere is an important and beautiful mathematical object called the *Hopf fibration*. This is well beyond the scope of this class, but it is so important and beautiful that we cannot help but mention it here.

1.8 Quantum observables and measurements

We now state the axioms of quantum mechanics/quantum information for quantum observables and measurements.

Axioms 1.8.1 (Axioms of Quantum Observables I) **—** Consider a quantum system consisting of a single qubit.

(Q^I1) A quantum state is a vector state $|\psi\rangle\langle\psi| \in M_2(\mathbb{C})$ where $|\psi\rangle \in \mathbb{C}^2$ is a state vector.

(Q^I2) A quantum observable is a self-adjoint operator $A \in M_2(\mathbb{C})$, whose eigenvalues must be real by Exercise 1.9.4.

(Q^I3) There are two possibilities when measuring the observable A:

- Suppose $|\psi\rangle$ is an *eigenstate* of A, meaning there is some $\lambda \in \mathbb{R}$ called the corresponding *eigenvalue* of A such that

$$A|\psi\rangle = \lambda|\psi\rangle.$$

This property is unchanged multiplying $|\psi\rangle$ by a phase. In this case, the

outcome of the measurement is the corresponding eigenvalue, and the state remains unchanged.

- If $|\psi\rangle$ is not an eigenvector of A (which is again a property unchanged multiplying $|\psi\rangle$ by a phase), the outcome of the measurement is probabilistic, and occurs by the *Born rule* as follows. Let $\{|e_1\rangle, |e_2\rangle\}$ be an ONB of eigenstates of A corresponding to eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ respectively. The outcome of measuring A in state $|\psi\rangle\langle\psi|$ is λ_i with probability $|\langle e_i|\psi\rangle|^2$ for $i = 1, 2$. After measuring, the state *changes* to $|e_i\rangle\langle e_i|$ if the value λ_i is observed.

The above axioms say that if we want to measure the observable A in an eigenstate, the result is completely determined. However, if we want to measure A in a state that is not an eigenstate, the result is not determined, and we get a random result according to the particular probability distribution corresponding to our state.

Example 1.8.2 — If we measure Z in the state $|0\rangle$, the result is $+1$ with probability 1, and $|0\rangle$ remains unchanged. However, if we measure X in the state $|0\rangle$, the outcome will be ± 1 with probability $1/2$, and the state will change to $|\pm\rangle$ depending on the outcome.

Definition 1.8.3 — Given an observable A with an ONB of eigenstates $\{|e_1\rangle, |e_2\rangle\}$ and a quantum state $|\psi\rangle$ which is not an eigenvector of A, we say $|\psi\rangle$ is in a *superposition* of $|e_1\rangle, |e_2\rangle$.

Warning — It does not make sense to say that a state $|\psi\rangle$ is in a superposition unless you have chosen a quantum observable to measure. We can only say $|\psi\rangle$ is in a superposition relative to some basis of eigenstates of a quantum observable.

1.9 Eigenvalues and eigenvectors

Given their importance to measuring quantum observables, we now study eigenvalues and eigenvectors of operators in $M_2(\mathbb{C})$ systematically.

Definition 1.9.1 — Given a 2×2 matrix A, a $\lambda \in \mathbb{C}$ is called an *eigenvalue* of A if there is a non-zero vector $|\psi\rangle \in \mathbb{C}^2$ called an *eigenvector* of A such that $A|\psi\rangle = \lambda|\psi\rangle$.

Example 1.9.2 — Suppose $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ is a diagonal matrix. Then by Trick 1.3.2,

the eigenvalues of D are λ_1, λ_2 with corresponding eigenvectors $|0\rangle, |1\rangle$. For a specific example, the eigenvalues of Z are ± 1.

Example 1.9.3 — The eigenvalues of X are also ± 1 with corresponding eigenvectors $|\pm\rangle$.

Exercise 1.9.4 — Suppose A is normal.

(1) Show that $\|A|\psi\rangle\|^2 = \|A^*|\psi\rangle\|^2$ for all $|\psi\rangle \in \mathbb{C}^2$.

(2) Show that $A - \lambda I$ is also normal for all $\lambda \in \mathbb{C}$.

(3) Use (1) applied to the normal matrix $A - \lambda I$ to prove that if $|\psi\rangle$ is an eigenvector for A with eigenvalue λ, then $|\psi\rangle$ is also an eigenvector for A^* with eigenvalue $\overline{\lambda}$.

(4) Use (3) to prove that if A is self-adjoint, then the eigenvalues of A are real.

Exercise 1.9.5 — Use Theorem 1.3.15 to prove that 0 is an eigenvalue of A if and only if A is not invertible.

Algorithm 1.9.6 (Eigenvalue/Eigenvector Algorithm) — The following algorithm produces the eigenvalues and eigenvectors of a 2×2 matrix A:

Step 1: By (Sq5) in Theorem 1.3.15, the values of λ for which $A - \lambda I$ are not invertible are exactly the $\lambda \in \mathbb{C}$ for which there is a non-zero $|\psi\rangle \in \mathbb{C}^2$ such that $(A - \lambda I)|\psi\rangle = \mathbf{0}$, i.e., $A|\psi\rangle = \lambda|\psi\rangle$.

Step 2: Compute the *characteristic polynomial* of A, given by $\chi_A(\lambda) := \det(A - \lambda I)$. The values of $\lambda \in \mathbb{C}$ such that $\chi_A(\lambda) = 0$ are the eigenvalues of A by Step 1.

Step 3: For each eigenvalue λ of A, find a non-zero $|\psi\rangle \in \mathbb{C}^2$ such that $(A - \lambda I)|\psi\rangle = \mathbf{0}$. This can be done quite quickly, since the rows of $A - \lambda I$ must be multiples of each other. If $A - \lambda I = 0$, the zero matrix, then $|0\rangle, |1\rangle$ are both eigenvectors. If $A - \lambda I \neq 0$, then pick a non-zero row of A, say $\begin{bmatrix} x & y \end{bmatrix}$, and set $|\psi\rangle = \begin{bmatrix} y \\ -x \end{bmatrix}$.

Note: if you cannot find a non-zero $|\psi\rangle$ in Step 3 for λ, you have made an algebraic mistake, and you should go back to Step 2.

Remark 1.9.7 — Let's actually compute $\chi_A(\lambda) = \det(A - \lambda I)$. Observe:

$$\det \begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix} = (a-\lambda)(d-\lambda) - bc = \lambda^2 - (a+d)\lambda + (ad-bc) = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A).$$

Now remember that the eigenvalues are the roots of $\chi_A(\lambda)$, i.e., the λ_1, λ_2 such that we have a factorization $\chi_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)$. Now expanding this factorization, we see

$$\chi_A(\lambda) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2.$$

We conclude that we must have that $\mathrm{Tr}(A) = \lambda_1 + \lambda_2$ and $\det(A) = \lambda_1\lambda_2$!

Example 1.9.8 — Consider the matrix

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The characteristic polynomial is $\chi_N(\lambda) = \lambda^2$, so 0 is the only eigenvalue. We see there is only one eigenvector up to scaling, namely $|0\rangle$.

Example 1.9.9 — Consider the matrix

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}.$$

Its characteristic polynomial is

$$\chi_A(\lambda) = (5-\lambda)(2-\lambda) - 4 = \lambda^2 - 7\lambda + 6 = (\lambda - 1)(\lambda - 6),$$

and so the eigenvalues are $1, 6$. We compute

$$A - I = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \qquad \Longrightarrow \qquad |e_1\rangle = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

is an eigenstate for $\lambda = 1$. We compute

$$A - 6I = \begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \qquad \Longrightarrow \qquad |e_1\rangle = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

is an eigenstate for $\lambda = 6$. Setting

$$U := \frac{1}{\sqrt{5}}\begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix} \qquad \text{and} \qquad D = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix},$$

we see U is unitary, D is diagonal, and $AU = UD$.

Theorem 1.9.10 (2×2 Spectral Theorem) — For a 2×2 matrix A, the following are equivalent:

(Sp1) There is an ONB consisting of eigenvectors of A.

(Sp2) There is a diagonal matrix D and a unitary matrix U such that $AU = UD$.

(Sp3) There are non-zero pairwise orthogonal projections P_1, P_2 with $P_1P_2 = 0 = P_2P_1$ and $P_1 + P_2 = I_2$ and scalars $\lambda_1, \lambda_2 \in \mathbb{C}$ such that $A = \lambda_1 P_1 + \lambda_2 P_2$.

(Sp4) A is normal.

Proof.

(Sp1)⇒(Sp2): Let $\{|e_1\rangle, |e_2\rangle\}$ be such an ONB corresponding to eigenvalues λ_1, λ_2. Then

$$A\underbrace{\begin{bmatrix} |e_1\rangle & |e_2\rangle \end{bmatrix}}_{=:U} = \begin{bmatrix} A|e_1\rangle & A|e_2\rangle \end{bmatrix} = \begin{bmatrix} \lambda_1|e_1\rangle & \lambda_2|e_2\rangle \end{bmatrix} = \underbrace{\begin{bmatrix} |e_1\rangle & |e_2\rangle \end{bmatrix}}_{=U}\underbrace{\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}}_{=:D}.$$

Observe that U is unitary by Exercise 1.6.8.

(Sp2)⇒(Sp3): Suppose $AU = UD$, and write $U = \begin{bmatrix} |e_1\rangle & |e_2\rangle \end{bmatrix}$ as in the previous proof. Then

$$A = UDU^* = \begin{bmatrix} |e_1\rangle & |e_2\rangle \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} \langle e_1| \\ \langle e_2| \end{bmatrix} = \begin{bmatrix} |e_1\rangle & |e_2\rangle \end{bmatrix}\begin{bmatrix} \lambda_1\langle e_1| \\ \lambda_2\langle e_2| \end{bmatrix} = \lambda_1 \underbrace{|e_1\rangle\langle e_1|}_{=:P_1} + \lambda_2 \underbrace{|e_2\rangle\langle e_2|}_{=:P_2}.$$

Now observe that the ket-bras P_1, P_2 are projections by Exercise 1.6.5, and they satisfy $P_1P_2 = 0 = P_2P_1$ and $P_1 + P_2 = I_2$ by Exercise 1.6.6.

(Sp3)⇒(Sp4): Suppose $A = \lambda_1 P_1 + \lambda_2 P_2$. Then $A^* = \overline{\lambda_1}P_1^* + \overline{\lambda_2}P_2^* = \overline{\lambda_1}P_1 + \overline{\lambda_2}P_2$, and since $P_1^2 = P_1$, $P_2^2 = P_2$, and $P_1P_2 = 0 = P_2P_1$, we calculate

$$A^*A = (\overline{\lambda_1}P_1 + \overline{\lambda_2}P_2)(\lambda_1 P_1 + \lambda_2 P_2) = |\lambda_1|^2 P_1 + |\lambda_2|^2 P_2 = (\lambda_1 P_1 + \lambda_2 P_2)(\overline{\lambda_1}P_1 + \overline{\lambda_2}P_2) = AA^*.$$

(Sp4)⇒(Sp1): By Algorithm 1.9.6, we know A has an eigenvector $|\psi_1\rangle$ corresponding to an eigenvalue λ_1. Since $|\psi_1\rangle$ is non-zero, we may normalize it to get a length one vector $|e_1\rangle$ which is still an eigenvector corresponding to the eigenvalue λ_1.

Complete $|e_1\rangle$ to an ONB by finding $|e_2\rangle$ of length 1 with $\langle e_2|e_1\rangle = 0$. We claim $|e_2\rangle$ is also an eigenvector of A. First, there are $\alpha, \beta \in \mathbb{C}$ such that

$$A|e_2\rangle = \alpha|e_1\rangle + \beta|e_2\rangle.$$

We can calculate α using (1.5.7):

$$\alpha = \langle e_1|A|e_2\rangle = \langle A^* e_1|e_2\rangle \underset{\text{(Ex. 1.9.4)}}{=} \langle \overline{\lambda_1} e_1|e_2\rangle = \lambda_1 \langle e_1|e_2\rangle = 0.$$

We conclude that $A|e_2\rangle = \beta|e_2\rangle$, and thus $|e_2\rangle$ is an eigenvector as claimed. Since $\{|e_1\rangle, |e_2\rangle\}$ is an ONB, we are finished. ∎

Exercise 1.9.11 — Suppose U is a 2×2 self-adjoint unitary with distinct eigenvalues.

(1) Prove that the eigenvalues of U are ± 1.

(2) Show that $P = \dfrac{I - U}{2}$ is an orthogonal projection.

(3) Show that if $|e_{\pm 1}\rangle$ is the eigenvector for U with eigenvalue ± 1, then $P|e_{+1}\rangle = 0$ and $P|e_{-1}\rangle = |e_{-1}\rangle$.

1.10 Random variables

We saw in Axioms 1.8.1 that given an observable A and a state $|\psi\rangle$, the outcome of measuring A is probabilistic and given by the Born rule. In the language of probability theory, we would say the outcome (λ_1 or λ_2) is a *random variable* on some *event set* (outcome 1 or 2) with a given *probability distribution* ($p_1 = |\langle e_i|\psi\rangle|^2$ and $p_2 = |\langle e_2|\psi\rangle|^2$).

We now give a more formal treatment of these basic concepts of probability theory.

Definition 1.10.1 — An *event set* Ω is a (finite) set of outcomes of a given experiment. For example, if we are flipping a coin, the outcomes are heads and tails, so $\Omega = \{H, T\}$.

A *probability distribution* on Ω is a function from Ω to $[0, 1]$ which assigns a probability value p_x to each outcome $x \in \Omega$ such that $\sum_{x\in\Omega} p_x = 1$. Typically we assume that $p_x > 0$ for every $x \in \Omega$; otherwise we remove points from Ω with zero probability of occurrence.

We call an event set Ω equipped with a probability distribution $\{p_x\}_{x\in\Omega}$ a *probability space*.

A *random variable* R on a probability space $(\Omega, \{p_x\}_{x\in\Omega})$ is a function which assigns real numbers to elements of Ω.

Beyond observing operators in quantum states, here are some more examples of probability spaces and random variables.

Example 1.10.2 — A *Bernoulli trial* is an experiment with two outcomes *success* and *failure*, where 'success' has probability p and 'failure' has probability $1-p$. Typically a Bernoulli trial corresponds to the random variable

$$R(\text{success}) = 1 \qquad\qquad R(\text{failure}) = 0.$$

Typical examples include flipping a weighted coin with probability of heads p.

Example 1.10.3 — Iterating Bernoulli trials with probability of success p gives the *binomial distribution*. Typically, we have a fixed number n of Bernoulli trials, and the event set is $\Omega = \{0, \dots, n\}$ corresponding to the possible number of successes. The probability of exactly k successes is

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}.$$

That is, out of the n trials, to get exactly k successes, we pick k of the n trials to succeed, where each has probability p, and $n-k$ of the trials to fail, where each has probability $1-p$. There are exactly $\binom{n}{k}$ many ways for this to happen.

Example 1.10.4 — Given a set $\Omega = \{x_1, \dots, x_n\}$, the *uniform distribution* assigns equal probability $p_i = 1/n$ to each outcome x_i. We will see in §1.16 below that this distribution is the 'most random' in some sense.

Definition 1.10.5 — Given a random variable R on a probability space $(\Omega, \{p_x\}_{x\in\Omega})$, the *expected* or *average value* of R is

$$\mathbb{E}[R] := \sum_{x\in\Omega} R(x) p_x.$$

The *variance* of R is

$$\operatorname{Var}[R] := \mathbb{E}[(R - \mathbb{E}[R])^2] = \mathbb{E}[R^2] - \mathbb{E}[R]^2 \geq 0.$$

The *standard deviation* of R is $\sigma[R] := \sqrt{\operatorname{Var}[R]}$.

Exercise 1.10.6 — Compute the expected value, variance, and standard deviation of

the random variable $X(H) = 1$ and $X(T) = 0$ for a weighted coin with probability p of heads. What value of p maximizes the variance?

Exercise 1.10.7 — Alice and Bob work at the same office and like to gamble. Each day, they each wear one of two hats to work, a red hat or a blue hat. If both hats are red, Alice gives Bob \$3. If both hats are blue, Alice gives Bob \$1. If the hats do not match, Bob gives Alice \$2. Would you rather be Alice or Bob? Why?

In the above formula for the variance, we used the following result.

Proposition 1.10.8 — Expected value is *linear*. That is, given random variables X, Y on $(\Omega, \{p_x\}_{x\in\Omega})$ and $r \in \mathbb{R}$,

$$\mathbb{E}[rX + Y] = r\mathbb{E}[X] + \mathbb{E}[Y].$$

Proof. We compute

$$\mathbb{E}[rX + Y] = \sum_{x\in\Omega}(rX(x) + Y(x))p_x = r\sum_{x\in\Omega} X(x)p_x + \sum_{x\in\Omega} Y(x)p_x = r\mathbb{E}[X] + \mathbb{E}[Y]. \qquad \blacksquare$$

Remark 1.10.9 — The above proof should look vaguely like linearity of the integral in calculus; indeed, it is an integral over a discrete space, where the values p_x play the role of dx. This is made rigorous in *measure theory*.

Exercise 1.10.10 — Let $X : \{1, 2\} \to \mathbb{R}$, and consider the two probability distributions $\{p_1 = 1/4, p_2 = 3/4\}$ and $\{p_1 = 1/2, p_2 = 1/2\}$. Compute $\mathrm{Var}(X)$ with respect to both distributions, and compare them. Is one always bigger than the other?

Suppose now we have an experiment with outcomes Ω that we iterate N times without knowing ahead of time the probability distribution $\{p_x\}_{x\in\Omega}$. What is our best estimate for the p_x? Given a random variable $X : \Omega \to \mathbb{R}$, what is our best estimate for $\mathbb{E}[X]$, $\mathrm{Var}[X]$, and $\sigma[X]$?

- To estimate p_x, we perform N experiments (with N large) and count the number of times N_x that x occurs. We would then estimate $\widehat{p}_x = N_x/N$.
- To estimate $\mathbb{E}[X]$ for a random variable X, we use our estimated $\widehat{p}_x$ to get $\widehat{\mathbb{E}}[X] = \sum_{x\in\Omega} X(x)\widehat{p}_x$.
- Estimation of $\mathrm{Var}(X)$ and $\sigma[X]$ is similar.

In the field of *statistics*, we also have formulas for 'error estimates' for these estimates.

Example 1.10.11 — If we have a weighted coin with probability p of heads and we flip it 10000 times and get 3535 heads, we would estimate that $p \approx 3535/10000$.

1.11 Expected value of observables and the Uncertainty Principle

Here is our most important example of expected value.

Proposition 1.11.1 — The expected value of the observable A in the state $|\psi\rangle$ is $\langle A\rangle := \langle\psi|A|\psi\rangle$.

Proof. We calculate the expected value of the random variable R which is λ_i with probability $|\langle e_i|\psi\rangle|^2$ where $\{|e_1\rangle, |e_2\rangle\}$ is an ONB of eigenstates for A corresponding to eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$. We see

$$\begin{aligned}\mathbb{E}[R] &= \lambda_1|\langle e_1|\psi\rangle|^2 + \lambda_2|\langle e_2|\psi\rangle|^2 \\ &= \lambda_1\langle\psi|e_1\rangle\langle e_1|\psi\rangle + \lambda_2\langle\psi|e_2\rangle\langle e_2|\psi\rangle \\ &= \langle\psi|\Big(\lambda_1|e_1\rangle\langle e_1| + \lambda_2|e_2\rangle\langle e_2|\Big)|\psi\rangle \\ &= \langle\psi|A|\psi\rangle.\end{aligned}$$

The last equality above follows from the Spectral Theorem 1.9.10. ∎

Exercise 1.11.2 — Prove that the expected value of the observable A in the state $|\psi\rangle$ is also given by $\mathrm{Tr}(A|\psi\rangle\langle\psi|)$.

We can now use the notion of expected value and variance to state the *Uncertainty Principle*, which gives a lower bound for the product of the variances of two observables in terms of the commutator $[A, B] := AB - BA$. Roughly speaking, this means that we cannot know measurements for both A and B in the same state if A and B do not commute. For an observable A and a chosen state vector $|\psi\rangle \in \mathbb{C}^2$, we denote by σ_A^2 the variance of A viewed as a random variable:

$$\sigma_A^2 := \langle\psi|(A - \mathbb{E}[A])^2|\psi\rangle = \mathbb{E}[A^2] - \mathbb{E}[A]^2.$$

Theorem 1.11.3 (Uncertainty Principle) **—** For observables $A, B \in M_2(\mathbb{C})$ and a state vector $|\psi\rangle \in \mathbb{C}^2$,

$$\sigma_A\sigma_B \geq \left|\frac{1}{2i}\langle\psi|AB - BA|\psi\rangle\right| = \left|\mathbb{E}\left[\tfrac{AB-BA}{2i}\right]\right|.$$

Proof. Let $|\psi_1\rangle = (A - \mathbb{E}[A])|\psi\rangle$ and $|\psi_2\rangle = (B - \mathbb{E}[B])|\psi\rangle$. By the Cauchy-Schwarz Inequality (Theorem 1.2.14),

$$|\langle\psi_1|\psi_2\rangle|^2 \leq \langle\psi_1|\psi_1\rangle \cdot \langle\psi_2|\psi_2\rangle = \sigma_A^2\sigma_B^2.$$

Since $\text{Im}(z)^2 \leq |z|^2$ for all complex numbers $z \in \mathbb{C}$, we have

$$|\langle\psi_1|\psi_2\rangle|^2 \geq \text{Im}(\langle\psi_1|\psi_2\rangle)^2 = \left(\frac{\langle\psi_1|\psi_2\rangle - \langle\psi_2|\psi_1\rangle}{2i}\right)^2.$$

We calculate that

$$\langle\psi_1|\psi_2\rangle = \langle\psi|(A - \mathbb{E}[A])(B - \mathbb{E}[B])|\psi\rangle = \langle\psi|AB|\psi\rangle - \mathbb{E}[A]\mathbb{E}[B]$$
$$\langle\psi_2|\psi_1\rangle = \langle\psi|(B - \mathbb{E}[B])(A - \mathbb{E}[A])|\psi\rangle = \langle\psi|BA|\psi\rangle - \mathbb{E}[A]\mathbb{E}[B],$$

so their difference is exactly $\langle\psi|AB - BA|\psi\rangle$. The result follows. ∎

The following result stands in contrast to the Uncertainty Principle; if A, B commute, there are states where we can know with 100% certainty measurements for both A, B simultaneously!

Theorem 1.11.4 — If $A, B \in M_2(\mathbb{C})$ are commuting observables, there is a common ONB of eigenstates $\{|e_1\rangle, |e_2\rangle\}$ for A and B.

Proof. First, by unitarily diagonalizing B, we may assume B is diagonal. Indeed, by the Spectral Theorem 1.9.10, there is a diagonal matrix D and a unitary matrix U such that $B = UDU^*$. Then $AUDU^* = AB = BA = UDU^*A$, so multiplying by U^* on the left and U on the right, we have $U^*AUD = DU^*AU$. If $\{|e_1\rangle, |e_2\rangle\}$ is a simultaneous ONB of eigenstates for U^*AU and D, then $\{U|e_1\rangle, U|e_2\rangle\}$ will be a simultaneous ONB of eigenstates for A and B.

Now assuming B is diagonal, we compute AB and BA:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} b_{11} & \\ & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{22} \\ a_{21}b_{11} & a_{22}b_{22} \end{bmatrix} \qquad \begin{bmatrix} b_{11} & \\ & b_{22} \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{11} \\ a_{21}b_{22} & a_{22}b_{22} \end{bmatrix}.$$

Since these are equal, we must have $a_{12}b_{11} = a_{12}b_{22}$, or equivalently, $a_{12}(b_{11} - b_{22}) = 0$.

Case 1: If $a_{12} = 0$, then since A is self-adjoint, we have $a_{21} = 0$, so A is also diagonal. Hence $\{|0\rangle, |1\rangle\}$ is a simultaneous ONB of eigenstates for A and B, as desired.

Case 2: If $b_{11} - b_{22} = 0$, then $B = bI_2$ for some $b \in \mathbb{C}$. Thus any ONB of eigenstates for A is also an ONB of eigenstates for B, as desired. ∎

This means that if A, B are commuting observables, and $|\psi\rangle$ is a common eigenstate for both A and B, then we know the outcome of measuring A and B simultaneously with 100% certainty.

1.12 Mutually unbiased bases and Stern-Gerlach revisited

Two ONBs are mutually unbiased if they are as 'far away from each other' as possible. We make this rigorous using the inner product, since for any two length-one vectors $|\psi\rangle, |\eta\rangle \in \mathbb{C}^2$, the angle $\theta \in [0, \pi/2]$ between them is determined by $\cos(\theta) = |\langle\eta|\psi\rangle|$.

Definition 1.12.1 — Two ONBs $\{|e_1\rangle, |e_2\rangle\}$ and $\{|f_1\rangle, |f_2\rangle\}$ of $\mathbb{C}^2$ are said to be *mutually unbiased* if

$$|\langle e_i|f_j\rangle|^2 = \frac{1}{2} \qquad \forall 1 \leq i, j \leq 2.$$

There is an analogous definition for mutually unbiased ONBs of $\mathbb{C}^n$ replacing $1/2$ with $1/n$.

Exercise 1.12.2 — Show that $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$ are mutually unbiased.

Exercise 1.12.3 — Find another ONB in $\mathbb{C}^2$ which is mutually unbiased with respect to both $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$. Can you find any more?

Remark 1.12.4 — One can find $p + 1$ mutually unbiased ONBs in $\mathbb{C}^p$ for every prime number p. At this time, it is an *open problem* how many mutually unbiased ONBs can be found in $\mathbb{C}^6$!

We now revisit the Stern-Gerlach experiment from §1.1 and explain the phenomena there in terms of mutually unbiased bases. We saw that if we have two light polarization filters oriented the same way, then all light that passes through the first filter will pass through the second. However, if we rotate the second filter by 90°, no light gets through.

Mathematically, this can be explained as follows.[1] The initial photon is some qubit $|\psi\rangle$. As we are free to declare what 'zero rotation' ought to mean, we declare that our first light polarization filter acts as the spectral projection of the Pauli Z operator for the $\lambda = 1$ eigenspace, namely

$$p_{Z=1} = |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Assuming no rotation for the filter, when the photon meets the filter, it passes through with probability $|\langle\psi|0\rangle|^2$, after which it is in state $|0\rangle$.

A light polarization filter rotated by θ acts as the $p_{Z=1}$ operator conjugated by the unitary R_θ representing the rotation. If the second filter is not rotated relative to the first, it is another $p_{Z=1}$ operator, and since $|0\rangle$ has eigenvalue $+1$, the outcome is $+1$ with 100%

[1] There is also a classical description of light polarization and attenuation (Malus' Law), which gives the same predictions as the quantum mechanical description here.

probability, meaning every photon passing through the first filter passes through the second. However, if we rotate the filter by 90°, we must conjugate $p_{Z=1}$ by

$$R_{\pi/2} = \begin{bmatrix} \cos(\pi/2) & -\sin(\pi/2) \\ \sin(\pi/2) & \cos(\pi/2) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

meaning we have the observable

$$R_{\pi/2} Z R_{\pi/2}^* = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

This is exactly the spectral projection $p_{Z=-1}$, which is orthogonal to $p_{Z=1}$. For this operator, $|0\rangle$ has eigenvalue 0, so it passes through with probability $|\langle 1|0\rangle|^2 = 0$. This means no light passes through the 90°-rotated filter.

Now what happens when we add a filter between these two rotated by 45°? The $\pi/4$-rotation means we conjugate by $R_{\pi/4}$:

$$\begin{aligned} \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos(\pi/4) & \sin(\pi/4) \\ -\sin(\pi/4) & \cos(\pi/4) \end{bmatrix} &= \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = |+\rangle\langle +|, \end{aligned}$$

which is exactly the spectral projection $p_{X=1}$ onto $\mathbb{C}|+\rangle$. Note that the ONB of eigenstates $\{|+\rangle, |-\rangle\}$ for $p_{X=1}$ is *mutually unbiased* with respect to the ONB of eigenstates $\{|0\rangle, |1\rangle\}$ for $p_{Z=1} = |0\rangle\langle 0|$.

Now when the photon in state $|0\rangle$ meets the second filter rotated by $\pi/4$ corresponding to $p_{X=1}$, it passes through with probability $|\langle +|0\rangle|^2 = 1/2$ and changes into the $|+\rangle$ state. Now when it meets the third filter rotated by another $\pi/4$ (which is $\pi/2$ from the original Pauli Z), we apply $p_{Z=-1}$, so the photon passes through with probability $|\langle 1|+\rangle|^2 = 1/2$. Thus the initial beam of light is reduced by 1/2 intensity each time, resulting in a total reduction by $1/8 > 0$.

Exercise 1.12.5 — Suppose $\{|e_1\rangle, |e_2\rangle\}$ is an ONB of eigenstates for the observable A and U is a unitary matrix. Show that $\{U|e_1\rangle, U|e_2\rangle\}$ is an ONB of eigenstates for the observable UAU^*. Deduce that $\{R_\theta|0\rangle, R_\theta|1\rangle\}$ is an ONB of eigenstates for the observable $R_\theta Z R_\theta^*$.

Exercise 1.12.6 — Show that the operator corresponding to rotating the light polarization filter by $22.5° = \pi/8$ is the *Hadamard gate*

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

Then compute an ONB of eigenstates for this operator.

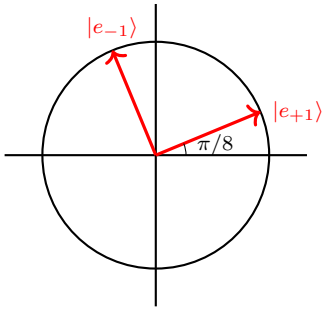
Exercise 1.12.7 — Show that $ZHZ = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$. What are its eigenvalues and corresponding eigenstates?

1.13 Towards violations of Bell's Inequality

This subsection is adapted from material from Bell's Theorem: The Quantum Venn Diagram Paradox https://www.youtube.com/watch?v=zcqZHYo7ONs.

Using the sequential Stern-Gerlach style experiments, we can provide evidence that measurements of observables are random, and not pre-determined by some kind of *hidden variable*. We consider light polarization filters with angles 0, $22.5° = \pi/8$, and $45° = \pi/4$. We will call these type A, B, C respectively. We already saw in the previous section that A corresponds to the Pauli Z operator, B corresponds to the Hadamard gate H, and C corresponds to Pauli X.

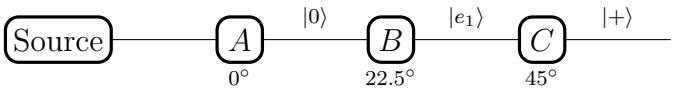
We denote the eigenstate corresponding to the eigenvalue ± 1 for H by $|e_{\pm 1}\rangle$. This means that if a photon in state $|\psi\rangle$ meets a B filter, the filter allows the photon to pass through with probability $|\langle\psi|e_{+1}\rangle|^2$, after which it is in state $|e_{+1}\rangle$. By Exercises 1.12.5 above, we see that the ONB from Exercise 1.12.6 is a 22.5° rotation of the standard basis, up to choices of phases.



Exercise 1.13.1 — Calculate the probability that a photon passes through:

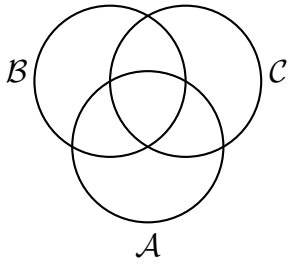
- a type A filter and then a type A filter,
- a type A filter and then a type B filter,
- a type A filter and then a type C filter,
- a type B filter and then a type C filter,
- a type A filter and then a type B filter, but then is then blocked by a type C filter.

We now consider three *sequential* light polarization filters with angles 0, $22.5^\circ = \pi/8$, and $45^\circ = \pi/4$.

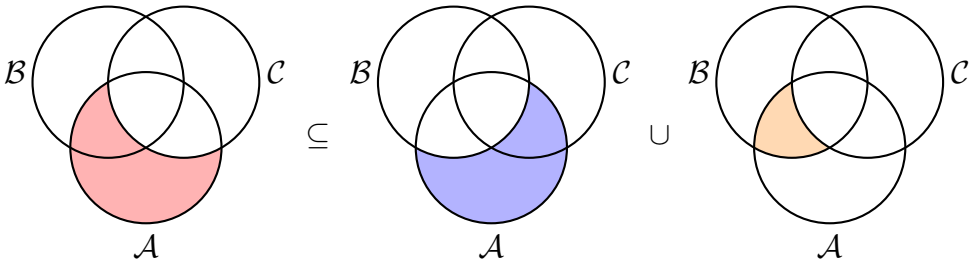


At this point, we make an assumption of *realism*, that is, we assume every photon has a *pre-determined* answer $a, b, c \in \{0, 1\}$ to whether it passes through the A, B, C filters respectively. For example, if a photon has $a = 1$, $b = 1$, and $c = 0$, then it will pass through the 0° and 22.5° filters, but not the 45° filter. We will see this assumption leads to a contradiction.

We repeat this experiment for a large number N of photons, and we let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ denote the sets of photons whose values for a, b, c respectively are $+1$. Consider the following Venn diagram:



Intuitively, the only way a photon can pass through the initial A filter and not the final C filter is if it is blocked at the B or C filters. This means we have an inclusion of subsets $\mathcal{A} \setminus \mathcal{C} \subseteq (\mathcal{A} \setminus \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B}) \setminus \mathcal{C}$:



Note that all three of these are subsets of $\mathcal{A}$. This containment gives the following inequality:

$$|\mathcal{A} \setminus \mathcal{C}| \leq |\mathcal{A} \setminus \mathcal{B}| + |(\mathcal{A} \cap \mathcal{B}) \setminus \mathcal{C}|. \tag{1.13.2}$$

This means the number of particles which satisfy $a = +1$ and $b = -1$ is less than or equal to the sum of the number of particles for which $a = +1$ and $b = -1$ and the number of particles for which $a = +1$ and $b = +1$, but $c = -1$.

Now if we were to actually set up these polarization filters, we would see that for every 100 photons in $\mathcal{A}$:

- about 50 of them are blocked by C, meaning $|\mathcal{A} \setminus \mathcal{C}| \approx 50$,
- about 15 of them are blocked by B, meaning $|\mathcal{A} \setminus \mathcal{B}| \approx 15$, and

- of all the photons that pass through B filter, about 15% are blocked at C, so $|(\mathcal{A} \cap \mathcal{B}) \setminus \mathcal{C}| \lesssim 15$.

But this means that the inequality (1.13.2) would be $50 \lesssim 15 + 15 = 30$, a contradiction.

Another way to see this apparent contradiction is as follows. You saw in Exercise 1.13.1 above that after passing through A, a photon has a 50% chance at passing through C. But if it first approaches B, it has a 85% chance to pass through B, after which it has another 85% chance to pass through C. But note that $.5 < (.85)^2$, so introducing an *additional* filter has somehow *increased* the number of particles which pass through. Such a simple hidden variable model above would not allow for this.

To make this mathematically rigorous, we will rephrase the above in probability theory, in particular random variables and expectations. To do so, we introduce the notion of *conditional probability*.

Definition 1.13.3 — Given a probability space $(\Omega, \{p_x\}_{x\in\Omega})$, an *event* is a subset $A \subset X$. The *probability of the event* A is $\mathbb{P}(A) := \sum_{a\in A} p_a$.

Example 1.13.4 — Given a random variable X on a probability space $(\Omega, \{p_x\}_{x\in\Omega})$, we can consider the event that $X = r \in \mathbb{R}$. The probability of this event occurring is

$$\mathbb{P}(X = r) := \sum_{x\in\Omega} \{p_x | X(x) = r\}.$$

Example 1.13.5 — Suppose we have a fair 6-sided die. The probability of the event that we roll 1 or 2 is

$$\mathbb{P}(\text{roll is 1 or 2}) = 2\frac{1}{6} = \frac{1}{3}.$$

Definition 1.13.6 — Given two events $A, B \subset \Omega$, the event that both A and B occur is the intersection $A \cap B$. We define the *conditional probability* that A occurs given that B occurs as

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Here, since B is assumed to have occurred, we have assumed $\mathbb{P}(B) > 0$. There is a similar definition for the conditional probability that B occurs given A occurs.

Exercise 1.13.7 — Show that conditional probability gives a probability distribution on Ω. That is, show that for a fixed event B, the numbers

$$q_x := \begin{cases} \dfrac{p_x}{\mathbb{P}(B)} & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$$

give a probability distribution Ω.

Example 1.13.8 — Given two random variables X, Y on $(\Omega, \{p_x\}_{x\in\Omega})$, we have

$$\mathbb{P}(X = r \text{ and } Y = s) := \sum_{x\in\Omega} \{p_x | X(x) = r \text{ and } Y(x) = s\}.$$

We have a that the conditional probability that $X = r$ given that $Y = s$ is given by the formula

$$\mathbb{P}(X = r | Y = s) := \frac{\mathbb{P}(X = r \text{ and } Y = s)}{\mathbb{P}(Y = s)}.$$

Surprisingly, we can compute the conditional probability $\mathbb{P}(B|A)$ in terms of the conditional probability $\mathbb{P}(A|B)$! Indeed, observe that

$$\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A).$$

As long as $\mathbb{P}(A) > 0$, we can divide to obtain the following result.

Theorem 1.13.9 (Bayes) — For events A, B in $(\Omega, \{p_x\}_{x\in\Omega})$, if $\mathbb{P}(A) > 0$, then

$$\mathbb{P}(B|A) = \mathbb{P}(A|B)\frac{\mathbb{P}(B)}{\mathbb{P}(A)}.$$

Example 1.13.10 (Monty Hall problem) — In the famous game show *Let's Make a Deal*, the host Monty Hall would present contestants with a dilemma. A prize would be hiding behind one of three doors. After the contestant picked a door, say door 1, Monty would open one of the doors 2 or 3 not hiding the prize. (Monty never opens the door the contestant picks.) At this point, the contestant is given the opportunity to switch doors. Should they do so?

Since the prize is behind door 1 exactly 1/3 of the time, if the contestant stays with door 1, they will still win exactly 1/3 of the time. However, 2/3 of the time the contestant picked the wrong door, and Monty shows them the correct door by opening one of the two alternatives! Thus switching wins 2/3 of the time.

This result may seem counter-intuitive; there are only 2 doors left after Monty opens

one, so why should switching have any effect whatsoever? To answer this, let's discuss the Monty Hall problem using Bayes' Theorem. Let A be the event that Monty opens one of the doors the contestant does not choose, and let B be the event that the contestant has chosen the correct door. Observe that $P(A) = 1$, $P(B) = 1/3$, and $P(A|B) = 1$. Thus

$$\mathbb{P}(B|A) = \mathbb{P}(A|B)\frac{\mathbb{P}(B)}{\mathbb{P}(A)} = 1 \cdot \frac{1/3}{1} = \frac{1}{3}.$$

The reason this problem is counter-intuitive is that $P(A|B) = 1$! Monty *always* opens one of the doors the contestant does not pick.

We are now in the position to give a rigorous mathematical treatment of the above contradiction. Consider a set of photons traveling towards the filters of type A, B, C set up in series. We can view the values of a, b, c for each photon as a random variable, and (1.13.2) becomes the following inequality:

$$\mathbb{P}(c = -1|a = 1) \leq \mathbb{P}(b = -1|a = 1) + \mathbb{P}(c = -1|a = 1 \text{ and } b = 1).$$

In real-world experiments, we have observed that these probabilities are 50%, 15%, and 15% respectively, violating the expected inequality.

The mathematical reason is clear for this violation. For a given photon, if $a = 1$, the state becomes $|0\rangle$, so if we measure a type C filter, there is a 50% chance we get -1 as $|\langle -|0\rangle|^2 = 1/2$. If $a = 1$ and we measure a type B filter, the probability we get -1 is $|\langle e_{-1}|0\rangle|^2 \approx .15$. Similarly, if $a = 1$ and $b = 1$, the state becomes $|e_{+1}\rangle$, and if we measure a type C filter, the chance we get -1 is $|\langle -|e_{-1}\rangle|^2 \approx .15$.

1.14 Mixed states

Suppose Alice has an electron in state $|\psi\rangle$, and she measures its spin the the z-direction by measuring Pauli Z. She then leaves the room, and passes by Bob in the hallway. She says, "I measured the electron in the room in the z-direction," but she does not mention the outcome of her experiment. Bob then walks into the room to perform measurements on the electron. What state best represents the state of the electron? Alice measured either ± 1, so the answer is either $|0\rangle$ or $|1\rangle$, but Bob has no idea which value she measured.

In this scenario we say the state is a *classical mixture* of $|0\rangle$ and $|1\rangle$, meaning it can be modeled by classical probability theory. Indeed, there is some $p \in [0, 1]$ such that the state is $|0\rangle$ with probability p and $|1\rangle$ with probability $1 - p$. We represent this by the *density matrix*

$$\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| = \begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix}.$$

Definition 1.14.1 — A *density matrix* for one qubit is a $\rho \in M_2(\mathbb{C})$ such that

- ρ is *positive* ($\rho \geq 0$), meaning $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathbb{C}^2$, and

- $\mathrm{Tr}(\rho) = \rho_{11} + \rho_{22} = 1$.

A density matrix is called a *pure state* if there is a state vector $|\psi\rangle \in \mathbb{C}^2$ such that $\rho = |\psi\rangle\langle\psi|$. Otherwise, it is called a *mixed state*.

Exercise 1.14.2 — For a state vector $|\psi\rangle \in \mathbb{C}^2$, prove that $\rho = |\psi\rangle\langle\psi|$ is a density matrix.

Example 1.14.3 — The matrix $\frac{1}{2}I_2$ is a mixed state.

In these notes, we will take the point of view that *all quantum states are pure states*, and *mixed states* reflect incomplete information about the state of a system.

Recall from Exercise 1.11.2 that $\langle\psi|A|\psi\rangle = \mathrm{Tr}(A|\psi\rangle\langle\psi|)$ for all observables $A \in M_2(\mathbb{C})$ and state vectors $|\psi\rangle \in \mathbb{C}^2$. This formula also gives the expected value of A in a mixed state ρ.

Proposition 1.14.4 — The expected value of an observable A in a mixed state ρ is $\mathrm{Tr}(A\rho)$.

Proof. Using the spectral decomposition of ρ,

$$\mathrm{Tr}(A\rho) = p\,\mathrm{Tr}(A|e_1\rangle\langle e_1|) + (1-p)\,\mathrm{Tr}(A|e_2\rangle\langle e_2|) = p\langle e_1|A|e_1\rangle + (1-p)\langle e_2|A|e_2\rangle.$$

This latter quantity is the expected value of A in the mixed state ρ. ∎

Exercise 1.14.5 — Suppose $\rho = |\psi\rangle\langle\psi|$ is a pure state.

- Show ρ is not invertible.
- Show ρ has rank one, i.e., the columns of A are multiples of each other.

The next results proves the converse of the previous exercise.

Proposition 1.14.6 — Suppose $\rho \in M_2(\mathbb{C})$ is a density matrix which is not invertible. There is a $|\psi\rangle \in \mathbb{C}^2$ such that $\rho = |\psi\rangle\langle\psi|$.

Proof. First, since ρ is self-adjoint and has trace 1, we can write it as

$$\rho = \begin{bmatrix} t & z \\ \overline{z} & 1-t \end{bmatrix}$$

for some $t \in \mathbb{R}$ and $z \in \mathbb{C}$. Since ρ is positive, $\langle 0|\rho|0\rangle = t \geq 0$ and $\langle 1|\rho|1\rangle = 1 - t \geq 0$, so $t \in [0, 1]$. Now since ρ is not invertible,

$$\det(\rho) = t - t^2 - |z|^2 = 0,$$

so $|z| = \sqrt{t(1-t)}$, and there is a phase α with $|\alpha| = 1$ such that $z = \alpha|z|$. This means we can write

$$\rho = \begin{bmatrix} t & \alpha\sqrt{t(1-t)} \\ \overline{\alpha}\sqrt{t(1-t)} & 1-t \end{bmatrix}.$$

Finally, we leave it to the reader to check that $\rho = |\psi\rangle\langle\psi|$ where $|\psi\rangle = \sqrt{t}|0\rangle + \alpha\sqrt{1-t}|1\rangle$. ■

We now give several more equivalent conditions to a density matrix ρ being a pure state. Observe that since ρ is self-adjoint, the Spectral Theorem 1.9.10 applies. This means there is an ONB $\{|e_1\rangle, |e_2\rangle\}$ of eigenstates for ρ and eigenvalues $0 \leq \lambda_1, \lambda_2 \leq 1$ such that

$$\rho = \lambda_1|e_1\rangle\langle e_1| + \lambda_2|e_2\rangle\langle e_2|.$$

Since $\lambda_1 + \lambda_2 = \mathrm{Tr}(\rho) = 1$ by Remark 1.9.7, using this spectral decomposition, we see the following are equivalent:

- ρ is pure,
- 0 is an eigenvalue of ρ, and
- 1 is an eigenvalue of ρ.

There are a couple other conditions as well in terms of ρ^2. Since $\langle e_1|e_2\rangle = \langle e_2|e_1\rangle = 0$, we see

$$\rho^2 = (\lambda_1|e_1\rangle\langle e_1| + \lambda_2|e_2\rangle\langle e_2|)(\lambda_1|e_1\rangle\langle e_1| + \lambda_2|e_2\rangle\langle e_2|) = \lambda_1^2|e_1\rangle\langle e_1| + \lambda_2^2|e_2\rangle\langle e_2|.$$

Recall ρ is pure if and only if λ_1, λ_2 are $0, 1$ (up to swapping) if and only if $\lambda_1^2 = \lambda_1$ and $\lambda_2^2 = \lambda_2$. We can now add the next two equivalent conditions to ρ being pure:

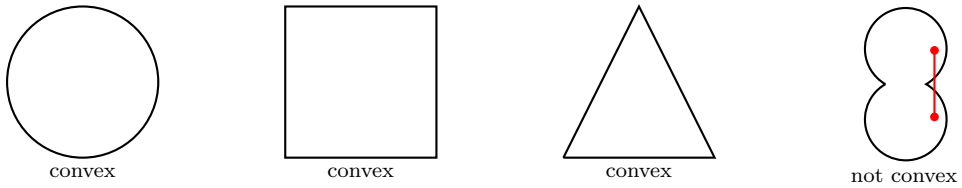
- $\rho^2 = \rho$, and
- $\mathrm{Tr}(\rho^2) = 1$.

Indeed, if $0 < \lambda_1 < 1$, then $\rho^2 \neq \rho$ and $\mathrm{Tr}(\rho^2) = \lambda_1^2 + \lambda_2^2 < 1$.

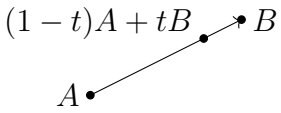
1.15 Quantum state tomography

So far, we have given several characterizations of when a state ρ is pure, but we have not seen many examples of mixed states. We now give two methods to construct lots of mixed states.

The first uses *convex combinations*. Recall that a shape in $\mathbb{R}^2$ is called *convex* if given any two points in the shape, the line segment between them is also contained in the shape.



A convex combination of matrices $A, B \in M_2(\mathbb{C})$ (or more generally vectors in any vector space) is a sum of the form $(1-t)A + tB$ for $t \in [0,1]$. Geometrically, as t goes from 0 to 1, we travel along a straight line from A to B in $M_2(\mathbb{C})$:



The point A corresponds to $t = 0$ and the point B corresponds to $t = 1$.

Proposition 1.15.1 — Suppose $\rho_1, \rho_2 \in M_2(\mathbb{C})$ are density matrices and $0 \leq t \leq 1$. Then the convex combination $(1-t)\rho_1 + t\rho_2$ is a density matrix.

Proof. First, observe that $\mathrm{Tr} : M_2(\mathbb{C}) \to \mathbb{C}$ is a linear map, which means

$$\mathrm{Tr}(\alpha A + B) = \alpha \,\mathrm{Tr}(A) + \mathrm{Tr}(B) \qquad \forall\, A, B \in M_2(\mathbb{C}),\ \forall\, \alpha \in \mathbb{C}.$$

This means for any two densities $\rho_1, \rho_2 \in M_2(\mathbb{C})$ and $0 \leq t \leq 1$, the convex combination $(1-t)\rho_1 + t\rho_2$ also has trace 1. Indeed,

$$\mathrm{Tr}((1-t)\rho_1 + t\rho_2) = (1-t)\,\mathrm{Tr}(\rho_1) + t\,\mathrm{Tr}(\rho_2) = t + (1-t) = 1.$$

Second, for any fixed $|\psi\rangle \in \mathbb{C}^2$, the expectation map $\mathbb{E}_{|\psi\rangle}(A) := \langle\psi|A|\psi\rangle$ is a linear map $M_2(\mathbb{C}) \to \mathbb{C}$. This means for any two densities $\rho_1, \rho_2 \in M_2(\mathbb{C})$ and $0 \leq t \leq 1$, the convex combination $(1-t)\rho_1 + t\rho_2$ is also positive. Indeed, for any $|\psi\rangle \in \mathbb{C}^2$,

$$\langle\psi|(1-t)\rho_1 + t\rho_2|\psi\rangle = (1-t)\underbrace{\langle\psi|\rho_1|\psi\rangle}_{\geq 0} + t\underbrace{\langle\psi|\rho_2|\psi\rangle}_{\geq 0} \geq 0.$$

The result follows. ∎

We can now give lots of examples of mixed states.

Example 1.15.2 — We can write $\frac{1}{2}I_2$ as a convex combination in multiple ways:

$$\frac{1}{2}I_2 = \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$$
$$\frac{1}{2}I_2 = \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix} = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -|.$$

Example 1.15.3 — Consider the two states $|0\rangle$ and $|1\rangle$. We get a mixed state by taking a convex combination:

$$\frac{1}{4}|0\rangle\langle 0| + \frac{3}{4}|+\rangle\langle +| = \frac{1}{4}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + \frac{3}{8}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} = \frac{1}{8}\begin{bmatrix}5 & 3\\ 3 & 3\end{bmatrix}.$$

Now we also have a geometric interpretation of the space of states in $M_2(\mathbb{C})$: the Bloch sphere. In §1.7 above, we saw that the pure states are exactly the *surface/boundary* of the Bloch sphere (1.7.3). The mixed states, which are non-trivial convex combinations of pure states are then in the *interior* of the Bloch sphere.

There is a better way to represent these mixed states in the interior of the Bloch sphere than using convex combinations, which did not give a unique representation as seen in Example 1.15.2 above. First, we note that I_2, X, Y, Z form a *basis* of $M_2(\mathbb{C})$, meaning that every matrix can be written *uniquely* as a *linear combination* of these four matrices:

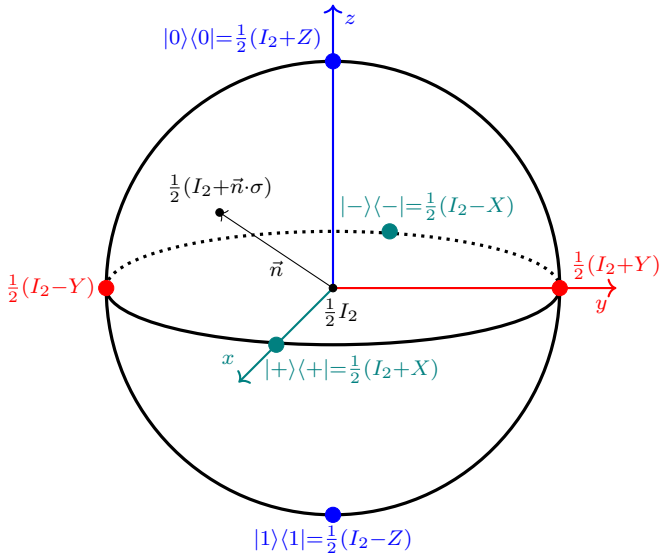
$$A = aI_2 + bX + cY + dZ \qquad\qquad a, b, c, d \in \mathbb{C}.$$

Moreover, each of I_2, X, Y, Z are self-adjoint, so they form a *real* basis for the real vector space of self-adjoint 2×2 matrices. This means every self-adjoint matrix can be written uniquely as a linear combination of these matrices where the coefficients are in $\mathbb{R}$ instead of $\mathbb{C}$.

Now, note that each of X, Y, Z have trace zero, so if we write any density matrix ρ in the I_2, X, Y, Z basis, we must have that the coefficient of I_2 is $1/2$. We thus incorporate an overall factor of $1/2$ in our coordinates:

$$\rho = \frac{1}{2}(I_2 + n_x X + n_y Y + n_z Z) = \frac{1}{2}(I_2 + \vec{n}\cdot\vec{\sigma}) \qquad\qquad n_x, n_y, n_z \in [-1, 1]. \quad (1.15.4)$$

Here, we write $\vec{n} = \begin{bmatrix}n_x\\ n_y\\ n_z\end{bmatrix}$ for the *Bloch vector*, and $\vec{\sigma} = \begin{bmatrix}X\\ Y\\ Z\end{bmatrix}$ as shorthand notation.



Now if we look again at the Bloch sphere geometrically, the center is $\frac{1}{2}I_2$, and we get to the surface by adding $\vec{n}\cdot\vec{\sigma}$ for some $\vec{n} \in \mathbb{R}^3$ with norm exactly 1. For example, $|0\rangle\langle 0| = \frac{1}{2}(I_2 + Z)$ and $|+\rangle\langle +| = \frac{1}{2}(I_2 + X)$. The interior points exactly correspond to those $\vec{n}$ with norm strictly less than one. We thus get the following equivalent characterizations for a state ρ to be pure:

- ρ is on the surface/boundary of the Bloch sphere, or
- the Bloch vector $\vec{n}$ of ρ has norm 1.

In order to calculate the Bloch vector for a given density ρ, we can use expectation values as in §1.11 above. Indeed, using Exercise 1.6.11 above, we see that in the state ρ,

$$\langle X \rangle = \operatorname{Tr}(X\rho) = \frac{1}{2}\operatorname{Tr}(X + n_x I_2 + n_y XY + n_z XZ) = \frac{n_x}{2}\operatorname{Tr}(I_2) = n_x$$

and similarly $n_y = \operatorname{Tr}(Y\rho)$ and $n_z = \operatorname{Tr}(Z\rho)$. This is another reason for the choice of overall multiplicative factor of $1/2$ for the expression (1.15.4).

In practice, if we want to prepare a particular state in a lab, we need to be able to check that we have prepared the correct state. A process to do this is *quantum state tomography*. We assume we can prepare as many copies of the same state $\rho \in M_2(\mathbb{C})$ as we would like, say a large number $3N$. We can then take measurements of the observables X, Y, Z numerous times, say N times each, and count how many times we measure ± 1 for each. We call these numbers $N^X_{\pm 1}, N^Y_{\pm 1}, N^Z_{\pm 1}$ respectively, where $N^X_{+1} + N^X_{-1} = N$, and similarly for Y, Z. Our best estimate for the Bloch vector is then

$$\widehat{n}_x \approx \frac{N^X_{+1} - N^X_{-1}}{N} \qquad \widehat{n}_y \approx \frac{N^Y_{+1} - N^Y_{-1}}{N} \qquad \widehat{n}_z \approx \frac{N^Z_{+1} - N^Z_{-1}}{N}.$$

Example 1.15.5 — Suppose we have an unknown density matrix ρ, and we measure the observables X, Y and Z 90 times each. For X, we obtain 1 on 73 measurements and -1 on 17 measurements; for Y, we obtain 1 and -1 on 45 measurements each; for Z, we obtain 1 on 10 measurements and -1 on 80 measurements. Our best estimate for ρ is

$$\widehat{\rho} \approx \frac{1}{2}\left(I_2 + \frac{56}{90}X + 0Y - \frac{70}{90}Z\right).$$

1.16 Shannon entropy

In Shannon's original paper [Sha48], he gives axioms for a function $\mathcal{I}$ from $(0,1]$ to $[0,\infty)$ meant to represent the amount of *information* or *surprise* that one gets from observing an event with probability $p \in [0,1]$. For example, if we observe an event with probability 1, we have not learned anything, since it was sure to happen. We cannot observe an event with probability 0, so it is is not in the domain of the function.

Axioms 1.16.1 (Axioms of information function) — We assume the function $\mathcal{I} : (0,1] \to [0,\infty)$ satisfies the following three properties:

(I1) If $p < q$, then $\mathcal{I}(p) > \mathcal{I}(q)$. That is, we gain more information when a more rare event occurs.

(I2) $\mathcal{I}(1) = 0$. That is, if an event is sure to happen, we get no information.

(I3) $\mathcal{I}(pq) = \mathcal{I}(p) + \mathcal{I}(q)$. That is, if we observe two independent events, with probabilities p and q, we get a sum of information from both events.

Proposition 1.16.2 — If $\mathcal{I}$ is twice differentiable, then $\mathcal{I}(p) = -\log(p)$ for all $p \in (0,1]$, up to choice of base for the logarithm.

Proof. First, we differentiate (I3) ($\mathcal{I}(pq) = \mathcal{I}(p)+\mathcal{I}(q)$) with respect to p (holding q constant) to get

$$q\mathcal{I}'(pq) = \mathcal{I}'(p).$$

Differentiating again with respect to q yields

$$\mathcal{I}'(pq) + pq\mathcal{I}''(pq) = 0.$$

Now introducing a change of variables $pq = r$, we have the differential equation

$$\mathcal{I}''(r) + r\mathcal{I}'(r) = (r\mathcal{I}'(r))' = 0.$$

This means that $r\mathcal{I}'(r) = c$, a constant, so $\mathcal{I}'(r) = c/r$, and thus $\mathcal{I}(r) = c\log(r) + k$ for some constant k. By (I2), $k = 0$, and by (I1), $c < 0$. The choice of c exactly corresponds to the choice of base for the logarithm. ∎

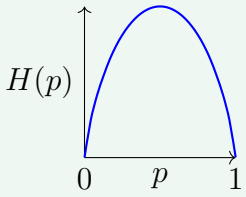
Given a probability distribution $\{p_1, \ldots, p_n\}$ on the set $\{x_1, \ldots, x_n\}$, the *Shannon entropy* is the *expected/average information* obtained from one experiment.

Definition 1.16.3 — The *Shannon entropy* of $\{p_1, \ldots, p_n\}$ is $H(\{p_i\}) := -\sum p_i \log(p_i)$.

Shannon entropy is is a measure of *uncertainty* for the distribution. We will compute two examples of this.

Example 1.16.4 — Suppose we have a weighted coin with probability of heads p and probability of tails $1-p$. The entropy of this distribution is

$$H(p) = -p\log(p) - (1-p)\log(1-p).$$



To maximize H, we take the derivative and set it equal to zero:

$$H'(p) = -1 - \log(p) - \big((-1)\log(1-p) + 1\big) = -\log(p) + \log(1-p) = 0.$$

We see that $\log(p) = \log(1-p)$ exactly when $p = 1/2$, i.e., the coin is fair. (Technically, we should check that $H'(p)$ changes sign at $p = 1/2$; we leave this to the reader.) This makes sense: we gain the most average information from a single coin flip if the coin is fair.

Example 1.16.5 — The entropy of the distribution $\{p_1, \ldots, p_n\}$ is

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log(p_i).$$

To maximize H, we use the method of Lagrange multipliers. We add to H the constraint $\sum p_i = 1$ multiplied by a scalar λ:

$$G(p_1, \ldots, p_n, \lambda) := H(p) + \lambda\left(\sum_{i=1}^{n} p_i - 1\right) = -\sum_{i=1}^{n} p_i \log(p_i). + \lambda\left(\sum_{i=1}^{n} p_i - 1\right).$$

Taking partial derivatives, we see

$$\frac{\partial G}{\partial p_i}(p_1, \ldots, p_n, \lambda) = 1 - \log(p_i) + \lambda \qquad \forall\, i = 1, \ldots, n$$
$$\frac{\partial G}{\partial \lambda}(p_1, \ldots, p_n, \lambda) = \sum_{i=1}^{n} p_i - 1.$$

Setting all these partial derivatives equal to zero, we obtain our constraint back, together with the equation $\lambda + 1 = \log(p_i)$ for all $i = 1, \ldots, n$. This means the critical point happens when all p_i are equal to the *same value*, i.e., $\{p_1, \ldots, p_n\}$ is the *uniform distribution*. (Technically, we have to check that something happens at this critical point to verify we have a local maximum, but this is left to the reader.)

One important feature of entropy in comparison to the variance of a random variable is that entropy only depends on the distribution, and not on any random variable.

Exercise 1.16.6 — Consider the two probability distributions $\{p_1 = 1/6, p_2 = 1/3, p_3 = 1/2\}$ and $\{p_1 = 1/6, p_2 = 1/6, p_3 = 2/3\}$ on $\{1, 2, 3\}$.

(1) Which distribution has greater entropy?

(2) Let $X : \{1, 2, 3\} \to \mathbb{R}$ be given by $X(1) = 10$, $X(2) = 1$ and $X(3) = 2$. Compute $\mathrm{Var}(X)$ with respect to both distributions. Which one is greater?

1.17 von Neumann entropy

States/density matrices have their own notion of entropy due to von Neumann.

Definition 1.17.1 — The *von Neumann entropy* of a state $\rho \in M_2(\mathbb{C})$ is

$$S(\rho) := -\operatorname{Tr}(\rho \log(\rho)).$$

Here, $\log(\rho)$ is defined in terms of the Spectral Theorem 1.9.10. Since ρ is a density matrix, we can write

$$\rho = p|e_1\rangle\langle e_1| + (1-p)|e_2\rangle\langle e_2|$$

for some ONB $\{|e_1\rangle, |e_2\rangle\}$ and $p \in [0, 1]$. We then define

$$-\rho \log(\rho) := -p \log(p)|e_1\rangle\langle e_1| - (1-p)\log(1-p)|e_2\rangle\langle e_2|.$$

Here, the trace Tr plays the role of the summation in the usual formula for $H(\{p_i\})$.

Lemma 1.17.2 — For any $A \in M_2(\mathbb{C})$ and ONB $\{|e_1\rangle, |e_2\rangle\}$ of $\mathbb{C}^2$,

$$\mathrm{Tr}(A) = \langle e_1|A|e_1\rangle + \langle e_2|A|e_2\rangle.$$

Proof. Form the unitary matrix

$$U = \begin{bmatrix} |e_1\rangle & |e_2\rangle \end{bmatrix}$$

and observe that for all $A \in M_2(\mathbb{C})$, $\mathrm{Tr}(A) = \mathrm{Tr}(I_2 A) = \mathrm{Tr}(UU^*A) = \mathrm{Tr}(U^*AU)$. But then we calculate

$$\mathrm{Tr}(U^*AU) = \langle 0|U^*AU|0\rangle + \langle 1|U^*AU|1\rangle = \langle e_1|A|e_1\rangle + \langle e_2|A|e_2\rangle,$$

and the result follows. ∎

Lemma 1.17.3 — For any state vector $|\psi\rangle \in \mathbb{C}^2$, $\mathrm{Tr}(|\psi\rangle\langle\psi|) = 1$.

Proof. Using the 2×2 Gram-Schmidt Algorithm 1.3.11, we can complete $|\psi\rangle$ to an ONB $\{|\psi\rangle, |\phi\rangle\}$. Using the previous lemma,

$$\mathrm{Tr}(|\psi\rangle\langle\psi|) = \langle\psi|\psi\rangle\langle\psi|\psi\rangle + \langle\phi|\psi\rangle\langle\psi|\phi\rangle = 1.$$ ∎

Exercise 1.17.4 — Use the strategy of the above proof to give another proof that $\langle\psi|A|\psi\rangle = \mathrm{Tr}(A|\psi\rangle\langle\psi|)$ for any state vector $|\psi\rangle$ and observable $A \in M_2(\mathbb{C})$.

We get the following immediate corollary.

Corollary 1.17.5 — The von Neumann entropy $S(\rho)$ is equal to the Shannon entropy $H(\{p, 1-p\})$ where

$$\rho = p|e_1\rangle\langle e_1| + (1-p)|e_2\rangle\langle e_2|$$

is the spectral decomposition of ρ.

Proof. We use Lemmas 1.17.2 and 1.17.3 to calculate

$$\begin{aligned} S(\rho) &= -\mathrm{Tr}(\rho \log(\rho)) = -p\log(p)\,\mathrm{Tr}(|e_1\rangle\langle e_1|) - (1-p)\log(1-p)\,\mathrm{Tr}(|e_2\rangle\langle e_2|) \\ &= -p\log(p) - (1-p)\log(1-p) = H(p). \end{aligned}$$ ∎

We now see from the above corollary that 1 is an eigenvalue of ρ if and only if $S(\rho) = 0$. This is an other equivalent characterization to ρ being pure. We summarize all the equivalent such characterizations in the theorem below.

Theorem 1.17.6 (Characterization of 2×2 pure states) — For a density matrix $\rho \in M_2(\mathbb{C})$, the following are equivalent:

(P1) ρ is pure,

(P2) ρ is not invertible,

(P3) $\det(\rho) = 0$,

(P4) $\operatorname{rank}(\rho) = 1$,

(P5) 0 is an eigenvalue of ρ,

(P6) 1 is an eigenvalue of ρ,

(P7) $\rho^2 = \rho$

(P8) $\operatorname{Tr}(\rho^2) = 1$

(P9) ρ is on the surface/boundary of the Bloch sphere,

(P10) the Bloch vector of ρ has norm one, and

(P11) $S(\rho) = 0$.

2 Entanglement: 2+ qubits

2.1 Introduction: Violations of Bell's inequality

Before learning about quantum mechanics or quantum information, most people believe that the world is made of real things which exist even when we don't look at them, and that these things can only be affected by other things close to it, within a reasonable amount of time. (For example, nothing moves faster than the speed of light.) These are notions of *realism* and *locality*. Taken together, the notion of *local realism* is a common belief held by most people at some point in their lives. And if they never deal with quantum things, this assumption is usually pretty good!

In the initial days of quantum mechanics [EPR35], Einstein and collaborators came up with a notion of 'spooky action at a distance' which would violate the concept of local realism. They posited that spins of particles are not random, but are governed by some 'hidden variable' that we cannot access. Later in 1964 [Bel64], Bell came up with an experiment that could falsify any hidden variable theory. He came up with an inequality, now referred to as a *Bell inequality*, which should hold under the assumptions of local realism. He then gave an experimental setup that one could perform to check whether this inequality holds in the real world.

In [FC72], Bell's prediction that his inequality would be violated in the real world came true. Numerous experiments have been carried out since providing more evidence against local realism. In Spring 2022, the Nobel Prize was awarded to Aspect, Clauser, Zeilinger "for experiments with entangled photons, establishing the violation of Bell inequalities and pioneering quantum information science,"[2] telling us that quantum entanglement is a real phenomenon in our world.

The entire premise of this module is quantum entanglement. In order to define the notion of entanglement between two particles, we discuss tensor product Hilbert spaces, and in particular, the tensor product space $\mathbb{C}^2 \otimes \mathbb{C}^2$. Not only will we see how to create entangled states, but we will use quantum entanglement as a resource for quantum communication, including the quantum teleportation protocol.

2.2 The tensor product space $\mathbb{C}^2 \otimes \mathbb{C}^2$

One qubit is represented mathematically by $\mathbb{C}^2$ with its standard inner product. For applications to quantum information, computation, and communication, we need to be able to work with more qubits. We represent 2 qubits by the vector space $\mathbb{C}^2 \otimes \mathbb{C}^2$ with its inner product, which we define below. There are several definitions we can give – some abstract and some concrete. We will make the most concrete definition possible and discuss a more abstract approach in a remark later on.

[2] Advanced information. NobelPrize.org. Nobel Prize Outreach AB 2023. https://www.nobelprize.org/prizes/physics/2022/advanced-information/

Definition 2.2.1 — Given an $m \times n$ matrix A and a $p \times q$ matrix B, the *(Kronecker) tensor product* matrix $A \otimes B$ is the $mp \times nq$ matrix obtained from A by substituting $A_{ij}B$ for the if-th entry of A.

$$A \otimes B = \begin{bmatrix} A_{11}B & \cdots & A_{n1}B \\ \ddots & & \ddots \\ A_{m1}B & \cdots & A_{mn}B \end{bmatrix}.$$

Example 2.2.2 — Here are some examples of tensor products of matrices.

$$Z \otimes X = \begin{bmatrix} 1X & 0X \\ 0X & -1X \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \qquad X \otimes Z = \begin{bmatrix} 0Z & 1Z \\ 1Z & 0Z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}$$

Exercise 2.2.3 — Compute $A \otimes B$ for the other 14 possibilities for $A, B \in \{I_2, X, Y, Z\}$.

Observe that

$$\begin{bmatrix} w \\ x \end{bmatrix} \otimes \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} w \begin{bmatrix} y \\ z \end{bmatrix} \\ x \begin{bmatrix} y \\ z \end{bmatrix} \end{bmatrix} = \begin{bmatrix} wy \\ wz \\ xy \\ xz \end{bmatrix} \in \mathbb{C}^4. \tag{2.2.4}$$

Such a 4×1 vector in $\mathbb{C}^4$ which is of the form $|\psi\rangle \otimes |\eta\rangle$ for $|\psi\rangle, |\eta\rangle \in \mathbb{C}^2$ is called a *product vector*. The space $\mathbb{C}^2 \otimes \mathbb{C}^2$ is the set of all *linear combinations* of product vectors, i.e., we can take sums of product vectors which are allowed to first be multiplied by scalars, e.g.:

$$\sum_{i=1}^{n} c_i |\psi_i\rangle \otimes |\eta_i\rangle \qquad\qquad |\psi_1\rangle, \ldots, |\psi_n\rangle, |\eta_1\rangle, \ldots, |\eta_n\rangle \in \mathbb{C}^2, \quad c_1, \ldots, c_n \in \mathbb{C}.$$

By (2.2.4), $\mathbb{C}^2 \otimes \mathbb{C}^2 \subseteq \mathbb{C}^4$.

Notation 2.2.5 — Product vectors in $\mathbb{C}^2 \otimes \mathbb{C}^2$ are sometimes denoted without the tensor product symbol as $|\psi\rangle|\eta\rangle$, or sometimes by juxtaposing the symbols in a single ket $|\psi\eta\rangle$. We will use this second notation as it involves the fewest mathematical symbols. In summary:

$$\mathbb{C}^2 \otimes \mathbb{C}^2 = \operatorname{span}\left\{ |\psi\eta\rangle \middle| |\psi\rangle, |\eta\rangle \in \mathbb{C}^2 \right\}.$$

Here, the word *span* means we take the set of all linear combinations of product vectors as explained above.

Example 2.2.6 — We calculate that

$$|00\rangle = \begin{bmatrix}1\\0\end{bmatrix} \otimes \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\begin{bmatrix}1\\0\end{bmatrix}\\0\begin{bmatrix}1\\0\end{bmatrix}\end{bmatrix} = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \qquad |01\rangle = \begin{bmatrix}1\\0\end{bmatrix} \otimes \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1\begin{bmatrix}0\\1\end{bmatrix}\\0\begin{bmatrix}0\\1\end{bmatrix}\end{bmatrix} = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$$

$$|00\rangle = \begin{bmatrix}0\\1\end{bmatrix} \otimes \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0\begin{bmatrix}1\\0\end{bmatrix}\\1\begin{bmatrix}1\\0\end{bmatrix}\end{bmatrix} = \begin{bmatrix}0\\0\\1\\0\end{bmatrix} \qquad |01\rangle = \begin{bmatrix}0\\1\end{bmatrix} \otimes \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\begin{bmatrix}0\\1\end{bmatrix}\\1\begin{bmatrix}0\\1\end{bmatrix}\end{bmatrix} = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

This means that every vector in $\mathbb{C}^4$ can be written as a linear combination of the vectors $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. Said another way, the vectors $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ *span* $\mathbb{C}^4$. Since $\mathbb{C}^2 \otimes \mathbb{C}^2 \subseteq \mathbb{C}^4$ by (2.2.4), we now see that in fact $\mathbb{C}^4 \subset \mathbb{C}^2 \otimes \mathbb{C}^2$, since each of $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ are in $\mathbb{C}^2 \otimes \mathbb{C}^2$. Thus $\mathbb{C}^2 \otimes \mathbb{C}^2 = \mathbb{C}^4$.

Note that not every element in $\mathbb{C}^4$ is a product vector.

Exercise 2.2.7 — Write down the form of an arbitrary product vector $|\psi\eta\rangle$. Then use proof by contradiction to show that $|00\rangle + |11\rangle$ is not a product vector.

Remark 2.2.8 — Some would define $\mathbb{C}^2 \otimes \mathbb{C}^2$ as a space of linear combinations of abstract symbols $|\eta\rangle \otimes |\psi\rangle$ and then say there is a map to $\mathbb{C}^4$ given by mapping the abstract tensor symbol to the Kronecker product symbol. This point of view is particularly useful when defining the tensor product of abstract Hilbert spaces, but it does not really afford that much more generality when working with concrete Hilbert spaces.

As before with $\mathbb{C}^2$, we can write bras $\langle\psi\eta|$ for the 4×1 matrix $\langle\psi| \otimes \langle\eta|$.

Definition 2.2.9 — The inner product on $\mathbb{C}^4 = \mathbb{C}^2 \otimes \mathbb{C}^2$ is given by

$$\left\langle \begin{bmatrix}w_1\\x_1\\y_1\\z_1\end{bmatrix} \middle| \begin{bmatrix}w_2\\x_2\\y_2\\z_2\end{bmatrix} \right\rangle = \begin{bmatrix}\overline{w_1} & \overline{x_1} & \overline{y_1} & \overline{z_1}\end{bmatrix} \begin{bmatrix}w_2\\x_2\\y_2\\z_2\end{bmatrix} = \overline{w_1}w_2 + \overline{x_1}x_2 + \overline{y_1}y_2 + \overline{z_1}z_2.$$

Observe that $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ is an ONB of $\mathbb{C}^2 \otimes \mathbb{C}^2$:

$$\langle ij|j\ell\rangle = \delta_{i=j}\delta_{k=\ell}$$

where $\delta_{i=j}$ is the *Kronecker delta* symbol which is 1 if $i = j$ and 0 otherwise. As before, every vector in $\mathbb{C}^2 \otimes \mathbb{C}^2$ can be written uniquely as a linear combination of $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$;

if

$$|\Psi\rangle = a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle,$$

then $a = \langle 00|\Psi\rangle$, $b = \langle 01|\Psi\rangle$, $c = \langle 10|\Psi\rangle$, and $d = \langle 11|\Psi\rangle$.

We have the following alternative formula for the inner product on $\mathbb{C}^2 \otimes \mathbb{C}^2$ which is useful for taking inner products of product vectors.

Exercise 2.2.10 — Suppose $|\psi_1\rangle, |\psi_2\rangle, |\eta_1\rangle, |\eta_2\rangle \in \mathbb{C}^2$. Use the first part of Exercise 2.2.7 to show that

$$\langle \psi_1\eta_1|\psi_2\eta_2\rangle = \langle \psi_1|\psi_2\rangle \cdot \langle \eta_1|\eta_2\rangle.$$

Exercise 2.2.11 — Find the 4×1 matrix representations of $|++\rangle, |+-\rangle, |-+\rangle, |--\rangle$ and show they form an ONB.

As with $\mathbb{C}^2$, the notion of length of a vector in $\mathbb{C}^2 \otimes \mathbb{C}^2$ is still given by $\|\psi\| = \sqrt{\langle\psi|\psi\rangle}$, and the notion of angle between two non-zero vectors is still given by the usual formula:

$$\cos(\theta) = \frac{\mathrm{Re}(\langle\psi|\eta\rangle)}{\||\psi\rangle\| \cdot \||\eta\rangle\|}.$$

Remark 2.2.12 — Larger quantum systems are represented by larger numbers of qubits. The space representing n qubits is

$$(\mathbb{C}^2)^{\otimes n} = \underbrace{\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{n \text{ copies}} = \mathbb{C}^{2^n}.$$

We will use this space later on when we need to.

Exercise 2.2.13 — Show that the vectors $\frac{1}{\sqrt{2}}(|+-\rangle - |-+\rangle)$,$|++\rangle$, $\frac{1}{\sqrt{2}}(|+-\rangle + |-+\rangle)$, $|--\rangle$ form an ONB. The first state is known as the "singlet" state where as the last three are known as the "triplet" states.

Exercise 2.2.14 — Show that the tensor product of two vectors is a bilinear operation. That is to say that for all vectors $|a\rangle, |b\rangle, |c\rangle, |d\rangle \in \mathbb{C}^2$ and $\alpha, \beta, \gamma, \delta \in \mathbb{C}$,

$$(\alpha|a\rangle + \beta|b\rangle) \otimes (\gamma|c\rangle + \delta|d\rangle) = \alpha\gamma|a\rangle \otimes |c\rangle + \alpha\delta|a\rangle \otimes |d\rangle + \beta\gamma|b\rangle \otimes |c\rangle + \beta\delta|b\rangle \otimes |d\rangle. \tag{2.2.15}$$

2.3 Tensor products of operators $M_2(\mathbb{C}) \otimes M_2(\mathbb{C})$

Just as the 2×2 matrices in $M_2(\mathbb{C})$ act as operators on $\mathbb{C}^2$, 4×4 matrices in $M_4(\mathbb{C})$ act as operators on $\mathbb{C}^4 = \mathbb{C}^2 \otimes \mathbb{C}^2$. Given $A, B \in M_2(\mathbb{C})$, we already discussed the meaning of the operator $A \otimes B$, which is an element of $M_4(\mathbb{C})$; we will call such an operator a *simple tensor*. We write $M_2(\mathbb{C}) \otimes M_2(\mathbb{C})$ for the subset of $M_4(\mathbb{C})$ spanned by operators of the simple tensors $A \otimes B$.

Proposition 2.3.1 — We have an equality of sets $M_2(\mathbb{C}) \otimes M_2(\mathbb{C}) = M_4(\mathbb{C})$.

Proof. We need to show we can write any 4×4 matrix as a linear combination of operators of the form $A \otimes B$. Let $E_{ij} \in M_2(\mathbb{C})$ be the matrix with a 1 in the ij-th entry and all other entries zero. Then $E_{ij} \otimes E_{k\ell}$ is a 4×4 matrix with exactly one non-zero entry, and that entry is equal to 1. The 16 different ways of choosing i, j, k, ℓ (which can all be either 1 or 2) give the 16 matrices in $M_4(\mathbb{C})$ with exactly one non-zero entry equal to 1. Thus we can write every matrix in $M_4(\mathbb{C})$ as a linear combination of the simple tensors $E_{ij} \otimes E_{k\ell}$. ∎

Exercise 2.3.2 — Compute the 4×4 matrices corresponding to each of the 16 simple tensors $E_{ij} \otimes E_{k\ell}$.

As with $\mathbb{C}^4$, not every 4×4 matrix in $M_4(\mathbb{C})$ is a simple tensor.

Exercise 2.3.3 — Let $S \in M_4(\mathbb{C})$ be the matrix corresponding to the swap operation $|\psi\eta\rangle \mapsto |\eta\psi\rangle$.

(1) Compute the 4×4 matrix S.
Hint: The i-th column of S is $S|e_i\rangle$ where $|e_i\rangle$ is the i-th standard basis element.

(2) Show that S is not of the form $A \otimes B$ for $A, B \in M_2(\mathbb{C})$.

Exercise 2.3.4 — Show that $(A \otimes B)(C \otimes D) = AC \otimes BD$ for $A, B, C, D \in M_2(\mathbb{C})$. Deduce that $(I_2 \otimes B)(A \otimes I_2) = (A \otimes I_2)(I_2 \otimes B)$.

As before, we can still take the conjugate transpose of a 4×4 matrix.

Exercise 2.3.5 — Show that $(A \otimes B)^* = A^* \otimes B^*$.

It is easy to calculate the eigenvalues and eigenvectors of simple tensors $A \otimes B$; we simply calculate the individual eigenvalues and eigenvectors for A and B separately, and then we 'multiply': take the products of the eigenvalues and the tensor products of their corresponding eigenvectors. Matrices which are non-trivial linear combinations of simple

tensors are more complicated; we will explain how to deal with them later on when we need to.

We end this section with an important theorem for quantum information. If we want to perform computations with qubits or transmit quantum information in some communication protocol, we will need a way to copy qubits. The operators that do not destroy any quantum information are the unitary operators. So it would be great if there were a unitary operator $U \in M_2(\mathbb{C}) \otimes M_2(\mathbb{C})$ such that for any state $|\psi\rangle \in \mathbb{C}^2$, there is a phase $\alpha_\psi \in \mathbb{C}$ (which is allowed to depend on ψ!) such that

$$U|\psi 0\rangle = \alpha_\psi |\psi\psi\rangle.$$

The phase α_ψ is to account for the fact that states are really only well-defined up to a phase. We start with $|\psi 0\rangle$ on the right because we have one qubit of information $|\psi\rangle$ we want to copy, and our second qubit is allowed to be any fixed ancillary qubit that we know how to prepare. (Without loss of generality, we can take it to be $|0\rangle$, but we could fix it as $|\eta\rangle$ for any other choice of ancillary qubit.)

Unfortunately this universal unitary U does not exist!

Theorem 2.3.6 (No Cloning) — There is no unitary $U \in M_4(\mathbb{C})$ such that for any state $|\psi\rangle \in \mathbb{C}^2$, there is a phase $\alpha_\psi \in \mathbb{C}$ such that

$$U|\psi 0\rangle = \alpha_\psi |\psi\psi\rangle.$$

Proof. Suppose for contradiction that such a U exists, and let α, β be phases such that

$$U|00\rangle = \alpha|00\rangle \qquad \text{and} \qquad U|+0\rangle = \beta|++\rangle.$$

Since $|\alpha| = 1$, rearranging the first equation and taking conjugate transpose, we have $\langle 00| = \langle 00|U^*\alpha$. Since $|\beta| = 1$, rearranging the second equation, we have $|++\rangle = \overline{\beta}U|+0\rangle$. Now we calculate:

$$\frac{1}{2} = \langle 0|+\rangle^2 = \langle 00|++\rangle = \langle 00|U^*\alpha\overline{\beta}U|+0\rangle = \alpha\overline{\beta}\langle 00|\underbrace{U^*U}_{I_2}|+0\rangle = \alpha\overline{\beta}\langle 00|+0\rangle = \alpha\overline{\beta}\langle 0|+\rangle = \frac{\alpha\overline{\beta}}{\sqrt{2}}.$$

But $|\alpha\overline{\beta}| = 1$, so taking modulus of both sides, we get a contradiction. ∎

Remark 2.3.7 — The above proof actually holds replacing $\mathbb{C}^2$ with $\mathbb{C}^n$ for any arbitrary n, so that no such $U \in M_{n^2}(\mathbb{C})$ exists. Indeed, $|0\rangle$ and $|+\rangle$ can be chosen to be any two length one vectors in $\mathbb{C}^n$ such that $\langle 0|+\rangle = 1/\sqrt{2}$, like $|0\rangle = |e_1\rangle$ and $|+\rangle = \frac{1}{\sqrt{2}}(|e_1\rangle + |e_2\rangle)$.

Remark 2.3.8 — The space of operators on the n qubit space $\mathbb{C}^{2^n}$ is $M_2(\mathbb{C})^{\otimes n} = M_{2^n}(\mathbb{C})$.

Exercise 2.3.9 — Is it true that for any $S \in M_4(\mathbb{C})$, the singlet state $|\psi\rangle = \frac{1}{\sqrt{2}}(|+-\rangle - |-+\rangle)$ and the triplet state $|\phi\rangle = \frac{1}{\sqrt{2}}(|+-\rangle + |-+\rangle)$ give

$$\langle\psi|S|\psi\rangle = \langle\phi|S|\phi\rangle? \tag{2.3.10}$$

If true, explain why. If false, find a counter example $S \in M_4(\mathbb{C})$ where S can be expressed in terms of sums and tensor products of Pauli matrices.

Exercise 2.3.11 (No Anti-Cloning) — Show that there is no unitary $U \in M_4(\mathbb{C})$ such that for any state $|\psi\rangle \in \mathbb{C}^2$, there is a phase $\alpha_\psi \in \mathbb{C}$ such that

$$U|\psi\psi\rangle = \alpha_\psi|\psi 0\rangle. \tag{2.3.12}$$

2.4 Pure states: entangled vs. product states

For one qubit, there were only 2 kinds of states: pure and mixed. With 2+ qubits, these types of states can be further divided by whether they are separable or entangled. In this section, we treat only pure states in $\mathbb{C}^2 \otimes \mathbb{C}^2$, and we will discuss mixed states later.

Definition 2.4.1 — A pure state $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ is called

- a *product state* if $|\Psi\rangle = |\psi\eta\rangle$ for some $|\psi\rangle, |\eta\rangle \in \mathbb{C}^2$.
- *entangled* if it is not of the form $|\psi\eta\rangle$.

As before, states are only defined up to phase. Alternatively, we could discuss the density matrices $|\Psi\rangle\langle\Psi| \in M_4(\mathbb{C})$ instead to get rid of this phase ambiguity.

We already saw in Exercise 2.2.7 that $|00\rangle + |11\rangle$ is not a product vector; normalizing provides an example of an entangled state. Here is another example.

Example 2.4.2 (Singlet states) — The *singlet state* is

$$\frac{1}{\sqrt{2}}(|e_1e_2\rangle - |e_2e_1\rangle)$$

for some ONB $\{|e_i\rangle, |e_j\rangle\}$ of $\mathbb{C}^2$. In Exercise 2.4.3 below, you will show that the singlet

state is *independent* of the choice of ONB! The singlet state describes two electrons whose spins are paired along some axis.

Exercise 2.4.3 — Consider a state $|e_1\rangle = \alpha|0\rangle + \beta|1\rangle \in \mathbb{C}^2$.

1. Show that $|e_2\rangle := -\overline{\beta}|0\rangle + \overline{\alpha}|1\rangle$ extends $|e_1\rangle$ to an ONB.
2. Show that $\frac{1}{\sqrt{2}}(|e_1e_2\rangle - |e_2e_1\rangle)$ is equal to $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$.
3. Deduce that the singlet state is independent of the choice of ONB.

Remark 2.4.4 — There is a deep reason that the singlet state is independent of the choice of ONB coming from the advanced mathematical area of *representation theory*. The main idea is that the *special unitary group* $SU(2)$ of unitary 2×2 matrices with determinant 1 acts on $\mathbb{C}^2$, and thus acts diagonally on $\mathbb{C}^2 \otimes \mathbb{C}^2$. This representation breaks up as a 1D representation spanned by the singlet state and a 3D representation. That the singlet state is independent of the choice of ONB corresponds to the fact that there is only one copy of this 1D representation inside $\mathbb{C}^2 \otimes \mathbb{C}^2$.

Example 2.4.5 (Spooky action at a distance) — Suppose Alice and Bob have a pair of photons or electrons in an entangled state, and separate them very far apart. To be as concrete as possible, we will assume the state is the singlet state

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle),$$

where the first tensor factor represents Alice's qubit and the second tensor factor represents Bob's qubit.

Now suppose Alice observes spin in the z-direction by measuring Pauli Z on her qubit. We will see in §2.5 below that this corresponds to measuring the observable $Z \otimes I_2$, and depending on Alice's measurement, the state will update to the (renormalized) *projection* of $|\Psi\rangle$ to the corresponding eigenspace.

In more detail, Alice will get an answer of $+1$ or -1, each with probability $1/2$. If Alice measures $+1$, then the state will update to the projection of $|\Psi\rangle$ to the $+1$ eigenspace of $Z \otimes I_2$, which is exactly $|01\rangle$. If Alice measures $+1$, then the state will update to the projection of $|\Psi\rangle$ to the -1 eigenspace of $Z \otimes I_2$, which is exactly $|10\rangle$.

Now observe that even though Bob has done nothing, his qubit is now in a state which is *completely determined* by Alice's measurement! It seems as though information has been transmitted instantaneously from Alice to Bob, faster than the speed of light,

which is impossible!

However, no information has been transferred; Bob is working with *incomplete information* about the state. He has no idea which value Alice measured, and so he cannot know his state unless Alice transmits information (which moves only as quickly as the speed of light) to Bob. Bob's best approximation of his state after Alice performs her measurement is the *mixed state*

$$\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = I_2.$$

This is the state with maximum entropy, so Bob still has as little information as possible about the state as possible until Alice sends him an update.

Example 2.4.6 (EPR paradox) — Example 2.4.5 was pointed out by Einstein-Podolsky-Rosen in [EPR35], and is part of what is now called the *EPR paradox*. Alice can also observe spin in the x-direction by measuring Pauli X on her qubit, which corresponds to the observable $X \otimes I_2$. Now we can also write the singlet state

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

and both $|-+\rangle$ and $|+-\rangle$ are eigenvectors of $(X \otimes I_2)$ corresponding to eigenvalues -1 and $+1$ respectively. This means that Alice will get $+1$ or -1 each with probability $1/2$, after which the state is updated to either $|+-\rangle$ or $|-+\rangle$ respectively. Again, even though Bob has done nothing, his qubit is in a state completely determined by Alice's measurement.

The supposed paradox arises from the same behavior when measuring either of the observables Z or X, which do not commute. Indeed, since Z and X do not commute, they satisfy the *Uncertainty Principle* from the previous module, saying that we cannot simultaneously know the z-spin and x-spin of an electron. Now assuming local realism so that the two qubits cannot 'communicate' faster than the speed of light, the state of Bob's electron cannot be instantaneously affected by Alice measuring in either the z- or x-direction. This means Alice could measure in the z-direction, and before the state has collapsed, Bob could measure in the x-direction, so that both the z- and x-spins of Bob's electrons would be determined simultaneously, giving a contradiction.

Today, we have observed the failure of local realism, in particular in violations of Bell's inequalities, which resolves the EPR paradox. We will study Bell's inequalities more in §2.8 below.

Example 2.4.7 (Bell states) — Certain inequalities called *Bell's inequalities* can be proven by assuming local realism, and *Bell states* violate these inequalities. These are

also referred to as *EPR states* as the same 'spooky' phenomenon can be observed using any of these states. The *Bell basis* of $\mathbb{C}^2 \otimes \mathbb{C}^2$ is given by

$$\left\{ \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle),\ \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle),\ \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle),\ \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \right\}.$$

All four of these states are entangled. The first basis element was our first example of an entangled state, and the last basis element is the singlet state. We sometimes call these states $|\beta_{00}\rangle, |\beta_{01}\rangle, |\beta_{10}\rangle, |\beta_{11}\rangle$.

Observe that we may obtain $|\beta_{ij}\rangle$ from $|\beta_{00}\rangle$ by applying a 'correction' simple tensor using Pauli matrices:

$$|\beta_{ij}\rangle = (C_{ij} \otimes I_2)|\beta_{00}\rangle \qquad \begin{array}{c|c} ij & C_{ij} \\ \hline 00 & I_2 \\ \hline 01 & X \\ \hline 10 & Z \\ \hline 11 & ZX \end{array} \tag{2.4.8}$$

Indeed, for $|\beta_{11}\rangle$, we have

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \overset{X\otimes I_2}{\longmapsto} \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \overset{Z\otimes I_2}{\longmapsto} \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

Exercise 2.4.9 — Consider the ONB $|e_1\rangle = \alpha|0\rangle + \beta|1\rangle$ and $|e_2\rangle := -\overline{\beta}|0\rangle + \overline{\alpha}|1\rangle$ as in Exercise 2.4.3. Define $|\overline{e_1}\rangle = \overline{\alpha}|0\rangle + \overline{\beta}|1\rangle$ and $|\overline{e_2}\rangle = -\beta|0\rangle + \alpha|1\rangle$. Show that $\frac{1}{\sqrt{2}}(|e_1\overline{e_1}\rangle + |e_2\overline{e_2}\rangle)$ is equal to $|\beta_{00}\rangle$.

Remark 2.4.10 — Again, there is a deep mathematical reason that the Bell state can be written with any ONB in this way. The main idea here is that $\mathbb{C}^2 \otimes \overline{\mathbb{C}^2}$ is canonically isomorphic to $M_2(\mathbb{C})$, and under this isomorphism,

$$\frac{1}{\sqrt{2}}(|e_1\overline{e_1}\rangle + |e_2\overline{e_2}\rangle) \longmapsto \frac{1}{\sqrt{2}}(|e_1\rangle\langle e_1| + |e_2\rangle\langle e_2| \in M_2(\mathbb{C}).$$

But by (3.2.1), this operator is exactly I_2, regardless of the choice of ONB.

Exercise 2.4.11 — There is an interpretation of Pauli matrices as operators which

rotate space. Determine which Bell state does not change (modulo a minus sign) after applying $X \otimes X$, $Y \otimes Y$, or $Z \otimes Z$. This state is "rotationally invariant."

Exercise 2.4.12 — As was seen in Example 2.4.7, any Bell state can be generated from $|\beta_{00}\rangle$. Show that any Bell state can also be generated from $|\beta_{11}\rangle$ by finding C'_{ij} such that

$$|\beta_{ij}\rangle = (C'_{ij} \otimes I_2)|\beta_{11}\rangle. \tag{2.4.13}$$

This exercise can be done starting with any Bell state.

2.5 Measurement and local measurement

We saw in §2.2 that a 2-qubit system is represented by $\mathbb{C}^2 \otimes \mathbb{C}^2$. This is true no matter how far apart the qubits are from one another. Often in quantum information, we will have a pair of qubits, one controlled by Alice and one controlled by Bob, who are separated by some distance. Alice and Bob can each measure an observable on their respective qubits locally which corresponds to measuring a simple tensor, and the outcome is either determined or probabilistic depending on our axioms of observables. There are also non-local measurements corresponding to observables which are not simple tensors. The axioms for measurements for 2 qubit systems is entirely analogous to those for 1 qubit systems.

Axioms 2.5.1 (Axioms of quantum observables II) — Consider a quantum system consisting of 2 qubits.

($Q^{II}1$) A quantum state is a vector state $|\Psi\rangle\langle\Psi| \in M_4(\mathbb{C})$ where $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ is a state vector.

($Q^{II}2$) A quantum observable is a self-adjoint operator $C \in M_4(\mathbb{C})$.

($Q^{II}3$) There are two possibilities when measuring the observable $C \in M_4(\mathbb{C})$:

- If $|\Psi\rangle$ is an eigenvector of C (which is a property unchanged multiplying $|\Psi\rangle$ by a phase), then the outcome of the measurement is the corresponding eigenvalue, and the state remains unchanged.
- If $|\Psi\rangle$ is not an eigenvector of C (which is again a property unchanged multiplying $|\Psi\rangle$ by a phase), the outcome of the measurement is probabilistic, and occurs as follows. We denote the distinct eigenvalues of C by $\lambda_1, \ldots, \lambda_k$ where $k \in \{2, 3, 4\}$ (if C is self-adjoint and has only one eigenvalue, then C is a scalar matrix, so all vectors are eigenvectors). We will see in Theorem

2.7.7 below that we can write $|\Psi\rangle$ *uniquely* as

$$|\Psi\rangle = \underbrace{a_1|\Psi_1\rangle + a_2|\Psi_2\rangle + \cdots}_{k \text{ terms}}$$

up to at most 4 terms where each $|\Psi_i\rangle$ is an eigenstate corresponding to eigenvalue λ_i. The outcome of measuring C in state $|\Psi\rangle$ is λ_i with probability $|a_i|^2 = |\langle\Psi_i|\Psi\rangle|^2$. After measuring λ_i, the state *changes* to $|\Psi_i\rangle$.

Example 2.5.2 — When C has 4 distinct eigenvalues $\lambda_1, \ldots, \lambda_4$, the axioms above are entirely similar to the case of 2 distinct eigenvalues in $\mathbb{C}^2$. There is an ONB $\{|e_1\rangle, |e_2\rangle, |e_3\rangle, |e_4\rangle\}$ of corresponding eigenvectors, and outcome of measuring C in state $|\Psi\rangle$ is λ_i with probability $|\langle e_i|\Psi\rangle|^2$ for $i = 1, 2, 3, 4$, after which the state changes to $|e_i\rangle$.

Example 2.5.3 — Suppose $A, B \in M_2(\mathbb{C})$ are observables with eigenvalues λ_1, λ_2 and μ_1, μ_2 respectively. Let $\{|e_1\rangle, |e_2\rangle\}$ and $\{|f_1\rangle, |f_2\rangle\}$ be respective ONBs of eigenstates for A, B. Then $A \otimes B$ has eigenvalues $\lambda_1\mu_1$, $\lambda_1\mu_2$, $\lambda_2\mu_1$, $\lambda_2\mu_2$ corresponding to eigenstates $|e_1f_1\rangle$, $|e_1f_2\rangle$, $|e_2f_1\rangle$, $|e_2f_2\rangle$, respectively. Observe these eigenstates form an ONB by Exercise 2.2.10. If the four products are distinct, then we may proceed as in the previous example.

There are many examples of $A \otimes B$ in these notes where we have $\lambda_i\mu_j = \pm 1$, each with a 2-dimensional eigenspace, including $A, B \in \{I_2, X, Y, Z\}$ (except $A = B = I_2$) and other examples in §2.8 below on the CHSH inequality.

Example 2.5.4 — We will observe $Z \otimes I_2$ in several different states to see how the process outlined in Axioms 2.5.1 works. First, by Example 2.5.3, the standard computational ONB is an ONB of eigenstates for $Z \otimes I_2$. We can separate them into two pairs depending on the corresponding eigenvalue: $|00\rangle, |01\rangle$ have eigenvalue $+1$ and $|10\rangle, |11\rangle$ have eigenvalue -1. This means that we have two *eigenspaces* (see Example 2.6.3 below):

$$E_1 = \text{span}\{|00\rangle, |01\rangle\} \qquad E_{-1} = \text{span}\{|10\rangle, |11\rangle\}.$$

- Suppose

$$|\Psi\rangle = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Then $|\Psi\rangle \in E_1$ as it is a linear combination of $|00\rangle$ and $|01\rangle$. Thus it is an eigenstate corresponding to $\lambda = 1$, When we observe $Z \otimes I_2$, we measure $+1$ and the state

remains unchanged.

- Suppose $|\Psi\rangle$ is the singlet state $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. Observe we have written $|\Psi\rangle$ as a linear combination of eigenstates: $|01\rangle \in E_1$ and $|10\rangle \in E_{-1}$. When we observe $Z \otimes I_2$, we measure ± 1 each with probability $|\frac{1}{\sqrt{2}}|^2 = \frac{1}{2}$.
- Suppose

$$|\Psi\rangle = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{3}}(|00\rangle + |01\rangle + |10\rangle).$$

We have written $|\Psi\rangle$ as a linear combination of eigenstates for $Z \otimes I_2$, but two of these eigenvalues are not distinct. We must *combine* these into one eigenstate to get the unique representation in order to determine what happens when we measure. Observe that

$$|\Psi\rangle = \frac{1}{\sqrt{3}}(|00\rangle + |01\rangle + |10\rangle) = \frac{\sqrt{2}}{\sqrt{3}} \left(\frac{|00\rangle + |01\rangle}{\sqrt{2}} \right) + \frac{1}{\sqrt{3}}|10\rangle.$$

Thus when we measure, we get $+1$ with probability $\frac{2}{3}$ after which the state changes to $\frac{1}{\sqrt{2}}(|00\rangle + |01\rangle)$, or we get -1 with probability $\frac{1}{3}$ after which the state changes to $|10\rangle$.

The above example is an instance of a *local measurement*, where we are only observing one qubit in our 2-qubit state. In the setting where Alice and Bob share a quantum state $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ and each control one qubit in this 2 qubit quantum system, Alice measuring observable A on her qubit means that she is measuring the operator $A \otimes I_2$ in state $|\Psi\rangle$. Bob measuring observable B on his qubit means he is measuring the operator $I_2 \otimes B$ in state $|\Psi\rangle$.

Definition 2.5.5 (local measurement) — The observable C is localized to the first qubit if $C = A \otimes I_2$ for some $A \in M_2(\mathbb{C})$; it is localized to the second qubit if $C = I_2 \otimes B$ for some $B \in M_2(\mathbb{C})$.

Remark 2.5.6 — We saw in the first module that if $A, B \in M_2(\mathbb{C})$ are commuting observables, there is an ONB of common eigenstates for A and B. This is also true for observables in $M_n(\mathbb{C})$ for any n. In particular, for the commuting local observables $A \otimes I_2$ and $I_2 \otimes B$, there is an ONB of common eigenstates.

As seen above in Example 2.5.3, when $A \in M_2(\mathbb{C})$ is an observable with 2 distinct eigenvalues λ_1, λ_2 and corresponding ONB of eigenvectors $\{|e_1\rangle, |e_2\rangle\}$, $A \otimes I_2$ also has 2 distinct eigenvalues λ_1, λ_2, and choosing any ONB $\{|f_1\rangle, |f_2\rangle\}$ of $\mathbb{C}^2$, $\{|e_1 f_1\rangle, |e_1 f_2\rangle\}$ is an

ONB of eigenstates for λ_1 and $\{|e_2 f_1\rangle, |e_2 f_2\rangle\}$ is an ONB of eigenstates for λ_2. The outcome of the measurement is λ_i with probability

$$|\langle e_i f_1|\Psi\rangle|^2 + |\langle e_i f_2|\Psi\rangle|^2.$$

In §2.11 below, we will connect this quantity above to the *reduced density matrix* of $|\Psi\rangle$ and the *partial trace*.

Exercise 2.5.7 — What are the possible measurement outcomes for the observable $(X + Z) \otimes Y$? What are their probabilities for the Bell state $|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$?

Exercise 2.5.8 — Suppose that you and another student of quantum information are each given one qubit from a Bell state. The Bell state has a probability of 1/2 of being $|\beta_{00}\rangle$, 1/3 of being $|\beta_{01}\rangle$, 7/12 of being $|\beta_{10}\rangle$, and 5/12 of being $|\beta_{11}\rangle$. You measure the observable X on your qubit and get the value +1. What is the probability that your friend measures X and gets +1 on their qubit?

2.6 Subspaces

In Axioms 2.5.1 above, we asserted that $|\Psi\rangle$ could be written uniquely as a linear combination of eigenstates corresponding to distinct eigenvalues. The next two sections provide the necessary mathematics to back up this claim. To do so, we will need the notion of a subspace (this section) and the operator which projects onto the subspace (the next section). Some of this material draws on the supplementary Appendix A at the end on more advanced linear algebra. In this section, we will work in $\mathbb{C}^n$, as nothing we say here is special to the case $n = 4$.

Definition 2.6.1 — A *subspace* of $\mathbb{C}^n$ is a non-empty collection S of vectors such that:

(S1) S is closed under addition, i.e., $|\Psi\rangle, |\Phi\rangle \in S$ implies $|\Psi\rangle + |\Phi\rangle \in S$.

(S2) S is closed under scalar multiplication, i.e., $|\Psi\rangle \in S$ and $\lambda \in \mathbb{C}$ implies $\lambda|\Psi\rangle \in S$.

Example 2.6.2 (Trivial subspaces) — The trivial subspaces of $\mathbb{C}^n$ are $V = \mathbb{C}^n$ and $V = \{\mathbf{0}\}$.

Example 2.6.3 (Eigenspaces) — If $A \in M_n(\mathbb{C})$ has eigenvalue $\lambda \in \mathbb{C}$, then the corre-

sponding *eigenspace* is

$$E_\lambda := \{|\psi\rangle \in \mathbb{C}^n | A|\psi\rangle = \lambda|\psi\rangle\}.$$

Eigenspaces are special examples of null spaces/kernels of matrices, which appear in the next example.

Example 2.6.4 (Null space/kernel and image) — Suppose A is an $m \times n$ matrix. The *null space/kernel* of A is

$$\ker(A) := \{|\psi\rangle \in \mathbb{C}^n | A|\psi\rangle = \mathbf{0}\}.$$

The *image* of A is

$$\mathrm{im}(A) := \{|\psi\rangle \in \mathbb{C}^m | \text{there is a } |\eta\rangle \in \mathbb{C}^n \text{ with } |\psi\rangle = A|\eta\rangle\} = \{A|\eta\rangle || \eta\rangle \in \mathbb{C}^n\} = A\mathbb{C}^n.$$

Observe that if $A \in M_n(\mathbb{C})$ is a square matrix with eigenvalue λ, then $E_\lambda = \ker(A - \lambda I_n)$.

Exercise 2.6.5 — Verify that $\ker(A) \subseteq \mathbb{C}^n$ and $\mathrm{im}(A) \subseteq \mathbb{C}^m$ are subspaces.

Example 2.6.6 (Span) — Any collection of vectors $S \subset \mathbb{C}^n$ generates or *spans* a subspace called

$$\begin{aligned}\mathrm{span}(S) :&= \{\text{(finite) linear combinations of elements of } S\} \\ &= \left\{ \sum_{i=1}^{k} c_i |\psi_i\rangle \,\middle|\, |\psi_i\rangle \in S \text{ and } c_i \in \mathbb{C} \text{ for } i = 1, \ldots, k \text{ where } k \geq 0 \right\}.\end{aligned}$$

Exercise 2.6.7 — Verify that $\mathrm{span}(S)$ is a subspace.

Example 2.6.8 (Orthogonal complement) — Given a subspace $V \subseteq \mathbb{C}^n$, we get another subspace $V^\perp$ called its *orthogonal complement* defined by

$$V^\perp := \{|\psi\rangle \in \mathbb{C}^n | \langle\eta|\psi\rangle = 0 \text{ for all } |\eta\rangle \in V\}.$$

Exercise 2.6.9 — Verify that $V^\perp$ is a subspace.

Exercise 2.6.10 — Suppose $S \subset \mathbb{C}^n$ is some collection of vectors and $V = \text{span}(S)$. Show that $|\psi\rangle \in V^\perp$ if and only if $\langle \eta | \psi \rangle = 0$ for all $|\eta\rangle \in S$.

Facts 2.6.11 — Every subspace $V \subseteq \mathbb{C}^n$ admits an ONB. All ONBs have the same size, and this size is called $\dim(V)$, the *dimension* of V. Moreover, every ONB for V can be extended to an ONB for $\mathbb{C}^n$. We will prove this in detail in Corollaries A.9.3 and A.9.4.

Remark 2.6.12 — In practical applications, e.g., to construct the projection onto V, we need to be able to actually find an ONB of V rather than just know one exists. Typically, this is done as follows:

- First, set up a matrix A corresponding to some system of linear equations defining V. For example, V could be $\ker(A)$ or $\text{im}(A)$ for some matrix.
- Perform the Gaussian Elimination Algorithm A.3.14 on A to obtain a matrix $\text{ref}(A)$ in row echelon form.
- Analyze $\text{ref}(A)$ (e.g., use Algorithm A.6.5 for $\text{im}(A)$) to obtain a basis for V.
- Perform the Gram-Schmidt Algorithm A.9.1 on this basis to make in orthonormal.

In practice, one can perform all these operations in some kind of computer algebra system, like `Mathematica` or `Python` using the `numpy.linalg` library.

Exercise 2.6.13 — Consider the collection of all entangled two qubit states. Do they form a subspace? Does the collection of all unentangled two qubit states form a subspace?

Exercise 2.6.14 — Find a value of $\alpha \in \mathbb{C}$ such that $X + \alpha Z$ has a non-trivial kernel.

2.7 Projections

Definition 2.7.1 — Suppose $V \subset \mathbb{C}^n$ is a subspace, and let $\{|e_1\rangle, \ldots, |e_k\rangle\}$ be an ONB of V. The *projection onto* V is the operator

$$P_V := |e_1\rangle\langle e_1| + |e_2\rangle\langle e_2| + \cdots |e_k\rangle\langle e_k| = \sum_{i=1}^{k} |e_i\rangle\langle e_i|.$$

It is independent of the choice of ONB. For every $|\psi\rangle \in V$, $P_V|\psi\rangle = |\psi\rangle$, and for every $|\eta\rangle \in V^\perp$, $P_V|\eta\rangle = 0$.

In fact, the above definition is *independent* of the choice of ONB, as we prove in the following proposition.

Proposition 2.7.2 — The vector $P_V|\psi\rangle$ is the unique closest vector in V to $|\psi\rangle \in \mathbb{C}^n$.

Proof. For our chosen ONB $\{|e_1\rangle, \ldots, |e_k\rangle\}$ of V, we may extend it to an ONB $\{|e_1\rangle, \ldots, |e_n\rangle\}$ of $\mathbb{C}^n$ by Corollary A.9.4. Given an arbitrary vector $\sum_{i=1}^k b_i|e_i\rangle \in V$, its distance squared to $|\psi\rangle = \sum_{i=1}^n a_i|e_i\rangle$ is

$$\left\|\sum_{i=1}^n a_i|e_i\rangle - \sum_{i=1}^k b_i|e_i\rangle\right\|^2 = \sum_{i=1}^k |a_i - b_i|^2 + \sum_{i>k}^n |a_i|^2 \geq \sum_{i>k}^n |a_i|^2 = \left\|\sum_{i=1}^n a_i|e_i\rangle - \sum_{i=1}^k a_i|e_i\rangle\right\|^2.$$

Hence the vector $P_V|\psi\rangle = \sum_{i=1}^k a_i|e_i\rangle \in V$ uniquely minimizes the distance to $|\psi\rangle$. ∎

Proposition 2.7.3 — Suppose $V \subseteq \mathbb{C}^n$ is a subspace. Every vector $|\psi\rangle \in \mathbb{C}^n$ can be written uniquely as

$$|\psi\rangle = |\psi_V\rangle + |\psi_{V^\perp}\rangle$$

where $|\psi_V\rangle \in V$ and $|\psi_{V^\perp}\rangle \in V^\perp$.

Proof. Suppose that $|\psi\rangle \in \mathbb{C}^n$ can be written as

$$|\psi\rangle = |\psi_V\rangle + |\psi_{V^\perp}\rangle$$

with $|\psi_V\rangle \in V$ and $|\psi_{V^\perp}\rangle \in V^\perp$. Let $\{|e_1\rangle, \ldots |e_k\rangle\}$ be an ONB for V, and define P_V as above. Applying P_V to both sides, we see that

$$P_V|\psi\rangle = P_V|\psi_V\rangle + P_V|\psi_{V^\perp}\rangle = |\psi_V\rangle,$$

and thus there is only one choice for $|\psi_V\rangle$. Moreover, we then must have $|\psi_{V^\perp}\rangle = |\psi\rangle - |\psi_V\rangle$, which is completely determined. It remains to prove that defining $|\psi_{V^\perp}\rangle := |\psi\rangle - |\psi_V\rangle$ gives a vector in $V^\perp$. To show this, it suffices to prove it is orthogonal to each $|e_i\rangle$ by Exercise 2.6.10:

$$\langle e_i|\psi_{V^\perp}\rangle = \langle e_i|\psi\rangle - \langle e_i|\psi_V\rangle = \langle e_i|\psi\rangle - \langle e_i|P_V|\psi\rangle = \langle e_i|\psi\rangle - \langle e_i|\psi\rangle = 0.$$

Here, we used that $\langle e_i|P_V = \langle e_i|$ by taking adjoints of the identity $P_V|e_i\rangle = |e_i\rangle$. ∎

Exercise 2.7.4 — Show that $I_n - P_V$ is the projection onto $V^\perp$.

Definition 2.7.5 — A *partition of unity* is a collection of projections $P_1, \ldots, P_k$ which are *mutually orthogonal* ($P_i P_j = 0$ if $i \neq j$) and sum to the identity: $\sum_{i=1}^k P_i = I_n$.

An *orthogonal direct sum decomposition* of $\mathbb{C}^n$ is a collection of subspaces $V_1, \ldots, V_k \subseteq \mathbb{C}^n$ which are mutually orthogonal ($\langle \psi_I | \psi_j \rangle = 0$ whenever $|\psi_i\rangle \in V_i$ and $|\psi_j\rangle \in V_j$ with $i \neq j$) and sum to $\mathbb{C}^n$ (every $|\psi\rangle \in \mathbb{C}^n$ can be written as a sum

$$|\psi\rangle = |\psi_1\rangle + \cdots + |\psi_k\rangle$$

with $|\psi_i\rangle \in V_i$ for all i).

There is a one-to-one correspondence between partitions of unity and orthogonal direct sum decompositions of $\mathbb{C}^n$. Given a partition of unity, we get an orthogonal direct sum decomposition by setting $V_i := P_i \mathbb{C}^n = \mathrm{im}(P_i)$. Given an orthogonal direct sum decomposition of $\mathbb{C}^n$, we get a partition of unity by setting $P_i := P_{V_i}$. Passing between these points of view is extremely helpful in understanding the general Spectral Theorem 2.7.7 for $\mathbb{C}^n$ below. We will only state the following theorem for now; we may provide a self-contained proof in an optional section later on.

Example 2.7.6 — We can now use the language of subspaces and projections to give a unified rule for measuring quantum observables $C \in M_4(\mathbb{C})$ in our 2-qubit state $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$, making Axioms 2.5.1 more precise. Denote the distinct eigenvalues of C by $\lambda_1, \ldots, \lambda_k$, the corresponding eigenspaces of C by $E_1, \ldots, E_k$, and the corresponding projections onto these eigenspaces by $P_1, \ldots, P_k$. The outcome of measuring C in state $|\Psi\rangle$ is λ_i with probability $\langle \Psi | P_i | \Psi \rangle$, which is called a *projective measurement*. After measuring λ_i, the state *changes* its projection onto E_i, renormalized to have unit length. That is, $|\Psi\rangle$ changes to

$$\frac{1}{\| P_i |\Psi\rangle \|} \cdot P_i |\Psi\rangle.$$

Note that the denominator cannot be zero as $\| P_i |\Psi\rangle \|^2 = \langle \Psi | P_i^2 | \Psi \rangle = \langle \Psi | P_i | \Psi \rangle$, which had to be non-zero in order to observe λ_i!

Theorem 2.7.7 ($n \times n$ Spectral Theorem) **—** For an $n \times n$ matrix A, the following are equivalent:

(Sp1) There is an ONB consisting of eigenvectors of A.

(Sp2) The eigenspaces of A give an orthogonal direct sum decomposition of $\mathbb{C}^n$.

(Sp3) There is a diagonal matrix D and a unitary matrix U such that $AU = UD$.

(Sp4) There is a partition of unity $P_1, \ldots, P_k$ and scalars $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ such that $A = \sum_{i=1}^{k} \lambda_i P_i$.

(Sp5) A is normal.

Proof. Most of the proof except (Sp5)⇒(Sp1) is similar to the 2×2 Spectral Theorem.

(Sp1)⇒(Sp3): Given our ONB of eigenstates $\{|e_1\rangle, \ldots, |e_n\rangle\}$, set

$$U := \begin{bmatrix} |e_1\rangle & \cdots & |e_n\rangle \end{bmatrix} \qquad \text{and} \qquad D := \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

where $|\psi_i\rangle$ is an eigenstate corresponding to the eigenvalue λ_i, where the λ_i may not all be distinct. The reader can check as in the 2×2 case that $AU = UD$ and U is unitary.

(Sp3)⇒(Sp4): Suppose $AU = UD$ where $U = \begin{bmatrix} |e_1\rangle & \cdots & |e_n\rangle \end{bmatrix}$ is unitary and D is diagonal. Then as in the 2×2 case,

$$A = UDU^* = \sum_{i=1}^{n} \lambda_i |\psi_i\rangle\langle\psi_i|$$

which is a linear combination of mutually orthogonal projections. Now it may be the case that not all λ_i are distinct, so we can group the projections $|\psi_i\rangle\langle\psi_i|$ into groups corresponding to the same eigenvalue. Summing these $|\psi_i\rangle\langle\psi_i|$ for the same eigenvalues gives the projections P_j.

(Sp4)⇔(Sp2): This is the content of the paragraph before the statement of the theorem. Just observe that if $A = \sum_{i=1}^{k} \lambda_i P_i$, then $\operatorname{im}(P_i)$ is exactly the eigenspace of A corresponding to λ_i.

(Sp4)⇒(Sp5): Just as in the 2×2 case,

$$\left(\sum_{i=1}^{k} \lambda_i P_i\right)^* \left(\sum_{j=1}^{k} \lambda_j P_j\right) = \sum_{i=1}^{k} |\lambda_i|^2 P_i = \left(\sum_{j=1}^{k} \lambda_j P_j\right) \left(\sum_{i=1}^{k} \lambda_i P_i\right)^*.$$

Here, it is important to use the fact that

$$P_i^* P_j = P_i P_j = \delta_{i=j} P_i = P_j P_i = P_j P_i^*.$$

(Sp5)⇒(Sp1): We omit this proof here as it would take us too far afield. The basic outline of the proof is as follows.

- First, the *characteristic polynomial* $\det(\lambda I - A)$ has a root $\lambda_1 \in \mathbb{C}$ by the Fundamental Theorem of Algebra.
- The root λ_1 is an eigenvalue of A with a corresponding eigenvector, say $|\psi_1\rangle$, which we may assume has norm 1. Since A is normal, $\overline{\lambda_1}$ is an eigenvalue of A^* with the same eigenvector $|\psi_1\rangle$.

- The subspace $\mathbb{C}|\psi\rangle \subset \mathbb{C}^n$ is invariant under both A and A^*, so we may restrict A to $\mathbb{C}|\psi_1\rangle^\perp$ to obtain another normal $(n-1)\times(n-1)$ matrix named B.
- We replace A with B and repeat the above steps, which yields an ONB of $\mathbb{C}^n$ consisting of eigenvectors of A. ■

Exercise 2.7.8 — Use the Spectral Theorem 2.7.7, to show that for every unitary U, there is a Hermitian/self-adjoint matrix H with the following two properties:

(1) The eigenspaces of U and H coincide.

(2) The eigenvalues λ_j of U are related to the eigenvalues μ_j of H by $\alpha_j = e^{i\beta_j}$.

In this sense, we say $U = \exp(iH)$. In fact, more is true; the series

$$\sum_{k\geq 0} \frac{(iH)^k}{k!} = I + iH - \frac{H^2}{2!} - i\frac{H^3}{3!} + \frac{H^4}{4!} + \cdots$$

converges entry-wise to U.

Exercise 2.7.9 — Suppose we start with the state $|{+}{+}\rangle \in \mathbb{C}^4$ and measure the observable $X\otimes I_2 + I_2 \otimes X$ and obtain the value 0. What is the probability of then measuring the observable $Z \otimes I_2$ and obtaining a value of $+1$?

2.8 The CHSH inequality

The exposition in this section is adapted from [NC00, p115-6].

Let's now revisit Alice and Bob who share an entangled 2-qubit state separated by a considerable distance. We now perform an experiment. We will need to suppose that Alice and Bob can generate as many pairs of entangled electrons or entangled photons as they would like, which is a reasonable assumption given experiments that can be carried out in laboratories today. Each of Alice and Bob will pick 2 observables to measure, whose possible measurement values are ± 1, for a total of 4 observables. We'll call these observables A_1, A_2 and B_1, B_2. For example, Alice may choose to measure her photon along the x- or z-axis. Bob can choose any other 2 observables he chooses.

On each round of the experiment, Alice and Bob choose at random to measure one of their two observables *independently*. That is, Alice chooses to measure A_1 or A_2, and Bob chooses to measure B_1 or B_2. After many repetitions, they compare their results.

We now suppose that local realism holds, which means that each photon has an *a priori* 'hidden answer' to whether it returns $+1$ or -1 when asked for its measurement for A_1, A_2 or B_1, B_2 locally.

We denote by a_i the value (± 1) that would be returned if Alice measures A_i for $i = 1, 2$. We denote by b_i the value (± 1) that would be returned if Bob measures B_i for $i = 1, 2$. Here, a_1, a_2, b_1, b_2 are the *hidden variables* possessed by the entangled photons shared by Alice and Bob. We will only ever be able to find out 2 of these values, as Alice only measures one of A_1, A_2Z and Bob only measures one of B_1, B_2.

Claim 2.8.1 — The random variable

$$R := a_1b_1 + a_2b_1 + a_1b_2 - a_2b_2 = (a_1 + a_2)b_1 + (a_1 - a_2)b_2$$

is always equal to ± 2.

Proof. Since a_1, a_2 are both ± 1, exactly one of $a_1 + a_2$ or $a_1 - a_2$ must be zero, and the other is ± 2. Since b_1, b_2 are both ± 1, the claim follows. ∎

We now repeat our experiment, and we calculate the expected value of our random variable R. It is clear that the expected value of R must lie between -2 and 2:

$$-2 \leq \mathbb{E}[R] \leq 2.$$

But since expected value $\mathbb{E}$ is *linear*, we also have that

$$-2 \leq \mathbb{E}[R] = \mathbb{E}[a_1b_1 + a_2b_1 + a_1b_2 - a_2b_2] = \mathbb{E}[a_1b_1] + \mathbb{E}[a_2b_1] + \mathbb{E}[a_1b_2] - \mathbb{E}[a_2b_2] \leq 2.$$

The right hand side is called a *Bell inequality*:

$$\mathbb{E}[a_1b_1] + \mathbb{E}[a_2b_1] + \mathbb{E}[a_1b_2] - \mathbb{E}[a_2b_2] \leq 2. \tag{2.8.2}$$

We now show that this Bell inequality can be *violated*, which means that local realism must fail. To be as concrete as possible, we will assume that Alice and Bob are generating as many pairs of photons in the Bell state

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

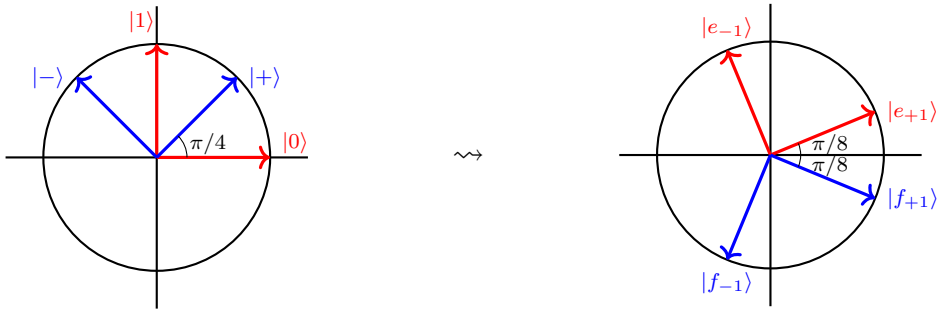
as they would like. Alice chooses her two observables to be measuring $A_1 = Z$ and $A_2 = X$ on her qubit. Bob chooses his two observables to be measuring

$$B_1 = H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \text{and} \qquad B_2 = ZHZ = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$$

on his qubit.

Exercise 2.8.3 — Show that Bob's two observables B_1, B_2 have possible measurement values ± 1. What are the corresponding eigenstates?

Recall that an ONB of eigenstates for Z is $\{|0\rangle, |1\rangle\}$, and an ONB of eigenstates for X is $\{|+\rangle, |-\rangle\}$; these ONBs are mutually unbiased. In $\mathbb{R}^2$, these ONBs are depicted in the diagram on the left below.



Observe that your eigenbasis $\{|e_{+1}\rangle, |e_{-1}\rangle\}$ for B_1 from Exercise 2.8.3 could be chosen to be a $\frac{\pi}{8}$-rotation of the standard basis. This $\pi/8$ is the same 22.5° that gave our previous violation of local realism in our sequential Stern-Gerlach experiment from the previous module. Your eigenbasis for B_2 could be chosen as $\{|f_{+1}\rangle = Z|e_{+1}\rangle, |f_{-1}\rangle = Z|e_{-1}\rangle\}$, which flips the y-coordinates. Multiplying $|f_{-1}\rangle$ by -1 , we see that again the eigenbases for B_1 and B_2 are mutually unbiased.

Exercise 2.8.4 — Compute the simple tensors $A_1 \otimes B_1$, $A_2 \otimes B_1$, $A_2 \otimes B_2$, and $A_1 \otimes B_2$ as elements of $M_4(\mathbb{C})$.

Exercise 2.8.5 — Compute $\langle\Psi|M|\Psi\rangle$ for $M = A_1 \otimes B_1, A_1 \otimes B_2, A_2 \otimes B_1, A_2 \otimes B_2$. What is

$$\langle\Psi|A_1 \otimes B_1|\Psi\rangle + \langle\Psi|A_1 \otimes B_2|\Psi\rangle + \langle\Psi|A_2 \otimes B_1|\Psi\rangle - \langle\Psi|A_2 \otimes B_2|\Psi\rangle?$$

What does this say about the CHSH Inequality (2.8.2)?

2.9 The GHSZ paradox

In this section, we provide a series of exercises guiding the reader through the GHSZ paradox [GHSZ90, §4-5]. Our presentation was adapted directly from [ZCZW19, §1.4.1]. In the exercises below, we consider the 3-qubit pure state

$$|GHSZ\rangle := \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle).$$

Exercise 2.9.1 — Prove that $|GHSZ\rangle$ is entangled.

Exercise 2.9.2 — Prove that the operators $Z \otimes Z \otimes I_2$, $X \otimes X \otimes X$, and $I_2 \otimes Z \otimes Z$ pairwise commute. There are exactly $2^3 = 8$ simple tensors of $\{I_2, X, Y, Z\}$ up to sign that one can obtain by multiplying copies of these 3 above operators together. (Why?) For example,

$$(Z \otimes Z \otimes I_2)(X \otimes X \otimes X) = ZX \otimes ZX \otimes X = iY \otimes iY \otimes X = -Y \otimes Y \otimes X.$$

So far, we have seen 4 of these operators:

$$Z \otimes Z \otimes I_2, \qquad X \otimes X \otimes X, \qquad I_2 \otimes Z \otimes Z, \qquad -Y \otimes Y \otimes X.$$

What are the other remaining 4?

Exercise 2.9.3 — Show that the set $\left\{(X^a \otimes Z^b \otimes X^c)|GHSZ\rangle \middle| a, b, c \in \{0, 1\}\right\}$ is simultaneously an ONB of eigenvectors for the three operators $Z \otimes Z \otimes I_2$, $X \otimes X \otimes X$, and $I_2 \otimes Z \otimes Z$. Then show that $\mathbb{C}|GHSZ\rangle$ is exactly the intersection of the $+1$ eigenspaces for each of these three operators.

Exercise 2.9.4 — Compute the expectation values $\langle X \otimes X \otimes X\rangle$, $\langle -Y \otimes Y \otimes X\rangle$, $\langle -Y \otimes X \otimes Y\rangle$, and $\langle -X \otimes Y \otimes Y\rangle$ in the GHSZ state. Do this in 2 ways: via direct computation and using Exercise 2.9.3.

Now suppose that each of the operators X, Y, Z have 'hidden values' in $\{\pm 1\}$ when acting on the first, second, or third qubit. Note here that these 'hidden values' can be different when the same operator acts on the first, second, or third qubits. We will denote these values by $v(X_i)$, $v(Y_j)$, and $v(Z_k)$ for $i, j, k \in \{A, B, C\}$ where A, B, C are the labels of the three qubits.

Suppose now that an experiment is performed which measures one of the four operators in Exercise 2.9.4. For instance, if $v(X_A) = -1, v(X_B) = +1, v(X_C) = -1$, then the observer would measure $+1$ for the value of $X \otimes X \otimes X$. However, there is no choice of the values $v(X_i)$, $v(Y_j)$, and $v(Z_k)$ for $i, j, k \in \{A, B, C\}$ for which all of the expectation values calculated in Exercise 2.9.4 can be true! That is, the following four equations are inconsistent:

$$\langle X \otimes X \otimes X\rangle = v(X_A)(X_B)v(X_C) \tag{2.9.5}$$
$$\langle -Y \otimes Y \otimes X\rangle = -v(Y_A)(Y_B)v(X_C) \tag{2.9.6}$$
$$\langle -Y \otimes X \otimes Y\rangle = -v(Y_A)(X_B)v(Y_C) \tag{2.9.7}$$
$$\langle -X \otimes Y \otimes Y\rangle = -v(X_A)(Y_B)v(Y_C). \tag{2.9.8}$$

This is the GHSZ paradox: there is no consistent choice of hidden variables which can mimic the quantum mechanical prediction of this experiment.

This demonstration may feel more troublesome to the reader than other quantum mechanical paradoxes. For example, Heisenberg's uncertainty principle for position and momentum measurements is troubling, but ultimately the acceptance of wave/particle duality reduces this principle to a simple fact about waves in Fourier analysis. A wave packet is similar enough to a particle that one can recover quickly from the initial shock.

More disturbing is violations Bell's inequalities. The nature of locality and causality is revealed to be inexorably linked to the randomness of quantum mechanics. As disturbing as these violations are, the discrepancy in the predicted expectations of measurements between classical and quantum mechanical theories are a matter of scale on a continuum of possibilities.

The GHSZ paradox forces us to confront the fact that our classical intuition about information is not merely a perturbation of a quantum truth, but is altogether incorrect when applied to the smallest scales of information. Bell's inequalities can be broken experimentally by using the "upon repeated sampling" approach to probability theory and its relation to scientific results. The GHSZ paradox shows us directly, without considering a probability distribution, that our classical conception of information is fundamentally incorrect.

Exercise 2.9.9 — Prove that that the system of equations (2.9.5), (2.9.6), (2.9.7), (2.9.8) is inconsistent.
Hint: Each of $v(X_A), v(X_B), v(X_c), v(Y_A), v(Y_B), v(Y_B)$ *appear exactly twice on the right hand side.*

2.10 Density matrices in $M_4(\mathbb{C})$

In the first module, we defined the notion of a density matrix ρ to define a mixed state, which described a classical mixture of states. In the 2 qubit setting, we can define density matrices and mixed states the same way, but as we will see, they capture more information about a quantum state than just a classical mixture. The following definition is made in $M_4(\mathbb{C})$, but it can easily be adapted to $M_n(\mathbb{C})$.

Definition 2.10.1 — A *density matrix* is a $\rho \in M_4(\mathbb{C})$ which

- is positive ($\rho \geq 0$), i.e., $\langle\Psi|\rho|\Psi\rangle \geq 0$ for all $|\Psi\rangle \in \mathbb{C}^4$, and
- has trace one, i.e., $\mathrm{Tr}(\rho) = \rho_{11} + \rho_{22} + \rho_{33} + \rho_{44} = 1$.

A density matrix ρ is called a *pure state* if there is a state vector $|\Psi\rangle \in \mathbb{C}^4$ such that $\rho = |\Psi\rangle\langle\Psi|$. Otherwise, we call ρ a *mixed state*.

Exercise 2.10.2 — Show that the following matrices are densities. Which ones are

pure?

$$\frac{1}{4}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} \qquad \frac{1}{4}\begin{bmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{bmatrix} \qquad \frac{1}{4}\begin{bmatrix}1&0&0&1\\0&1&1&0\\0&1&1&0\\1&0&0&1\end{bmatrix} \qquad \frac{1}{2}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}$$

Exercise 2.10.3 — Prove that $|\Psi\rangle\langle\Psi|$ is a density matrix when $|\Psi\rangle \in \mathbb{C}^4$ is a state vector.

Remark 2.10.4 — Since the image of $|\Psi\rangle\langle\Psi|$ is equal to $\mathbb{C}|\Psi\rangle$ which is 1-dimensional, we see that ρ is pure if and only if it has *rank 1*, i.e., the columns of ρ are all scalar multiples of each other. In this case, it is easy to find a state vector $|\Psi\rangle \in \mathbb{C}^4$ such that $\rho = |\Psi\rangle\langle\Psi|$: just take any non-zero column of ρ and normalize it to get a state vector.

Again, we can use convex combinations of densities to form more densities; this is an easy way to construct lots of examples of mixed states.

We now look at the equivalent conditions to being a pure state in $M_2(\mathbb{C})$, and we see which ones still remain true for $M_4(\mathbb{C})$. We can immediately remove any conditions referring to the Bloch sphere, as this was particular to $M_2(\mathbb{C})$. Recall that $\rho \in M_2(\mathbb{C})$ is pure if and only if one of the following equivalent conditions hold:

(P1) ρ is pure,

(P2) ρ is not invertible,

(P3) $\det(\rho) = 0$,

(P4) $\operatorname{rank}(\rho) = 1$,

(P5) 0 is an eigenvalue of ρ,

(P6) 1 is an eigenvalue of ρ,

(P7) $\rho^2 = \rho$

(P8) $\operatorname{Tr}(\rho^2) = 1$

The fourth example in Exercise 2.10.2 is a mixed state which is not invertible. This means (P2), (P3), and (P5) are no longer equivalent to (P1). However, all the rest are still equivalent.

Theorem 2.10.5 — The following are equivalent for a density matrix $\rho \in M_4(\mathbb{C})$.

(P1) ρ is pure,

(P2) $\text{rank}(\rho) = 1$,

(P3) 1 is an eigenvalue of ρ,

(P4) $\rho^2 = \rho$

(P5) $\text{Tr}(\rho^2) = 1$

Proof. The result follows from the Spectral Theorem 2.7.7 taking a spectral decomposition of ρ:

$$\rho = \lambda_1|e_1\rangle\langle e_1| + \lambda_2|e_2\rangle\langle e_2| + \lambda_3|e_3\rangle\langle e_3| + \lambda_4|e_4\rangle\langle e_4|$$

where, without loss of generality, we have

$$1 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq 0 \qquad \text{and} \qquad \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4.$$

Since the λ_i are between 0 and 1 and sum to 1, every condition in the statement of the theorem is equivalent to $\lambda_1 = 1$ and all other $\lambda_i = 0$. ∎

Exercise 2.10.6 — Show that if ρ_1 and ρ_2 are density matrices, so is $(\rho_1 + \rho_2)/2$.

Exercise 2.10.7 — Suppose that ρ_1 and ρ_2 are pure density matrices. When is it the case that $(\rho_1 + \rho_2)/2$ is pure?

Exercise 2.10.8 — Suppose that ρ_1 and ρ_2 are pure density matrices in $M_2(\mathbb{C})$. Is $\rho_1 \otimes \rho_2$ pure?

Exercise 2.10.9 — Give an example of a density matrix $\rho \in M_4(\mathbb{C})$ which is mixed, but not invertible.

By definition (see §2.4), a state vector $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ is either a product state of the form $|\psi\rangle \otimes |\eta\rangle = |\psi\eta\rangle$ or entangled. This is no longer the case for densities.

Definition 2.10.10 — A density matrix $\rho \in M_4(\mathbb{C})$ is called

- a *product density* if it is of the form $\rho_1 \otimes \rho_2$ for densities $\rho_1, \rho_2 \in M_2(\mathbb{C})$,
- *separable* if it is a convex combination of product densities, and

- *entangled* if it is not separable.

We note here that separable states are not considered entangled, but rather as classical mixtures of product states.

Exercise 2.10.11 — Show that a vector state $|\Psi\rangle\langle\Psi|$ is either a product state or entangled.

Exercise 2.10.12 — Suppose $\rho \in M_4(\mathbb{C})$ is a density and $U \in M_4(\mathbb{C})$ is a unitary.

(1) Show that $U^*\rho U \in M_4(\mathbb{C})$ is also a density.

(2) Show that ρ is a product density, separable, or entangled if and only if $U^*\rho U \in M_4(\mathbb{C})$ is a product density, separable, or entangled respectively.

The second part of the above exercise says that entanglement does not depend on a choice of basis, but rather is fundamental to the density itself. Moreover, entanglement can never be achieved from a separable state under unitary time evolution.

2.11 Reduced densities and partial traces

We saw how to take measurements of local operators in §2.5 above. However, if we are merely interested in *expectations* of local measurements, we can do this with only *local* knowledge of our state. This is made precise in the following theorem. In this section, we use the labels A, B to denote the first and second qubits in $\mathbb{C}^2 \otimes \mathbb{C}^2$.

Proposition 2.11.1 — Suppose $\rho \in M_4(\mathbb{C})$ is a density matrix. There is a unique *reduced density matrix* $\rho^A \in M_2(\mathbb{C})$ such that

$$\mathrm{Tr}_4(\rho(M \otimes I_2)) = \mathrm{Tr}_2(\rho^A M) \qquad \forall\, M \in M_2(\mathbb{C}). \tag{2.11.2}$$

Here, we use the subscripts $2, 4$ to denote which trace is applied above.

Proof. Assuming such a ρ^A exists, we can calculate its ij-th entry by

$$\begin{aligned}
\langle i|\rho^A|j\rangle &= \mathrm{Tr}_2(\rho^A|j\rangle\langle i|) \\
&= \mathrm{Tr}_4(\rho(|j\rangle|i\rangle \otimes I_2)) && \text{(by assumption)} \\
&= \sum_{k,\ell=0}^{1} \langle \ell k|\rho \cdot \underbrace{(|j\rangle\langle i| \otimes I_2)|\ell k\rangle}_{\delta_{\ell=i}\cdot|jk\rangle} \\
&= \sum_{k=0}^{1} \langle ik|\rho|jk\rangle.
\end{aligned}$$

So there is at most one operator $\rho^A \in M_2(\mathbb{C})$ that can satisfy (2.11.2), and it is given by

$$\langle i|\rho^A|j\rangle = \sum_{k=0}^{1} \langle ik|\rho|jk\rangle.$$

We now verify that this ρ^A actually does satisfy (2.11.2):

$$\begin{aligned}
\mathrm{Tr}_4(\rho(M \otimes I_2)) &= \sum_{i,k=0}^{1} \langle ik|\rho(M \otimes I_2)|ik\rangle \\
&= \sum_{i,k=0}^{1} \sum_{j,\ell=0}^{1} \langle ik|\rho \underbrace{|j\ell\rangle\langle j\ell|}_{=I_4}(M \otimes I_2)|ik\rangle \\
&= \sum_{i,k=0}^{1} \sum_{j,\ell=0}^{1} \langle ik|\rho|j\ell\rangle\langle j|M|i\rangle \underbrace{\langle \ell|I_2|k\rangle}_{\delta_{\ell=k}} \\
&= \sum_{i,j=0}^{1} \underbrace{\left(\sum_{k=0}^{1} \langle ik|\rho|jk\rangle \right)}_{=\langle i|\rho^A|j\rangle} \langle j|M|i\rangle \\
&= \sum_{i,j=0}^{1} \langle i|\rho^A \underbrace{|j\rangle\langle j|}_{=I_2} M|i\rangle \\
&= \sum_{i=0}^{1} \langle i|\rho^A M|i\rangle \\
&= \mathrm{Tr}_2(\rho^A M).
\end{aligned}$$

It remains to prove that ρ^A is again a density. First, to see ρ^A has trace 1, we calculate

$$\mathrm{Tr}_2(\rho^A) = \sum_{i=0}^{1} \langle i|\rho^A|i\rangle = \sum_{i,k=0}^{1} \langle ik|\rho|ik\rangle = \mathrm{Tr}_4(\rho) = 1.$$

To see ρ^A is positive, we observe that

$$\rho^A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \rho \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \rho \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.11.3}$$

Then for any $|\psi\rangle \in \mathbb{C}^2$, we have

$$\langle\psi|\rho^A|\psi\rangle = \langle\psi| \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \rho \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} |\psi\rangle}_{=:|\psi_1\rangle} + \langle\psi| \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \rho \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} |\psi\rangle}_{=:|\psi_2\rangle}$$
$$= \langle\psi_1|\rho|\psi_1\rangle + \langle\psi_2|\rho|\psi_2\rangle,$$

and each of these terms on the right are non-negative. ∎

We extract the following trick from the above proof.

Trick 2.11.4 — If $C \in M_n(\mathbb{C})$ is a positive matrix, then for any $m \times n$ matrix Q, $QCQ^* \in M_m(\mathbb{C})$ is also positive.

Observe that the construction of ρ^A in the proof of Proposition 2.11.1 did not rely on ρ being a density; the formula makes sense for *any* $C \in M_4(\mathbb{C})$.

Definition 2.11.5 — The *partial trace* $\mathrm{Tr}^B : M_4(\mathbb{C}) \to M_2(\mathbb{C})$ is defined by

$$\langle i|\,\mathrm{Tr}^B(C)|j\rangle := \sum_{k=0}^{1} \langle ik|C|jk\rangle.$$

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix} \longmapsto \begin{bmatrix} c_{11} + c_{22} & c_{13} + c_{24} \\ c_{31} + c_{42} & c_{33} + c_{44} \end{bmatrix}$$

The name *partial trace* comes from the following property.

Exercise 2.11.6 — Prove that for all $M, N \in M_2(\mathbb{C})$, $\mathrm{Tr}^B(M \otimes N). = \mathrm{Tr}_2(N)M$ In particular, $\mathrm{Tr}^B(M \otimes I_2) = 2M$ and $\mathrm{Tr}^B(I_2 \otimes N) = \mathrm{Tr}_2(N)I_2$.

Exercise 2.11.7 — Verify that Tr^B is also given by (2.11.3).

Using the formula (2.11.3), we see that Tr^B, we see that Tr^B is linear and positive. (In fact, this equation shows Tr^B is *completely positive*, which is an even stronger property important in quantum information. This is beyond the scope of these notes though.)

Exercise 2.11.8 — Verify that the partial trace Tr^B has the following additional properties:

- Tr^B is $M_2(\mathbb{C})$-*bimodular* in the first component, i.e., for all $M_1, M_2 \in M_2(\mathbb{C})$ and $C \in M_4(\mathbb{C})$,
$$\mathrm{Tr}^B((M_1 \otimes I_2)C(M_2 \otimes I_2)) = M_1 \,\mathrm{Tr}^B(C)M_2.$$
- Tr^B is *tracial* in the second component, i.e., for all $N \in M_2(\mathbb{C})$ and $C \in M_4(\mathbb{C})$,
$$\mathrm{Tr}^B((I_2 \otimes N)C) = \mathrm{Tr}^B(C(I_2 \otimes N)).$$

Similar to Proposition 2.11.1 above, for every density matrix $\rho \in M_4(\mathbb{C})$, there is a unique reduced density matrix $\rho^B \in M_2(\mathbb{C})$ such that

$$\mathrm{Tr}_4(\rho(I_2 \otimes N)) = \mathrm{Tr}_2(\rho^B N) \qquad \forall\, N \in M_2(\mathbb{C}). \tag{2.11.9}$$

Again, the map $\rho \mapsto \rho^A$ makes sense for every matrix in $M_4(\mathbb{C})$.

Definition 2.11.10 — The *partial trace* $\mathrm{Tr}^A : M_4(\mathbb{C}) \to M_2(\mathbb{C})$ is defined by

$$\langle i|\,\mathrm{Tr}^A(C)|j\rangle := \sum_{k=0}^{1} \langle ki|C|kj\rangle.$$

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix} \longmapsto \begin{bmatrix} c_{11}+c_{33} & c_{12}+c_{34} \\ c_{21}+c_{43} & c_{22}+c_{44} \end{bmatrix}$$

Exercise 2.11.11 — Verify that ρ^B satisfies (2.11.9).

Exercise 2.11.12 — Prove that for all $M, N \in M_2(\mathbb{C})$, $\mathrm{Tr}^A(M \otimes N) = \mathrm{Tr}_2(M)N$ In particular, $\mathrm{Tr}^A(I_2 \otimes N) = 2N$ and $\mathrm{Tr}^A(M \otimes I_2) = \mathrm{Tr}_2(M)I_2$.

Exercise 2.11.13 — Find a formula similar to (2.11.3) for Tr^A. Deduce that Tr^A is linear and positive.

Exercise 2.11.14 — Verify that the partial trace Tr^A has the following properties:

- Tr^A is $M_2(\mathbb{C})$-*bimodular* in the second component, i.e., for all $N_1, N_2 \in M_2(\mathbb{C})$ and $C \in M_4(\mathbb{C})$,
$$\mathrm{Tr}^B((I_2 \otimes N_1)C(I_2 \otimes N_2)) = N_1 \,\mathrm{Tr}^B(C)N_2.$$
- Tr^B is *tracial* in the first component, i.e., for all $M \in M_2(\mathbb{C})$ and $C \in M_4(\mathbb{C})$,
$$\mathrm{Tr}^B((M \otimes I_2)C) = \mathrm{Tr}^B(C(M \otimes I_2)).$$

We close this section by proving that there is nothing special about using the computational basis $\{|0\rangle, |1\rangle\}$ to compute the partial traces. This is an instance of Remark 3.2.2

Proposition 2.11.15 — Let $\{|e_1\rangle, |e_2\rangle\}$ be another ONB for $\mathbb{C}^2$. The partial traces $\mathrm{Tr}^A, \mathrm{Tr}^B$ are also given by
$$\langle i|\,\mathrm{Tr}^A(C)|j\rangle = \sum_{k=1}^{2}\langle e_k i|C|e_k j\rangle \qquad \text{and} \qquad \langle i|\,\mathrm{Tr}^B(C)|j\rangle = \sum_{k=1}^{2}\langle ie_k|C|je_k\rangle.$$

Proof. We prove the second formula and leave the first to the reader. First, given any ONB $\{|e_1\rangle, |e_2\rangle\}$ of $\mathbb{C}^2$, $I_2 = |e_1\rangle\langle e_1| + |e_2\rangle\langle e_2|$. This means that we can write
$$I_4 = I_2 \otimes I_2 = \sum_{i,k=0}^{1} |ie_k\rangle\langle ie_k| \qquad \text{and} \qquad \mathrm{Tr}_4(\,\cdot\,) = \sum_{i,k=0}^{1}\langle ie_k|\,\cdot\,|ie_k\rangle.$$
We now calculate that
$$\begin{aligned}
\mathrm{Tr}_4(\rho(M \otimes I_2)) &= \sum_{i,k=0}^{1}\langle ie_k|\rho(M \otimes I_2)|ie_k\rangle \\
&= \sum_{i,k=0}^{1}\sum_{j,\ell=0}^{1}\langle ie_k|\rho\underbrace{|je_\ell\rangle\langle je_\ell|}_{=I_4}(M \otimes I_2)|ie_k\rangle \\
&= \sum_{i,j=0}^{1}\sum_{k,\ell=0}^{1}\langle ie_k|\rho|je_\ell\rangle\langle j|M|i\rangle\underbrace{\langle e_\ell|I_2|e_k\rangle}_{\delta_{\ell=k}} \\
&= \sum_{i,j=0}^{1}\underbrace{\left(\sum_{k=0}^{1}\langle ie_k|\rho|je_k\rangle\right)}_{=:\langle i|(\rho^A)'|k\rangle}\langle j|M|i\rangle \\
&= \sum_{i,k=0}^{1}\langle i|(\rho^A)'\underbrace{|j\rangle\langle j|}_{=I_2}M|i\rangle
\end{aligned}$$

$$= \sum_{i=0}^{1} \langle i|(\rho^A)' M|i\rangle$$
$$= \mathrm{Tr}_2((\rho^A)' M).$$

In the sequence of equations above, we implicitly defined $(\rho^A)' \in M_2(\mathbb{C})$ by defining its ij-th entry. By the above calculation, we have $(\rho^A)'$ satisfies the criterion (2.11.2), i.e.,

$$\mathrm{Tr}_4(\rho(M \otimes I_2)) = \mathrm{Tr}_2((\rho^A)' M) \qquad \forall\, M \in M_2(\mathbb{C}),$$

so $(\rho^A)' = \rho^A$ by uniqueness in Proposition 2.11.1. Hence $\langle i|\,\mathrm{Tr}^B(\rho)|j\rangle = \sum_{k=0}^{1} \langle ie_k|\rho|je_k\rangle$ as claimed. ∎

2.12 Quantum state purification

We now show that every mixed state $\rho \in M_2(\mathbb{C})$ is a partial trace of a pure state in $\mathbb{C}^4$. This result is known as *quantum state purification*.

Theorem 2.12.1 (Quantum state purification) — For every density matrix $\rho \in M_2(\mathbb{C})$, there is a state vector $|\Psi\rangle \in \mathbb{C}^4$ such that $\rho = (|\Psi\rangle\langle\Psi|)^A = \mathrm{Tr}^B(|\Psi\rangle\langle\Psi|)$.

Proof. Let $\rho = \sum_{i=0}^{1} \lambda_i |e_i\rangle\langle e_i|$ be the spectral decomposition of ρ, and pick any ONB $\{|f_1\rangle, |f_2\rangle\}$ of $\mathbb{C}^2$. Setting $|\Psi\rangle := \sum_{i=0}^{1} \lambda_i^{1/2} |e_i f_i\rangle$, we have

$$|\Psi\rangle\langle\Psi| := \sum_{i,j=0}^{1} \lambda_i^{1/2} \lambda_j^{1/2} |e_i f_i\rangle\langle e_j f_j|.$$

Using the formula for Tr^B from Proposition 2.11.15 with the ONB $\{|f_1\rangle, |f_2\rangle\}$, we see

$$\mathrm{Tr}^B(|\Psi\rangle\langle\Psi|) = \sum_{i,j=0}^{1} \sum_{k=0}^{1} (I_2 \otimes \langle f_k|) \left(\lambda_i^{1/2} \lambda_j^{1/2} |e_i f_i\rangle\langle e_j f_j|\right) (I_2 \otimes |f_k\rangle) = \sum_{k=0}^{1} \lambda_k |e_k\rangle\langle e_k| = \rho$$

as desired. ∎

Remark 2.12.2 — There was nothing special about ρ being in $M_2(\mathbb{C})$ here; indeed, for a mixed state in $M_n(\mathbb{C})$, we can easily adapt the above proof by changing the top index of summation to get a $|\Psi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^n$ such that $\mathrm{Tr}^B(|\Psi\rangle\langle\Psi|) = \rho$.

Exercise 2.12.3 — Find a state in $\mathbb{C}^4$ which is a purification of the density matrix given by $I_2 + X/2$.

Exercise 2.12.4 — Show that if $|\psi\rangle \in \mathbb{C}^4$ is a purification of a density matrix $\rho \in M_2(\mathbb{C})$ and $U \in M_2(\mathbb{C})$ is unitary, then $(I_2 \otimes U)|\psi\rangle$ is also a purification of ρ.

2.13 Von Neumann entropy

A qubit in a mixed state can be thought of as being entangled with its environment. The system together with its environment is a composite system always assumed to be in a pure state. Von Neumann entropy is a useful measure of how entangled a system is with its environment.

Definition 2.13.1 — Let ρ be an $n \times n$ density matrix. As in §1.17, we define

$$S(\rho) := -\operatorname{Tr}(\rho \log(\rho)) = -\sum_{\substack{k=1 \\ \lambda_k \neq 0}}^{n} \lambda_k \log(\lambda_k).$$

where ρ has eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$. As before, $S(\rho)$ is exactly the Shannon entropy of the probability distribution $\{\lambda_i\}_{i=1}^n$.

The following theorem says that if we know all the information about our (pure) quantum state, then there is no entanglement with the environment.

Proposition 2.13.2 (Entropy of pure states) — For a density matrix $\rho \in M_n(\mathbb{C})$, $S(\rho) = 0$ if and only if ρ is a pure state.

Proof. First, we observe that whenever $0 \leq \lambda \leq 1$, $\lambda \log(\lambda) \leq 0$. Thus if

$$S(\rho) = -\sum_{\substack{k=1 \\ \lambda_k \neq 0}}^{n} \lambda_k \log(\lambda_k) = 0,$$

then $\lambda_k \log(\lambda_k) = 0$ for every k. This means $\lambda_k \in \{0, 1\}$ for all k. But because $\sum_k \lambda_k = \operatorname{Tr}(\rho) = 1$, exactly one $\lambda_k = 1$, and the rest are all zero. This means there is a unique eigenvalue of ρ equal to 1, corresponding to a 1-dimensional eigenspace $\mathbb{C}|\psi\rangle$ for some norm one state vector $|\psi\rangle$. By the Spectral Theorem 2.7.7, $\rho = |\psi\rangle\langle\psi|$.

For the other direction, suppose $\rho = |\psi\rangle\langle\psi|$. Then in the spectral decomposition of ρ, we have only one non-zero eigenvalue, and thus

$$S(\rho) = -\sum_{\substack{k=1 \\ \lambda_k \neq 0}}^{n} \lambda_k \log(\lambda_k) = 0$$

■

Example 2.13.3 — By Example 1.16.5, in $\mathbb{C}^d$, the maximum entropy of a density matrix is $\log(d)$, which is achieved at a unique density, namely $\frac{1}{d}I_d$.

Definition 2.13.4 (Bipartite entanglement entropy) — Suppose we have a quantum system comprised of two components called A and B, usually called a bipartite system. We will denote the total system as $A \cup B$. Furthermore, denote the Hilbert spaces of systems A and B as $\mathbb{C}^a$ and $\mathbb{C}^b$, respectively. When we say that $A \cup B$ is "comprised" of A and B, we mean that the total Hilbert space for our quantum system $A \cup B$ is $\mathbb{C}^a \otimes \mathbb{C}^b$. Let ρ be a density matrix describing the full system. We now define the bipartite entanglement entropy of A to be the quantity $S_A := S(\rho_A)$ where $\rho_A := \mathrm{Tr}^B(\rho)$. Similarly, we define the bipartite entanglement entropy of B to be $S_B := S(\rho_B)$ where $\rho_B := \mathrm{Tr}^A(\rho)$.

Our next task is to show that $S_A = S_B$ whenever ρ is pure. To do this, we use the following fact.

Fact 2.13.5 — Suppose $|\Psi\rangle \in \mathbb{C}^a \otimes \mathbb{C}^b$ is a state vector. The vector state $\rho = |\Psi\rangle\langle\Psi|$ can be obtained from either of its reduced densities ρ_A or ρ_B via Quantum State Purification 2.12.1.

The proof of this fact is beyond the scope of these notes, but it uses the *Schmidt decomposition* of $|\Psi\rangle$, which says there is a $d \leq \min\{a, b\}$ and $\lambda_1, \ldots, \lambda_d > 0$ with $\sum_{i=1}^d \lambda_i = 1$, and orthonormal sets $\{|e_1\rangle, \ldots, |e_d\rangle\} \subset \mathbb{C}^a$ and $\{|f_1\rangle, \ldots, |f_d\rangle\} \subset \mathbb{C}^b$ such that

$$|\Psi\rangle = \sum_{i=1}^{d} \sqrt{\lambda_i} |e_i f_i\rangle.$$

From this decomposition, we see that

$$\rho_A = \sum_{i,j=1}^{d} \mathrm{Tr}^B \left(\sqrt{\lambda_i}\sqrt{\lambda_j} |e_i f_i\rangle\langle e_j f_j| \right) = \sum_{i=1}^{d} \lambda_i |e_i\rangle\langle e_i|, \tag{2.13.6}$$

and similarly,

$$\rho_B = \sum_{i=1}^{d} \lambda_i |f_i\rangle\langle f_i|. \tag{2.13.7}$$

Observe that indeed $\rho = |\Psi\rangle\langle\Psi|$ is a purification of both ρ_A and ρ_B.

Corollary 2.13.8 — For a bipartite system $A \cup B$ in a pure state, $S_A = S_B$.

Proof. Observe that (2.13.6) and (2.13.7) are spectral decompositions of ρ_A and ρ_B respectively. From these formulas, we see that ρ_A and ρ_B have the same non-zero eigenvalues, giving the same probability distribution $\{\lambda_i\}$. Thus

$$S_A = S(\rho_A) = H(\{\lambda_i\}) = S(\rho_B) = S_B.$$

■

Exercise 2.13.9 — Compute the bipartite entanglement entropy in each of the four Bell states.

Exercise 2.13.10 — Given a system with 10 qubits separated into two groups of 5, give an example of a pure state with maximal bipartite entanglement entropy.

Exercise 2.13.11 — Suppose $\rho \in M_4(\mathbb{C})$ is a 2-qubit state with initial bipartite entanglement entropy S_A. Act a Pauli operator (X, Y, Z) on the first qubit to obtain $\rho' = (X \otimes I_2)\rho(X \otimes I_2)$. What is the bipartite entanglement entropy of ρ'?

3 Applications

3.1 Introduction: Applications of graphical calculus to quantum information

The mathematical subject of *linear algebra* gets its name from the study of linear equations, i.e., equations of the form

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_1x_1 + a_2x_2 + \cdots a_nx_n = 0,$$

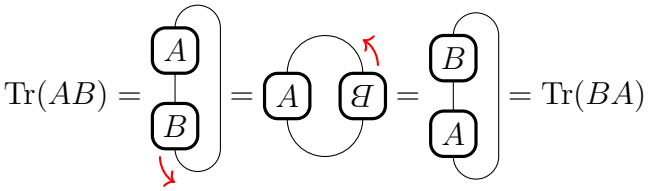
where each variable x_i appears without any exponents. Typically a homogeneous system of linear equations (i.e., a collection of equations of the above form) can be represented by its matrix A of coefficients, and the kernel/null space $\ker(A)$ of this matrix represents the vector subspace of $\mathbb{C}^n$ of solutions. That is, the subspace of solutions consists of the set of vectors $|\psi\rangle \in \mathbb{C}^n$ such that $A|\psi\rangle = \mathbf{0}$:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n &= 0 \\ \vdots \qquad\qquad & \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n &= 0 \end{aligned} \quad \leftrightarrow \quad \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{|\psi\rangle} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{0}}.$$

There is another way in which the adjective *linear* applies here; the literal equation $A|\psi\rangle = \mathbf{0}$ is physically one dimensional when written on a page. Similarly, the defining relation of the trace on $M_n(\mathbb{C})$

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$$

appears one dimensional when written on a page, but we will see later that this equation is better represented in *two dimensions*!



The operation of *tensor product* of vector/Hilbert spaces gives a second spatial dimension in which we can write equations. This second spatial dimension of equations is elegantly represented via the *graphical calculus* for finite dimensional vector/Hilbert spaces. Its origins trace back to research in category theory of Eilenberg and Kelly [EK66], Penrose diagrams for tensors [Pen71], and the Jones polynomial [Jon85] which in turn is deeply connected to

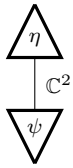
string diagrams appearing in work of Temperley-Lieb in statistical mechanics [TL71] and further developed by Kauffman [Kau87]. The graphical calculus has more recently been championed by Joyal-Street [JS91] and work in categorical quantum mechanics by Coecke and collaborators [CK17, HV19, CG23].

In this section, we will discuss in detail the two dimensional graphical calculus for finite dimensional Hilbert spaces. We will then give applications of the graphical calculus to quantum information, including quantum teleportation, superdense coding, and a non-trivial link between these two protocols. We end with a discussion on quantum key distribution which does not use the graphical calculus; this application could be presented right after §2.8 on the CHSH inequality if desired.

3.2 Graphical calculus for 1 qubit

There is a powerful and elegant graphical calculus for working with tensor product spaces $(\mathbb{C}^2)^{\otimes n}$. We denote a single copy of $\mathbb{C}^2$ by a line or *string*, which is sometimes drawn vertically and sometimes drawn horizontally depending on the application. For now, we will work with vertically drawn strings read *bottom-to-top*, but later on when we discuss quantum computation and quantum circuits, we will draw our pictures horizontally read *left-to-right*.

We represent kets/vectors in $\mathbb{C}^2$ by triangular coupons attached to the bottom of the string labelled by $\mathbb{C}^2$, and we represent bras/maps $\mathbb{C}^2 \to \mathbb{C}$ by triangular coupons attached to the top of the string. Note that the bras are the vertical reflections of the kets.

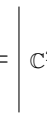
The inner product is represented by attaching the strings in the middle: 

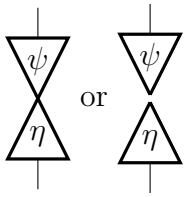
$:= \langle\eta|\psi\rangle$.

We represent operators in $M_2(\mathbb{C})$ by drawing boxes on the strings labelled by $\mathbb{C}^2$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{21} \end{pmatrix} = \boxed{A} \quad \begin{matrix} \mathbb{C}^2 \\ \\ \mathbb{C}^2 \end{matrix}.$$

The identity operator is represented simply by a string, and ket-bras are denoted by vertically joining the ket and the bra:

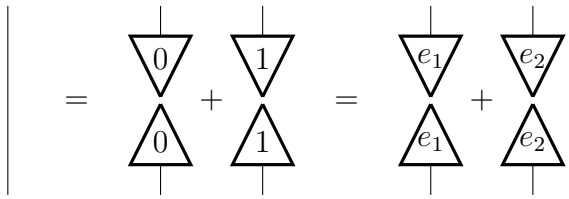
$I_2 =$ 

$|\psi\rangle\langle\eta| =$ 

or

.

The relation $I_2 = |0\rangle\langle 0| + |1\rangle\langle 1|$ means we may 'break' the string by summing over all indices:



(3.2.1)

for any ONB $\{|e_1\rangle, |e_2\rangle\}$ of $\mathbb{C}^2$.

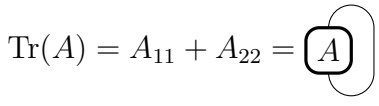
Remark 3.2.2 (*) — The fact that we may use any ONB above has a much wider generalization; any time we sum over any ONB appearing as kets and bras simultaneously, one time each, we may replace this ONB with any other ONB. This follows by the following computation in $\mathrm{Maps}(\mathbb{C}^2 \to \mathbb{C}) \otimes \mathbb{C}^2$:

$$\begin{aligned}
\sum_{i=1}^{2} \langle e_i| \otimes |e_i\rangle &= \sum_{i=1}^{2} \langle e_i| \otimes I_2 |e_i\rangle = \sum_{i=1}^{2} \langle e_i| \otimes \left(\sum_{j=1}^{2} |f_j\rangle\langle f_j| \right) |e_i\rangle \\
&= \sum_{i=1}^{2} \sum_{j=1}^{2} \langle e_i| \otimes \langle f_j|e_i\rangle |f_j\rangle = \sum_{i=1}^{2} \sum_{j=1}^{2} \langle e_i|\langle f_j|e_i\rangle \otimes |f_j\rangle \\
&= \sum_{i=1}^{2} \sum_{j=1}^{2} \langle f_j| \left(\sum_{j=1}^{2} |e_i\rangle\langle e_i| \right) \otimes |f_j\rangle = \sum_{j=1}^{2} \langle f_j| I_2 \otimes |f_j\rangle \\
&= \sum_{j=1}^{2} \langle f_j| \otimes |f_j\rangle.
\end{aligned}$$

We multiply matrices by stacking boxes, and we act on vectors by stacking vertically:

Exercise 3.2.3 — How would you represent $\langle\psi|A$ or $\langle\psi|A|\eta\rangle$?

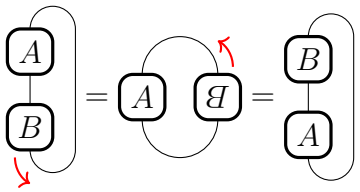
We represent the trace of a 2×2 matrix by 'capping off' on the right:



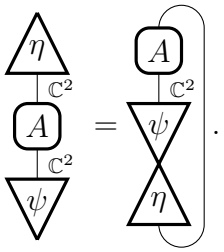
(3.2.4)

We will make this notation rigorous when we introduce our notation for tensor products below. Graphically, it should be justified that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$:

$$\mathrm{Tr}(AB) = \quad = \quad = \quad = \mathrm{Tr}(BA)\,. \tag{3.2.5}$$

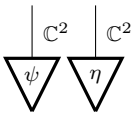


For the time being, the middle picture does not yet make sense, but we will revisit it in §3.4 below. We now see that that the formula $\langle\eta|A|\psi\rangle = \mathrm{Tr}(|\psi\rangle\langle\eta|A)$ has a nice graphical interpretation:

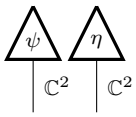


3.3 Graphical calculus for 2 qubits

To denote the tensor product $\mathbb{C}^2 \otimes \mathbb{C}^2$ of two copies of $\mathbb{C}^2$, we draw two vertical strings next to each other with space in the middle. Product vectors are denoted by drawing two kets in $\mathbb{C}^2$ side-by-side, and similarly for the corresponding bras.

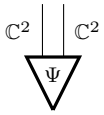


$:= |\psi\eta\rangle$

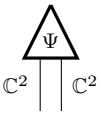


$:= \langle\psi\eta|.$

A vector $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ that is not known to be a product vector is denoted by a triangle coupon with two $\mathbb{C}^2$ lines emanating from the top:

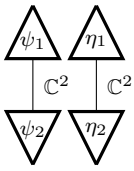


$:= |\Psi\rangle$



$:= \langle\Psi|$

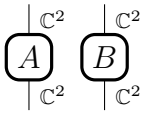
Again, the inner product is denoted by connecting both middle strings, and now the formula for the inner product for product vectors from Exercise 2.2.10 becomes visibly apparent:



$:= \langle\psi_1\eta_1|\psi_2\eta_2\rangle = \langle\psi_1|\psi_2\rangle \cdot \langle\eta_1|\eta_2\rangle.$

Simple tensors are denoted by two boxes on strings which are separated from each other:

$$A \otimes B = \quad .$$



Exercise 3.3.1 — Suppose $A, B \in M_2(\mathbb{C})$. Show that $A \otimes B = (A \otimes I_2)(I_2 \otimes B)$. Then show that $A \otimes B = (I_2 \otimes B)(A \otimes I_2)$. How do you represent this in the graphical calculus?

Simple tensors act on product vectors by stacking as usual. Operators that are not known to be simple tensors are denoted by a single longer box connected to both strings:

$$C = \begin{array}{c} \mathbb{C}^2 \quad \mathbb{C}^2 \\ \boxed{C} \\ \mathbb{C}^2 \quad \mathbb{C}^2 \end{array} .$$

The following exercise will be important when we revisit graphical calculus later on in §3.4.

Exercise 3.3.2 — Find $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ such that $\langle\Psi| : \mathbb{C}^2 \otimes \mathbb{C}^2 \to \mathbb{C}$ is the map $|ij\rangle \mapsto \delta_{i=j}$ for $i, j \in \{0, 1\}$. Then prove that $\langle\Psi|ij\rangle = \delta_{i=j}$ for $i, j \in \{+, -\}$.
Hint: Think of $\langle\Psi|$ as a 1×4 matrix and remember that the i-th column is exactly $\langle\Psi|e_i\rangle$ where $|e_i\rangle$ is the i-th basis vector.

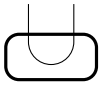
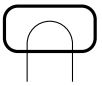
Exercise 3.3.3 — Use graphical calculus to show that for $A, U \in M_2(\mathbb{C})$, $Tr(UAU^\dagger)$ when U is unitary.

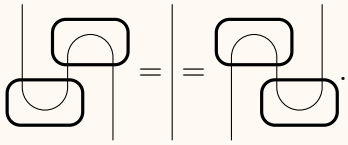
Exercise 3.3.4 — If $C \in M_4(\mathbb{C})$, how should $\mathrm{Tr}(C)$ be represented in graphical calculus? Check your idea by using it to confirm that for $A, B \in M_2(\mathbb{C})$, $\mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A)\,\mathrm{Tr}(B)$. Hint: In graphical calculus, a diagram consisting of two separate closed diagrams (in other words, two complex numbers) is the product of the diagrams (product of the two complex numbers).

3.4 Graphical calculus for Bell states

We now revisit the graphical calculus and give graphical representations of the partial traces. We will see that this graphical calculus guides our intuition about what properties we should expect from the partial traces.

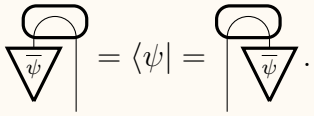
First, we denote the un-normalized Bell state $\sqrt{2}|\beta\rangle \in \mathbb{C}^4$ by a cup. We denote the corresponding bra by a cap.

 $:= |00\rangle + |11\rangle$  $:= \langle 00| + \langle 11|$

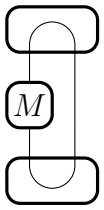
Exercise 3.4.1 — Show that  .

Hint: Evaluate each on $|i\rangle$ for $i = 0, 1$.

Exercise 3.4.2 — For $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \in \mathbb{C}^2$, define $|\overline{\psi}\rangle := \overline{\alpha}|0\rangle + \overline{\beta}|1\rangle$. Show that



Using this notation, we can now prove the correctness of (3.2.4):



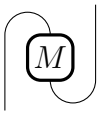
$$= (\langle 00| + \langle 11|)(M \otimes I_2)(|00\rangle + |11\rangle)$$

Now the right hand side is a sum of 4 terms, but the presence of the I_2 in the simple tensor $M \otimes I_2$ means the contribution will be zero unless the right indices agree:

$$\begin{aligned}
& \langle 00|M \otimes I_2|00\rangle + \langle 00|M \otimes I_2|11\rangle + \langle 11|M \otimes I_2|00\rangle + \langle 11|M \otimes I_2|11\rangle \\
= \; & \langle 0|M|0\rangle \underbrace{\langle 0|0\rangle}_{=1} + \langle 0|M|1\rangle \underbrace{\langle 0|1\rangle}_{=0} + \langle 1|M|0\rangle \underbrace{\langle 1|0\rangle}_{=0} + \langle 1|M|1\rangle \underbrace{\langle 1|1\rangle}_{=1} \\
= \; & M_{11} + M_{22} \\
= \; & \mathrm{Tr}_2(M).
\end{aligned}$$

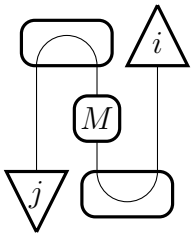
Exercise 3.4.3 — For $M \in M_2(\mathbb{C})$, show that $\mathrm{Tr}_2(M) =$  .

We now compute the *rotation*

$$R(M) := $$ 

by computing each ij-th entry:

$$\langle i|R(M)|j\rangle = \quad = (\langle 00i| + \langle 11i|)(I_2 \otimes M \otimes I_2)(|j00\rangle + |j11\rangle). \tag{3.4.4}$$



The right hand side of the above equation expands to

$$\underbrace{\langle 00i|I_2 \otimes M \otimes I_2|j00\rangle}_{\text{nonzero only if } i=0,\, j=0} + \underbrace{\langle 00i|I_2 \otimes M \otimes I_2|j11\rangle}_{\text{nonzero only if } i=1,\, j=0} + \underbrace{\langle 11i|I_2 \otimes M \otimes I_2)|j00\rangle}_{\text{nonzero only if } i=0,\, j=1} + \underbrace{\langle 11i|I_2 \otimes M \otimes I_2)|j11\rangle}_{\text{nonzero only if } i=1,\, j=1}.$$

Each of the four terms above is only non-zero for a specific value of i, j. In these cases, the above terms simplify to

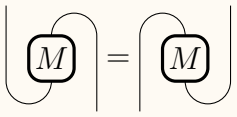
$$\underbrace{\langle 0|M|0\rangle}_{i=0,\, j=0} + \underbrace{\langle 0|M|1\rangle}_{i=1,\, j=0} + \underbrace{\langle 1|M|0\rangle}_{i=0,\, j=1} + \underbrace{\langle 1|M|1\rangle}_{i=1,\, j=1}.$$

We thus see that

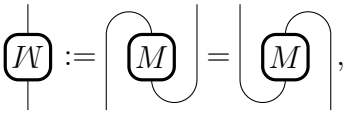
$$\langle i|R(M)|j\rangle = \langle j|M|i\rangle,$$

so $R(M) = M^T$. We record this fact here.

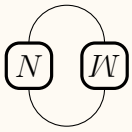
Proposition 3.4.5 — The rotation $R(M) =$  is equal to M^T. Moreover, $R(R(M)) = M$.

Exercise 3.4.6 — Show that  .

Now using the shorthand notation

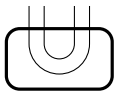


we can finally prove correctness of (3.2.5).

Exercise 3.4.7 — Prove that $\mathrm{Tr}(MN) =$ 

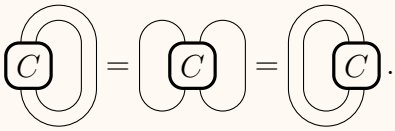
. Deduce $\mathrm{Tr}(MN) = \mathrm{Tr}(NM)$.

Using the graphical calculus, we also have a beautiful description of Tr_4. We use a double cap to represent the following un-normalized Bell state in $\mathbb{C}^4 \otimes \mathbb{C}^4$:

 $:= |0000\rangle + |0110\rangle + |1001\rangle + |1111\rangle,$

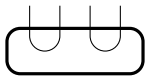
As the double cup is obtained from the single cup by inserting another cup inside of it, this un-normalized Bell state in $\mathbb{C}^4$ is obtained from inserting $\sqrt{2}|\beta\rangle$ inside $\sqrt{2}|\beta\rangle$.

Exercise 3.4.8 — Show that for all $C \in M_4(\mathbb{C})$,

$\mathrm{Tr}_4(C) =$ 

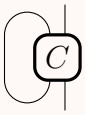
.

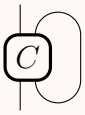
In the second diagram in the exercise above, we simply have the tensor product of two copies of $\sqrt{2}|\beta\rangle$:

 $:= \sqrt{2}|\beta\rangle \otimes \sqrt{2}|\beta\rangle = |0000\rangle + |0011\rangle + |1100\rangle + |1111\rangle.$

3.5 Graphical calculus for partial traces

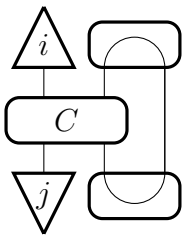
It might not be surprising that the partial traces $\mathrm{Tr}^A, \mathrm{Tr}^B$ correspond to only partially capping off C on the left/right respectively.

Proposition 3.5.1 — $\mathrm{Tr}^A(C) =$ 

and $\mathrm{Tr}^B(C) =$ 

.

Proof. We prove the second formula and leave the first to the reader. We compute the ij-th entry of the operator on the right hand side above by



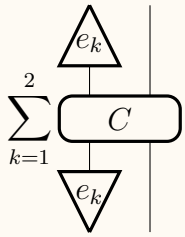
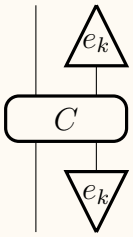
$= (\langle i00| + \langle i11|)(C \otimes I_2)(|j00\rangle + |j11\rangle).$

Again, the presence of the I_2 on the right hand side of the simple tensor means that the cross-terms do not contribute:

$$= \langle i00|(C \otimes I_2)|j00\rangle + \langle i11|(C \otimes I_2)|j11\rangle$$
$$= \langle i0|C|j0\rangle + \langle i1|C|j1\rangle = \sum_{k=0}^{1} \langle ik|C|jk\rangle =: \mathrm{Tr}^B(C)$$

as desired. ∎

Remark 3.5.2 — The graphical interpretation of Proposition 2.11.15 is the following alternative graphical formulas for $\mathrm{Tr}^A, \mathrm{Tr}^B$:

$\mathrm{Tr}^A(C) = \sum_{k=1}^{2}$  and $\mathrm{Tr}^B(C) = \sum_{k=1}^{2}$ 

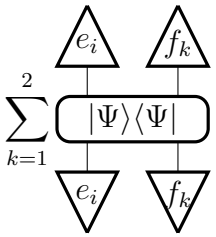
where $\{|e_1\rangle, |e_2\rangle\}$ is any ONB of $\mathbb{C}^2$.

Exercise 3.5.3 — Give a second graphical proof of the above formulas using Exercise 3.4.2.

We now connect the partial trace to the discussion of local measurement after Definition 2.5.5. When we want to measure an observable $M \in M_2(\mathbb{C})$ locally on the A qubit, and M has 2 distinct eigenvalues λ_1, λ_2 and corresponding ONB of eigenvectors $\{|e_1\rangle, |e_2\rangle\}$, we may choose any ONB $\{|f_1\rangle, |f_2\rangle\}$ of the B qubit space $\mathbb{C}^2$. Then $\{|e_1 f_1\rangle, |e_1 f_2\rangle\}$ is an ONB of eigenstates for λ_1 and $\{|e_2 f_1\rangle, |e_2 f_2\rangle\}$ is an ONB of eigenstates for λ_2, and the outcome of the measurement is λ_i with probability

$$|\langle e_i f_1|\Psi\rangle|^2 + |\langle e_i f_2|\Psi\rangle|^2.$$

Observe that this probability is exactly equal to

$\langle e_i f_1|\Psi\rangle\langle\Psi|e_i f_1\rangle + \langle e_i f_2|\Psi\rangle\langle\Psi|e_i f_2\rangle = \sum_{k=1}^{2}$  $= \langle e_i| \mathrm{Tr}^B(|\Psi\rangle\langle\Psi|)|e_i\rangle,$

which is exactly the expected value of the reduced density $(|\Psi\rangle\langle\Psi|)^A$ in the state $|e_i\rangle$, as expected. Indeed, we have

$$\begin{aligned}\langle e_i|\operatorname{Tr}^B(|\Psi\rangle\langle\Psi|)|e_i\rangle &= \operatorname{Tr}_2(\operatorname{Tr}^B(|\Psi\rangle\langle\Psi|)\cdot|e_i\rangle\langle e_i|)\\ &= \operatorname{Tr}_4(|\Psi\rangle\langle\Psi|\cdot(|e_i\rangle\langle e_i|\otimes I_2))\\ &= \langle\Psi|(|e_i\rangle\langle e_i|\otimes I_2)|\Psi\rangle,\end{aligned}$$

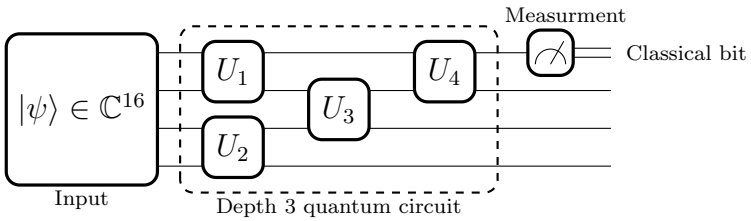
which is exactly the projective measurement of $|\Psi\rangle$ with respect to the orthogonal projection $|e_i\rangle\langle e_i|\otimes I_2$ onto the λ_i eigenspace of C as discussed in Example 2.7.6.

Exercise 3.5.4 — Repeat the above analysis for local measurements for the B qubit.

3.6 Quantum circuits

A *quantum circuit* is a sequence of initializations of quantum states, unitary operators, measurements, and classical information channels.

Usually, we represent a quantum circuit in the diagrammatic calculus by rotating our usual diagrams clockwise by $\pi/2 = 90°$ in order to save space on the page. Unitary gates are applied to an initial state, labelled on the left, and the process is read left to right. Here is an example:



The *depth* of a quantum circuit is the minimum number of unitary operators which need to be applied in sequence to achieve the circuit. The quantum circuit above has depth 3, since we may apply $U_1 \otimes U_2$ as the first operator instead of applying $U_1 \otimes I_4$ and then $I_4 \otimes U_2$ in sequence.

A *measurement* in a quantum circuit is an observation of a single qubit using some observable with eigenstates $|0\rangle, |1\rangle$, which produces a classical bit from the qubit. It is represented by the symbol in the above cartoon. A circuit may also use this classical bit as input to another unitary gate, as we will see with the quantum teleportation protocol in §3.7 below.

Warning — We typically write the output of a function f applied to an input x as $f(x)$. If we then apply another function g, we obtain $g(f(x))$. So the composite of first doing f and then doing g is $g \circ f$, where the symbols have been swapped. In computer science, notation for applying a function is often given as $x.f$, so applying g after f yields $x.f.g$, and the composite of applying f and then g is $f.g$. Thus if we are writing

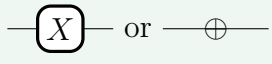
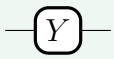
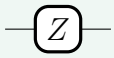
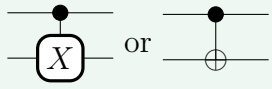
left-to-right, computer science notation seems more intuitive, since the order of symbols has not swapped.

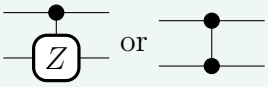
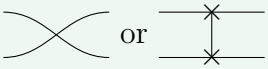
Quantum circuit diagrams are also written left-to-right, as they are computer programs for a quantum computer. We can think of this diagram as applying our quantum gates to our state $|\psi\rangle$ by writing them on the right, similar to the $x.f$ notation. But now something terrible has happened; if we apply a bra $\langle\eta|$ to a ket $|\psi\rangle$, we would obtain the notation $|\psi\rangle.\langle\eta|$, which looks like it should represent an operator, and not a scalar.

There are several ways to fix this; we could write our quantum circuits from right-to-left, similar to mathematics notation. This offers the advantage that $A|\psi\rangle$ would then be represented in a diagram with A to the left of $|\psi\rangle$, but it is against the convention of the rest of the literature on quantum circuits. We could write our quantum circuits from bottom-to-top, keeping our graphical calculus from the previous sections, but this takes too much vertical space on the page to be efficient. Finally, we can just admit defeat and live with the consequences of our choices and those that came before us.

Unfortunately, we go with this third option; such is life. As such, the reader should be prepared to swap orders of operators in circuit diagrams written from left-to-right with formal mathematical equations which are written right-to-left (input to output).

Example 3.6.1 — Here are examples of common unitary gates.

Name	Graphical notation	Operator
Identity I_2		$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
Pauli X		$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
Pauli Y		$\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$
Pauli Z		$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Hadamard H		$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$
CNOT or CX		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

CZ		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$
Swap		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Exercise 3.6.2 — The *controlled M gate* C_1M is the operator satisfying the rule

$$|ij\rangle \longmapsto \begin{cases} |0j\rangle & \text{if } i = 0 \\ |1\rangle \otimes M|j\rangle & \text{if } i = 1. \end{cases}$$

Compute the 4×4 matrix of C_1M.

Exercise 3.6.3 — The *controlled M gate* C_2M is the operator satisfying the rule

$$|ij\rangle \longmapsto \begin{cases} |i0\rangle & \text{if } j = 0 \\ M|i\rangle \otimes |1\rangle & \text{if } j = 1. \end{cases}$$

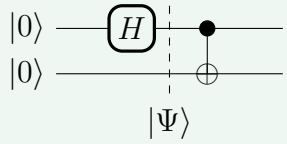
Compute the 4×4 matrix of C_2M.

Exercise 3.6.4 — Compute C_1Z and C_2Z.

Exercise 3.6.5 — Find all $M \in M_2(\mathbb{C})$ such that $C_1M = C_2M$.

Example 3.6.6 — The following quantum circuit prepares the Bell state $|\beta\rangle = \frac{1}{\sqrt{2}}(|00\rangle +$

$|11\rangle$) from the initial state $|00\rangle$.

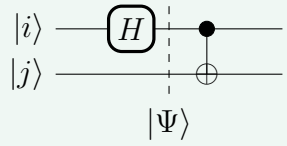


After applying H to the first $|0\rangle$, we obtain

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|0\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle),$$

and the CNOT gate preserves $|00\rangle$, but swaps $|10\rangle$ with $|11\rangle$.

Example 3.6.7 — Indeed, by changing our initial state, we can prepare any element in the Bell basis. The following quantum circuit prepares the Bell state $|\beta_{ij}\rangle$ from the initial state $|ij\rangle$.



After applying H to the first $|0\rangle$, we obtain

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (-1)^i|1\rangle)|j\rangle = \frac{1}{\sqrt{2}}(|0j\rangle + (-1)^i|1j\rangle).$$

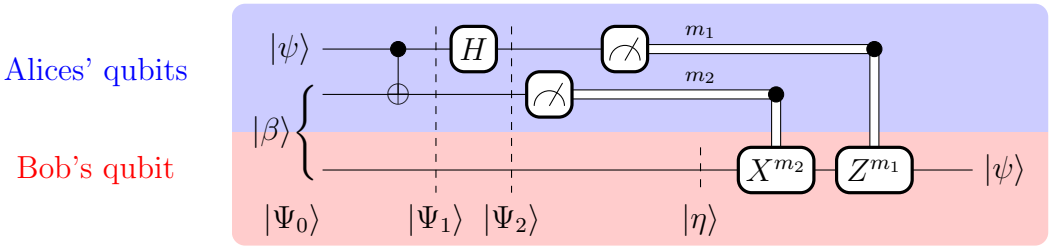
The CNOT gate preserves $|0j\rangle$, but swaps $|1j\rangle$ with $|1(1-j)\rangle$. Thus we obtain

$$|\beta_{ij}\rangle = \frac{1}{\sqrt{2}}(|0j\rangle + (-1)^i|1(1-j)\rangle).$$

3.7 Quantum teleportation protocol

The diagram below represents the quantum teleportation protocol [BBC+93] which transports a state $|\psi\rangle$ controlled by Alice safely across any distance to Bob. As input, Alice and Bob must share an entangled Bell state. The top two qubits below drawn in red are controlled by Alice, and the bottom qubit drawn in blue is controlled by Bob. Alice performs two unitary gates to her two qubits, and then measures both qubits; the measurements can happen in either order, or at the same time, as long as she gets a classical bit for both

channels. (Any diagonal self-adjoint operator with distinct eigenvalues will do.)



(3.7.1)

Suppose Alice's initial state is $|\psi\rangle = a|0\rangle + b|1\rangle$, so that the initial 3-qubit state is

$$|\Psi_0\rangle = |\psi\rangle|\beta\rangle = (a|0\rangle + b|1\rangle)\left(\frac{|00\rangle + |11\rangle}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}(a|0\rangle(|00\rangle + |11\rangle) + b|1\rangle(|00\rangle + |11\rangle)).$$

After Alice applies the CNOT gate, the 3-qubit state is

$$|\Psi_1\rangle = \frac{1}{\sqrt{2}}(a|0\rangle(|00\rangle + |11\rangle) + b|1\rangle(|10\rangle + |01\rangle)).$$

After she applies the H gate, the the 3-qubit state is

$$\begin{aligned}|\Psi_2\rangle &= \frac{1}{2}(a(|0\rangle + |1\rangle)(|00\rangle + |11\rangle) + b(|0\rangle - |1\rangle)(|10\rangle + |01\rangle)).\\ &= \frac{a}{2}(|000\rangle + |011\rangle + |100\rangle + |111\rangle) + \frac{b}{2}(|001\rangle + |010\rangle - |101\rangle - |110\rangle).\end{aligned}$$

We now write the state $|\Psi_2\rangle$ as a linear combination of eigenstates for an operator of the form $M \otimes I_2$ where $M \in M_4(\mathbb{C})$ is diagonal with 4 distinct eigenvalues, which corresponds to the 4 outcomes 00, 01, 10, 11 of measuring both of Alice's qubits.

$$|\Psi_2\rangle = \frac{1}{2}(a|000\rangle + b|001\rangle) + \frac{1}{2}(a|011\rangle + b|010\rangle) + \frac{1}{2}(a|100\rangle - b|101\rangle) + \frac{1}{2}(a|111\rangle - b|110\rangle).$$

Observe that all of these states have the same amplitude, so all outcomes are equally likely. This means when Alice measures her two qubits to obtain classical bits, we obtain the state $|\eta\rangle$ as follows.

Alice's measurements	Bob's state $\|\eta\rangle$	Bob's correction
00	$a\|0\rangle + b\|1\rangle$	I_2
01	$a\|1\rangle + b\|0\rangle$	X
10	$a\|0\rangle - b\|1\rangle$	Z
11	$a\|1\rangle - b\|0\rangle$	ZX

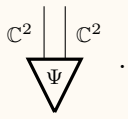
(3.7.2)

If Alice measures 00, then Bob has Alice's state $|\psi\rangle$ as planned. Otherwise, he must apply a correction to obtain $|\psi\rangle$. This correction factor is exactly $Z^{m_1}X^{m_2}$, where $m_1, m_2 \in \{0, 1\}$ are Alice's measurements. Thus Alice transmits her measurement via a classical channel to Bob, which allows him to reconstruct $|\psi\rangle$ exactly.

Exercise 3.7.3 — Explain the connection between the quantum teleportation protocol, (2.4.8), Exercise 3.4.1, and Example 3.6.6.

Exercise 3.7.4 — How do Bob's correction operators change if Alice and Bob initially share one of the other 3 Bell states?

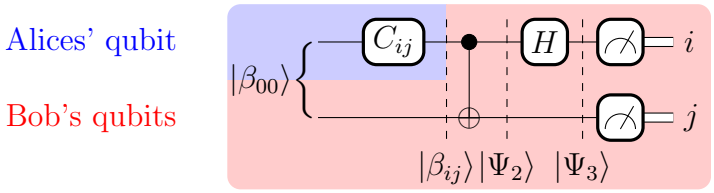
Exercise 3.7.5 — Design a quantum teleportation protocol which allows Alice to transport a 2-qubit state $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ to Bob. Recall we may represent $|\Psi\rangle$ in the graphical calculus by



Can we simply transport the first qubit via (3.7.1) and then transport the second qubit via (3.7.1)? Why or why not?

3.8 Superdense coding

The superdense coding protocol [BW92] allows Alice to communicate 2 classical bits to Bob using 1 qubit, provided this qubit is entangled with a qubit controlled by Bob. As in the quantum teleportation protocol, Alice and Bob share an initial Bell state $|\beta_{00}\rangle$.



(3.8.1)

Alice encodes her 2 classical bits ij in her qubit using the encoding operator C_{ij} from (2.4.8) or (3.7.2) above. The total 2-qubit state then becomes $|\beta_{ij}\rangle$ as in Example 2.4.7, and Alice transmits her qubit over to Bob. Bob then applies a CNOT gate and then a Hadamard gate H on the first qubit to obtain

ij	$\|\beta_{ij}\rangle$	$\|\Psi_2\rangle$	$\|\Psi_3\rangle$
00	$\frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$	$\frac{1}{\sqrt{2}}(\|00\rangle + \|10\rangle)$	$\|00\rangle$
01	$\frac{1}{\sqrt{2}}(\|01\rangle + \|10\rangle)$	$\frac{1}{\sqrt{2}}(\|01\rangle + \|11\rangle)$	$\|01\rangle$
10	$\frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$	$\frac{1}{\sqrt{2}}(\|00\rangle - \|10\rangle)$	$\|10\rangle$
11	$\frac{1}{\sqrt{2}}(\|01\rangle - \|10\rangle)$	$\frac{1}{\sqrt{2}}(\|01\rangle - \|11\rangle)$	$\|11\rangle$

To see how to easily compute $|\Psi_3\rangle$ from $|\Psi_2\rangle$ when applying $H \otimes I_2$, we note that H is a self-adjoint unitary, and thus equal to its own inverse. This means that under H, we have

$$|0\rangle \overset{H}{\longmapsto} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \overset{H}{\longmapsto} |0\rangle \qquad \text{and} \qquad |1\rangle \overset{H}{\longmapsto} \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \overset{H}{\longmapsto} |1\rangle.$$

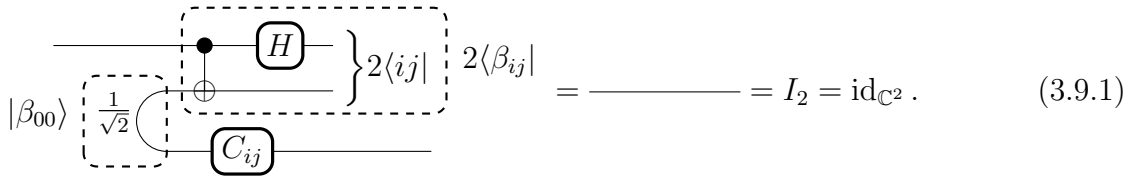
We now observe that $|\Psi_3\rangle = |ij\rangle$, so when Bob measures his qubits, he exactly obtains Alice's bits ij.

As a final remark, we could read the quantum circuit (3.8.1) *backwards* by taking adjoints. We would then see the preparation of the Bell state $|\beta_{ij}\rangle$ from the initial state $|ij\rangle$ by applying $(H \otimes I_2)$ and then CNOT as in Example 3.6.7. This again motivates the choice of C_{ij} as in (3.7.2).

3.9 Quantum teleportation correctness implies superdense coding correctness

In this section, we show that one can deduce the correctness of the superdense coding protocol from correctness of the quantum teleportation protocol using the graphical calculus.

The correctness of the quantum teleportation protocol (3.7.1) when the first two qubits are measured to be $i, j \in \{0, 1\}$ respectively is given by the following equation of string diagrams.



$$= \text{————} = I_2 = \mathrm{id}_{\mathbb{C}^2}\,. \tag{3.9.1}$$

Now the diagram on the left in (3.9.1) above looks very different than (3.7.1); the main differences are:

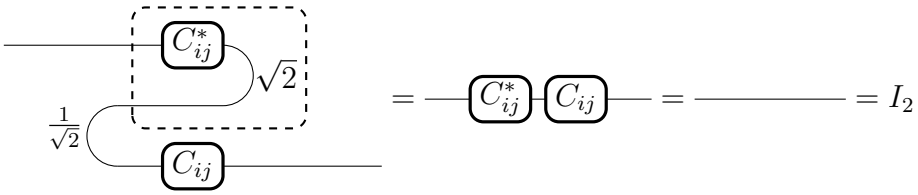
- We have drawn the Bell state $|\beta\rangle$ as a cup, but multiplied by a factor of $1/\sqrt{2}$.
- Instead of measuring i, j for the first two qubits, we have included a bra labelled $2\langle ij|$. In the graphical calculus, this *ensures* that we only measure i, j at this point. The reason we must introduce a multiplicative factor of 2 is the normalization that occurs when we measure i, j as the first two qubits in $|\Psi_2\rangle$ in (3.7.1). Instead of using the Born rule that says we always obtain a state when measuring quantum observables, applying the operator $\langle ij| \otimes I_2$ to $|\Psi_2\rangle$ gives

  $$(\langle ij| \otimes I_2)|\Psi_2\rangle = \frac{1}{2}|\eta\rangle,$$

  as $|\Psi_2\rangle$ is a superposition of 4 eigenstates corresponding to the 4 outcomes 00, 01, 10, 11 of Alice's local measurement.

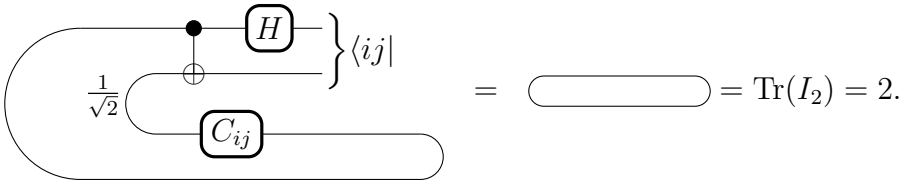
By taking adjoints in Example 3.6.7, we see that the operator from $\mathbb{C}^2 \otimes \mathbb{C}^2 \to \mathbb{C}$ in the top right dashed box is exactly $2\langle\beta_{ij}|$. Now by (2.4.8), the correction operator C_{ij} satisfies $|\beta_{ij}\rangle = (C_{ij} \otimes I_2)|\beta_{00}\rangle$, so taking adjoints, we have $\langle\beta_{ij}| = (C_{ij}^* \otimes I_2)\langle\beta_{00}|$. Replacing $2\langle\beta_{ij}|$

by $2(C_{ij}^* \otimes I_2)\langle\beta_{00}|$ in (3.9.1), we obtain

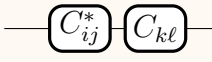


as promised above.

We now take the left hand side of (3.6.7), multiply by $1/2$, and take the *trace* to close off the diagram to obtain:



Exercise 3.9.2 — Show that if we replace C_{ij} with $C_{k\ell}$ in (3.9.1), we would obtain



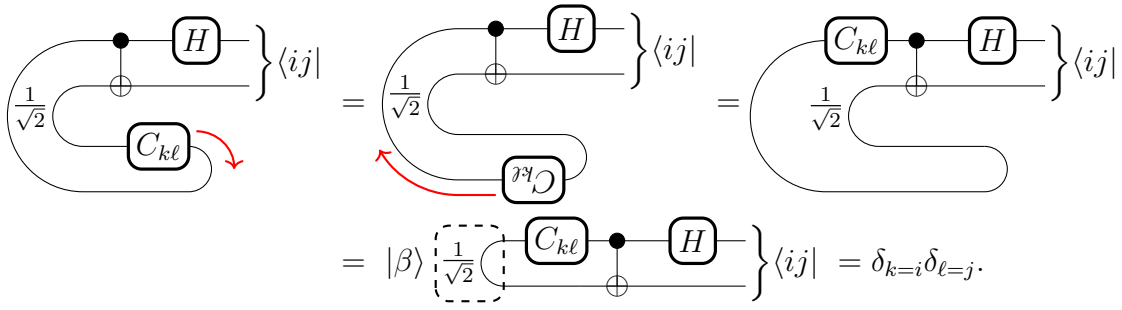
$= C_{k\ell}C_{ij}^* =$

$ij\backslash k\ell$	00	01	10	11
00	I_2	X	Z	ZX
01	X	I_2	ZX	Z
10	Z	XZ	I_2	$-X$
11	XZ	Z	$-X$	I_2

.

(Note that $ZXZ = -X$.) Deduce that applying the trace, we get zero unless $i = k$ and $j = \ell$.

We will see that the above exercise proves correctness of the superdense coding protocol. Indeed, using the graphical calculus, we deform the left hand side of (3.6.7) with C_{ij} replaced by $C_{k\ell}$ to first move the $C_{k\ell}$ around the bends, and second to contract the zig-zag:



The final diagram is exactly the superdense coding protocol, with measurements replaced by applying the bra $\langle ij|$. Since we began with the state $|\beta\rangle$ and applied the unitary gates $C_{k\ell} \otimes I_2$, then CNOT, and finally $H \otimes I_2$, we obtain another state in the span of $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. The above formula says that we in fact obtain the state $|k\ell\rangle$.

3.10 Quantum key distribution

The structure of this section is modeled after https://mpl.mpg.de/fileadmin/user_upload/Chekhova_Research_Group/Lecture_4_12.pdf and https://cklixx.people.wm.edu/teaching/QC2021/QC-chapter4.pdf, [[second probably based on Quantum Computing 1st Edition by Mikio Nakahara and Tetsuo Ohmi looking at table of contents.]]

Suppose Alice has a *secret message*, a bit string of length n, that she wants to securely transmit to Bob. Using the *one-time pad* cryptography method, if Alice and Bob share a secret *code word*, another random bit string of length n that only Alice and Bob know, then Alice can code her secret message by adding it to the code word modulo 2.

secret message		0110001101
code word	+	1011010111
public message		1101011010

Alice then publicly transmits the *public message*, which Bob can then decode by adding the code word back again modulo 2.

public message		1101011010
code word	+	1011010111
secret message		0110001101

If Alice and Bob are the only people to ever know this random code word, and if they never reuse this key, this system is impossible to break.

The two protocols BB84 and E91 that follow below give a way that Alice and Bob can share a secret random code word.

3.10.1 BB84 protocol

In the BB84 protocol [BB84], Alice chooses a random bit string $a_1a_2\cdots a_k$ of length considerably longer that n. She then chooses a second random bit string $b_1b_2\cdots b_k$ of the same length. Using b_i, she encodes a_i as a qubit in a particular ONB; if $b_i = 0$, she encodes a_i using the Z-basis, and if $b_1 = 1$, she encodes a_i using the X-basis:

$$|\psi_{a_ib_i}\rangle := \begin{cases} \begin{cases} |0\rangle & \text{if } a_i = 0 \\ |1\rangle & \text{if } a_i = 1 \end{cases} & \text{if } b_i = 0 \\ \begin{cases} |+\rangle & \text{if } a_i = 0 \\ |-\rangle & \text{if } a_i = 1 \end{cases} & \text{if } b_i = 1 \end{cases}$$

She then transmits each state $|\psi_{a_ib_i}\rangle$ over to Bob, who chooses the decoding basis randomly. That is, Bob chooses another bit string $b'_1b'_2\cdots b'_k$ randomly and measures the qubit $|\psi_{a_ib_i}\rangle$ in the ONB for b'_i, resulting in a measurement a'_i. If $b'_i = b_i$, then $a'_i = a_i$, so Bob gets Alice's bit back. Otherwise, since the Z and X-bases are mutually unbiased, there is only a 50% chance that $a'_i = a_i$.

Once Bob has measured every qubit sent by Alice, Alice and Bob both publicly announce the bit strings $b_1 b_2 \cdots b_k$ and $b'_1 b'_2 \cdots b'_k$ so that they know when they have chosen the same bases. They throw out all the i such that $b_i \neq b'_i$, and they are left with bits in which $a_j = a'_j$.

a_i	1	0	1	1	0	1	0	1	1	1
b_i	0	1	1	0	0	0	1	1	0	1
$\lvert\psi_{a_i b_i}\rangle$	$\lvert 1\rangle$	$\lvert +\rangle$	$\lvert -\rangle$	$\lvert 1\rangle$	$\lvert 0\rangle$	$\lvert 1\rangle$	$\lvert +\rangle$	$\lvert -\rangle$	$\lvert 1\rangle$	$\lvert -\rangle$
a'_i	?	**0**	?	?	**0**	?	**0**	?	**1**	**1**
b'_i	1	**1**	0	1	**0**	0	**1**	0	**0**	**1**

In the table above, the bold bits represent those which agree between Alice and Bob; we write ? for any a'_i for which $b_i \neq b'_i$.

In order to test for eavesdroppers, they reveal part of these shared bits in order to see if they agree close to 100% of the time. (Usually this is around 20% of the remaining bit string.) If they correlate very well, Alice and Bob conclude they now share a secret code word of some length, and they can use this code word to transmit messages. If they do not correlate well, they conclude the existence of some eavesdropper Eve. Since Eve cannot clone qubits by the No Cloning Theorem 2.3.6, her best strategy is to transmit exactly what she measures, but since she also does not know b_i, she will likely introduce error into the system.

3.10.2 E91 protocol

One advantage of the E91 protocol [Eke91] is that we can use the qubits where Alice and Bob choose different bases to perform the eavesdropper test. In this protocol, we suppose that Alice and Bob share as many copies of the singlet state $|\Psi^-\rangle$ as they need, each controlling one qubit. Alice measures her qubit in one of the following three bases at random:

operator	Z	H	X
eigenbasis	$\{\lvert 0\rangle, \lvert 1\rangle\}$	$\{\lvert h_+\rangle, \lvert h_-\rangle\}$	$\{\lvert +\rangle, \lvert -\rangle\}$

Bob measures his qubit in one of the following three bases at random:

operator	Z	H	ZHZ
eigenbasis	$\{\lvert 0\rangle, \lvert 1\rangle\}$	$\{\lvert h_+\rangle, \lvert h_-\rangle\}$	$\{Z\lvert h_+\rangle, Z\lvert h_-\rangle\}$

We note that when Alice or Bob measures $+1$, they record this as 0, and when they measure -1, they record this as 1. (They are thus actually measuring a projection associated to a unitary from Exercise 1.9.11.)

Now Alice and Bob measure as many qubits as they like, recording the basis in which they measured together with the outcome. When Alice and Bob both choose the same ONB, then their results are 100% anti-correlated by Exercise 2.4.3.

Alice's basis	Z	Z	X	**H**	X	H	**Z**	X	H
Alice's measurement	?	?	?	**0**	?	?	**1**	?	?
Bob's basis	H	ZHZ	H	**H**	Z	ZHZ	**Z**	H	ZHZ
Bob's measurement	?	?	?	**1**	?	?	**0**	?	?

Exercise 3.10.1 — What is the probability that Alice and Bob choose the same ONB in which to measure?

Now with the remaining measurements, Alice and Bob have chosen distinct bases. Notice, however, that a subset of Alice's measurements are in the eigenbases for $A_1 = Z$ and $A_2 = X$, and a subset of Bob's measurements are in the eigenbases for $B_1 = H$ and $B_2 = ZHZ$. They can thus take these measurements and perform a CHSH test to check for the presence of an eavesdropper!

Exercise 3.10.2 — With $A_1 = Z, A_2 = X$ and $B_1 = H, B_2 = ZHZ$, calculate the expected values of $A_1 \otimes B_1$, $A_1 \otimes B_2$, $A_2 \otimes B_1$, and $A_2 \otimes B_2$ in the singlet state $|\Psi^-\rangle$. Then show that the expected value of the observable

$$-A_1 \otimes B_1 - A_1 \otimes B_2 + A_2 \otimes B_1 - A_2 \otimes B_2 = -A_1 \otimes (B_1 + B_2) + A_2 \otimes (B_1 - B_2)$$

is $2\sqrt{2}$.

If no eavesdropper is present, the CHSH test should reveal a number close to $2\sqrt{2}$ for the above observable. (Note here that instead of their measurements $0, 1$, Alice and Bob should use $+1, -1$ when performing the CHSH test.) However, if eavesdropper Eve is present, her best strategy is to transmit exactly what she hears, which effectively introduces a hidden variable into the model. By similar reasoning as for the CHSH inequality, this hidden variable would make the expectation for this observable lie between -2 and 2. Thus if Alice and Bob do not witness a violation of this Bell inequality, they can infer the existence of an eavesdropper begin the process anew.

A Appendix: Supplemental linear algebra

A.1 Introduction

In this appendix, we provide an extremely short presentation of various abstract notions from linear algebra in order to supplement the material in the previous three sections. We only work in the vector/Hilbert space $\mathbb{C}^n$ rather than an abstract vector/Hilbert space for simplicity. Sometimes, we work in an arbitrary subspace $V \subset \mathbb{C}^n$, and proofs are written in such a way as to apply verbatim to the case when V is an abstract vector/Hilbert space.

We try to present results/theorems as algorithms whenever possible. These algorithms can typically be implemented in a computer algebra program like `Mathematica`, `MATLAB`, `Maple`, or `Python` using the `numpy.linalg` library.

This appendix leaves out several important topics, such as determinants and characteristic polynomials; these were presented for 2×2 matrices only in §1.3 and Algorithm 1.9.6 above. While a typical linear algebra course will teach 3×3 determinants, quantum information theory quickly jumps from 2×2 matrices to 4×4 matrices, where determinants become quite cumbersome and time consuming.

A.2 Linear independence

In this section, we give a rigorous treatment of linear independence and span.

Definition A.2.1 — Vectors $|\psi_1\rangle, \ldots, |\psi_k\rangle \in \mathbb{C}^n$ are said to be *linearly independent* if

$$\sum_{i=1}^{k} c_i |\psi_i\rangle = c_1|\psi_1\rangle + \cdots c_k|\psi_k\rangle = 0 \qquad \Longrightarrow \qquad c_1 = c_2 = \cdots = c_k = 0.$$

If $|\psi_1\rangle, \ldots, |\psi_k\rangle$ are not linearly independent, we say they are *linearly dependent*.

Example A.2.2 — Any set if vectors $|\psi_1\rangle, \ldots, |\psi_k\rangle$ whose zero entries form a *strict staircase pattern* are linearly independent, e.g.,

$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 2 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 6 \\ 0 \\ 7 \\ 1 \\ 0 \end{bmatrix}.$$

Here, we notice that there are 4 zeroes at the bottom in the vector on the left, then 2 zeroes at the bottom in the vector in the middle, and then 1 zero at the bottom in the vector on the right. Moreover, no two vectors have the same block of zeroes at the bottom.

To see that these vectors are linearly independent, we set an arbitrary linear combination equal to zero:

$$c_1 \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 2 \\ 5 \\ 0 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 6 \\ 0 \\ 7 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2c_1 + c_2 + 6c_3 \\ 2c_2 \\ 5c_2 + 7c_3 \\ c_3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Looking at the 4th entry, we see that $c_3 = 0$.

$$\begin{bmatrix} 2c_1 + c_2 + 6c_3 \\ 2c_2 \\ 5c_2 + 7c_3 \\ c_3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \implies c_3 = 0 \implies \begin{bmatrix} 2c_1 + c_2 \\ 2c_2 \\ 5c_2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Now looking at the 2nd or 3rd entry, we see $c_2 = 0$. We then conclude that $c_1 = 0$, so the only linear combination which gives the zero vector is the trivial linear combination.

The fully general case can be proven using the *principle of mathematical induction.*

Example A.2.3 — The following vectors are linearly dependent:

$$\begin{bmatrix} 1 \\ 4 \\ 7 \\ 10 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 8 \\ 11 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 9 \\ 12 \end{bmatrix}.$$

Indeed, we calculate that

$$\begin{bmatrix} 1 \\ 4 \\ 7 \\ 10 \end{bmatrix} - 2 \begin{bmatrix} 2 \\ 5 \\ 8 \\ 11 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \\ 9 \\ 12 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Thus there is a non-trivial linear combination which is equal to the zero vector.

Exercise A.2.4 — Suppose $|\psi\rangle \in \mathbb{C}^n$ is linearly dependent. What is $|\psi\rangle$?

Remark A.2.5 — If $|\psi_1\rangle, \ldots, |\psi_k\rangle$ are linearly dependent, then there is a minimal $1 \leq j \leq k$ such that $|\psi_j\rangle$ is a linear combination of $|\psi_1\rangle, \ldots, |\psi_{j-1}\rangle$.

Exercise A.2.6 — Suppose $|\psi_1\rangle, \ldots, |\psi_k\rangle \in \mathbb{C}^n$ and A is an $m \times n$ matrix such that $A|\psi_1\rangle, \ldots, A|\psi_k\rangle \in \mathbb{C}^m$ are linearly independent. Prove $|\psi_1\rangle, \ldots, |\psi_k\rangle$ are linearly independent.

Proposition A.2.7 — The vectors $|\psi_1\rangle, \ldots, |\psi_k\rangle \in \mathbb{C}^n$ are linearly independent if and only if the matrix

$$A := \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_k\rangle \end{bmatrix}$$

has trivial null space/kernel, i.e., $\ker(A) = \{0\}$.

Proof. The null space of A is exactly the vectors

$$|c\rangle = \begin{bmatrix} c_1 \\ \ddots \\ c_k \end{bmatrix} \in \mathbb{C}^k$$

such that $A|c\rangle = 0$. But $A|c\rangle$ is exactly the corresponding linear combination of the columns of A:

$$A|c\rangle = \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_k\rangle \end{bmatrix} \begin{bmatrix} c_1 \\ \ddots \\ c_k \end{bmatrix} = c_1|\psi_1\rangle + \cdots c_k|\psi_k\rangle.$$

This means the null space exactly consists of the vectors corresponding to linear combinations of $|\psi_1\rangle, \ldots, |\psi_k\rangle$ which give the zero vector in $\mathbb{C}^n$. ∎

A.3 Elementary row operations and Gaussian elimination

Definition A.3.1 — Suppose A is an $m \times n$ matrix. There are three *elementary row operations* one can perform on A:

(E1) Multiply a row by a non-zero scalar,

(E2) Swap two rows, and

(E3) Replace row j of A with a scalar multiple $c \in \mathbb{C}$ times row i added to row j.

These row operations are performed by multiplying A on the left by an *elementary*

matrix, which is the matrix obtained from the identity matrix I_m by performing the same elementary row operation.

Example A.3.2 (Elementary operation (E1)) — We multiply the first row by 2, which is achieved by multiplying on the left by the corresponding elementary matrix:

$$\begin{bmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\end{bmatrix} \rightsquigarrow \begin{bmatrix}2&4&6&8\\5&6&7&8\\9&10&11&12\end{bmatrix} = \underbrace{\begin{bmatrix}2&0&0\\0&1&0\\0&0&1\end{bmatrix}}_{\text{Elementary matrix}} \begin{bmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\end{bmatrix}$$

Example A.3.3 (Elementary operation (E2)) — We swap the first two rows, which is achieved by multiplying on the left by the corresponding elementary matrix:

$$\begin{bmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\end{bmatrix} \rightsquigarrow \begin{bmatrix}5&6&7&8\\1&2&3&4\\9&10&11&12\end{bmatrix} = \underbrace{\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}}_{\text{Elementary matrix}} \begin{bmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\end{bmatrix}$$

Example A.3.4 (Elementary operation (E3)) — We replace the second row $R2$ with $4 \cdot R1 + R2$ (4 times the first row plus the second row), which is achieved by multiplying on the left by the corresponding elementary matrix:

$$\begin{bmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\end{bmatrix} \rightsquigarrow \begin{bmatrix}1&2&3&4\\9&14&19&24\\9&10&11&12\end{bmatrix} = \underbrace{\begin{bmatrix}1&0&0\\4&1&0\\0&0&1\end{bmatrix}}_{\text{Elementary matrix}} \begin{bmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\end{bmatrix}$$

Observe each of elementary row operation is invertible, and its inverse is an elementary row operation of the same type. This means each elementary matrix is invertible as well. Here is the formal definition of invertibility.

Definition A.3.5 — We call an $n \times n$ matrix A *invertible* if there is another $n \times n$ matrix B such that $AB = I_n$ and $BA = I_n$. We call B *an inverse* of A.

Observe that being a square matrix is a necessary condition to being invertible.

Example A.3.6 — We give inverses of the elementary matrices from the above exam-

ples:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Exercise A.3.7 — Suppose that A, B, C are $n \times n$ matrices such that $AB = I_n$ and $BC = I_n$. Prove that $A = C$. Deduce that inverses are unique when they exists.

Exercise A.3.8 — Prove that if A, B are two invertible $n \times n$ matrices, then AB is also invertible.
Hint: Try $B^{-1}A^{-1}$.

Exercise A.3.9 — Show that $(AB)^T = B^T A^T$. Deduce that A is invertible if and only if A^T is invertible, in which case $(A^T)^{-1} = (A^{-1})^T$.

Definition A.3.10 — We say two $m \times n$ matrices A, B are *row equivalent* if B can be obtained from A by performing elementary row operations, equivalently, if B can be obtained from A by multiplying by elementary matrices.

Since elementary matrices are invertible, one can go back and forth between row equivalent matrices.

Exercise A.3.11 — Prove that row equivalence is an *equivalence relation*, i.e., show:

- (reflexive) for every matrix A, A is row equivalent to A,
- (symmetric) if A is row equivalent to B, then B is row equivalent to A, and
- (transitive) if A is row equivalent to B and B is row equivalent to C, then A is row equivalent to C.

Definition A.3.12 — An $m \times n$ matrix A is said to be in *row echelon form (ref)* if

- The first non-zero entry of every row (if one exists) is equal to 1. We call these 1's *pivots*.
- If row $i+1$ has a pivot, then row i has a pivot *strictly to the left* of the pivot in row $i+1$.

We further say that A is in *reduced row echelon form (rref)* if

- Every entry above a pivot is equal to zero.

Example A.3.13 — The first matrix below is in row echelon form, but the second and third are not.

$$\underbrace{\begin{bmatrix} 1 & 2 & 6 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{in ref}} \qquad \underbrace{\begin{bmatrix} 1 & 2 & 6 \\ 0 & 2 & 0 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{not in ref}} \qquad \underbrace{\begin{bmatrix} 6 & 1 & 2 \\ 7 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{not in ref}}$$

In fact, every matrix is row equivalent to one in row echelon form, and every matrix is row equivalent to a *unique* matrix in reduced row echelon form. The proof of this fact is the Gaussian Elimination Algorithm A.3.14 below.

Algorithm A.3.14 (Gaussian Elimination) — The following algorithm takes an $m \times n$ matrix A and produces a row equivalent matrix $\text{ref}(A)$ in row echelon form.

Step 1: If A is in row echelon form, then return A.

Step 2: Find the first column with a non-zero entry. Pick any row R_i with a non-zero entry and perform an elementary row operation of type (E1) on row i to make this entry 1. Call this new matrix A_1.

$$\begin{bmatrix} 0 & \cdots & 0 & * & * & \cdots & * \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & * & \cdots & * \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & * & * & \cdots & * \end{bmatrix}$$

Step 3: Apply elementary row operations of type (E2) to swap rows 1 and i so that the 1 appears in the first row. Call this new matrix A_2.

$$\begin{bmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & * & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & * & * & \cdots & * \end{bmatrix}$$

Step 4: Apply elementary row operations of type (E3) to zero out all entries below this 1 in the first row. Call this new matrix A_3.

$$\left[\begin{array}{cccc|ccc} 0 & \cdots & 0 & 1 & * & \cdots & * \\ \hline 0 & \cdots & 0 & 0 & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & * & \cdots & * \end{array}\right]$$

Step 5: Let A' be the lower right-hand corner A' of A_3 consisting of unknown entries. Plug A' into Step 1 to obtain $\operatorname{ref}(A')$.

$$\left[\begin{array}{cccc|ccc} 0 & \cdots & 0 & 1 & * & \cdots & * \\ \hline 0 & \cdots & 0 & 0 & & & \\ \vdots & & \vdots & \vdots & & A' & \\ 0 & \cdots & 0 & 0 & & & \end{array}\right]$$

Step 6: Replace A' with $\operatorname{ref}(A')$ in A_3, and observe that this matrix is now in row echelon form.

Remark A.3.15 — To obtain a matrix in reduced row echelon form, at the end, one can apply elementary row operations of type (E3) to zero out all entries above a pivot. Although $\operatorname{ref}(A)$ is not unique, the location of the pivots is unique. This can be deduced from the uniqueness of $\operatorname{rref}(A)$, whose pivots are in the same location as the pivots in $\operatorname{ref}(A)$.

Example A.3.16 — Here is an example of applying the Gaussian Elimination Algorithm A.3.14:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - 5R_1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 9 & 10 & 11 & 12 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3 - 9R_1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & -8 & -16 & -24 \end{bmatrix}$$

$$\xrightarrow{R_2 \leftarrow -\frac{1}{4}R_2} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & -8 & -16 & -24 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3 + 8R_2} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

This final matrix is in row echelon form. To obtain a matrix in reduced row echelon form, we perform one final elementary row operation.

$$\xrightarrow{R_1 \leftarrow R_1 - 2R_2} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

A.4 Linear independence and spanning tests

We now give two important applications of the Gaussian Elimination Algorithm A.3.14. We begin with two important lemmas.

Lemma A.4.1 — Suppose A, B are $m \times n$ matrices and U is an invertible $m \times m$ matrix such that $B = UA$. Then $\ker(A) = \ker(B)$.

In particular, if A and B are row equivalent, then $\ker(A) = \ker(B)$.

Proof. If $|\psi\rangle \in \ker(A)$, then

$$B|\psi\rangle = UA|\psi\rangle = U0 = 0,$$

so $|\psi\rangle \in \ker(B)$. Conversely, if $|\psi\rangle \in \ker(B)$, then since $A = U^{-1}B$,

$$A|\psi\rangle = U^{-1}B|\psi\rangle = U0 = 0,$$

so $|\psi\rangle \in \ker(A)$. We deduce $\ker(A) = \ker(B)$.

If A and B are row equivalent, then there are elementary row operations $e_k, \ldots, e_1$ such that B is obtained from A by applying elementary row operations (starting with 1 and then going to k). This is equivalent to multiplying A on the left by the corresponding invertible elementary matrices $E_k, \ldots, E_1$, i.e.,

$$B = E_k \cdots E_1 A.$$

Since each E_i is invertible, we also have $A = E_1^{-1} \cdots E_k^{-1} B$. ∎

Lemma A.4.2 — Suppose the $m \times n$ matrix $A = \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_n\rangle \end{bmatrix}$ is in row echelon form, and suppose there are pivots in columns $1, \ldots, j$.

(1) $|\psi_1\rangle, \ldots, |\psi_j\rangle$ are linearly independent.

(2) If there is no pivot in column $j+1$, then $|\psi_{j+1}\rangle$ is a linear combination of $|\psi_1\rangle, \ldots, |\psi_j\rangle$.

Proof.

(1) This follows immediately from Example A.2.2, as the zeroes at the bottom of A form a strict staircase pattern.

(2) The first $j+1$ columns of A are of the form

$$\left[\begin{array}{ccccc|c} 1 & * & \cdots & * & * & c_1 \\ & \ddots & \ddots & \vdots & \vdots & \vdots \\ & & & 1 & * & c_{j-1} \\ & & & & 1 & c_j \\ \hline 0 & \cdots & \cdots & 0 & 0 & 0 \\ \vdots & & & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & 0 \end{array}\right] \xrightarrow{\text{rref}} \begin{bmatrix} I_j & |d\rangle \\ 0 & 0 \end{bmatrix}.$$

Further passing to reduced row echelon form, we get a block matrix whose upper right corner is I_j, the upper right corner is some vector $|d\rangle \in \mathbb{C}^j$, and whose lower blocks are appropriately sized zero matrices. This matrix clearly contains the vector

$$\begin{bmatrix} d_1 \\ \vdots \\ d_j \\ -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{C}^n$$

in its null space. By Lemma A.4.1, this vector is also in the null space of A, which says that

$$d_1|\psi_1\rangle + \cdots d_j|\psi_j\rangle - |\psi_{j+1}\rangle = 0 \iff |\psi_{j+1}\rangle = d_1|\psi_1\rangle + \cdots d_j|\psi_j\rangle.$$

We have thus demonstrated that $|\psi_{j+1}\rangle$ is in the span of $|\psi_1\rangle, \ldots, |\psi_j\rangle$. ■

Algorithm A.4.3 (Linear Independence Test) — The following algorithm determines whether a set of vector $|\psi_1\rangle, \ldots, |\psi_k\rangle \in \mathbb{C}^n$ are linearly independent.

Step 1: Form the matrix $A := \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_k\rangle \end{bmatrix}$.

Step 2: Row reduce the matrix to obtain $\text{ref}(A)$.

Step 3: $|\psi_1\rangle, \ldots, |\psi_k\rangle$ are linearly independent if and only if $\text{ref}(A)$ has a pivot in every column.

Proof. If $\text{ref}(A)$ has a pivot in every column, then by Lemma A.4.2(1), its columns are linearly independent. By Proposition A.2.7, $\ker(\text{ref}(A)) = \{0\}$, so $\ker(A) = \{0\}$ by Lemma A.4.1. Again by Proposition A.2.7, the columns of A are linearly independent.

Conversely, if a column of $\text{ref}(A)$ has no pivot, then by Lemma A.4.2(2), the columns of $\text{ref}(A)$ are linearly dependent. (Apply Lemma A.4.2(2) to the first column without a pivot.) By Proposition A.2.7, $\ker(\text{ref}(A)) \neq \{0\}$, so $\ker(A) \neq \{0\}$ by Lemma A.4.1. Again by Proposition A.2.7, the columns of A are linearly dependent. ■

The Linear Independence Test A.4.3 proves the following essential lemma.

Corollary A.4.4 — Any collection of $n+1$ vectors in $\mathbb{C}^n$ is linearly dependent. In other words, if $S \subset \mathbb{C}^n$ has more than n elements, then S is linearly dependent.

Proof. Suppose we have $n+1$ vectors $|\psi_1\rangle, \ldots, |\psi_{n+1}\rangle \in \mathbb{C}^n$. Apply the Linear Independence Test A.4.3 to the matrix

$$A := \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_{n+1}\rangle \end{bmatrix}.$$

Since there are only n rows, we get at most n pivots in the row echelon form of A. Thus there is at least one column without a pivot, and thus the columns of A are linearly dependent. ■

The contrapositive of Corollary A.4.4 statement is that if $S \subset \mathbb{C}^n$ is linearly independent, then S has at most n elements.

Additional important facts can be deduced from the proof of the Linear Independence Test A.4.3.

Facts A.4.5 — Suppose $|\psi_1\rangle, \ldots, |\psi_k\rangle \in \mathbb{C}^n$ and form the matrix $A := \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_k\rangle \end{bmatrix}$.

(1) Looking at the columns of $\text{ref}(A)$ with pivots, the corresponding columns of A are linearly independent. This follows by removing all original columns of A corresponding to the columns of $\text{ref}(A)$ without pivots and reapplying the Linear Independence Test A.4.3.

(2) If column $j+1$ of $\text{ref}(A)$ has no pivot, then $|\psi_{j+1}\rangle$ is a linear combination of $|\psi_1\rangle, \ldots, |\psi_j\rangle$. This follows by removing the columns to the left of $|\psi_{j+1}\rangle$ whose columns did not have pivots in $\text{ref}(A)$ and all columns to the right of $|\psi_{j+1}\rangle$ and reapplying the Linear Independence Test A.4.3.

(3) Adding additional columns to the right of $|\psi_k\rangle$ in A does not change the location of the pivots in $\text{ref}(A)$, so the total number of pivots can only increase.

Example A.4.6 — We row reduce the following matrix to find a maximal subset of the columns which are linearly independent.

$$\begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 2 & 2 & -1 & 5 & 4 \\ 3 & 3 & -1 & 8 & 6 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & -1 & -1 & 2 \\ 3 & 3 & -1 & 8 & 6 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3 - 3R_1} \begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & -1 & -1 & 2 \\ 0 & 0 & -1 & -1 & 3 \end{bmatrix}$$

$$\xrightarrow{R_2 \leftarrow -R_2} \begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & -1 & -1 & 3 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3 + R_2} \begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This means the first, second, and third column of our initial matrix are linearly independent: By Corollary A.4.4, we know we have the maximal number of linearly independent columns.

A.5 Computing inverses and the Square Matrix Theorem

In this section, we give an algorithm for computing the inverse of a matrix when it exists. We begin with some important facts about inverses.

Exercise A.5.1 — Suppose A is an $m \times n$ matrix and B is an $n \times m$ matrix such that $BA = I_n$. Show that the columns of A are linearly independent, and the columns of B span $\mathbb{C}^n$. Deduce that if A is an invertible $n \times n$ matrix, then its columns are both linearly independent and they span $\mathbb{C}^n$.
Hint: First, prove that $\ker(A) = \{0\}$ *and apply Proposition A.2.7. Then show that each* $|e_i\rangle$ *lies in the span of the columns of* B *by multiplying* B *on the left of the* i*-th column of* A*.*

Exercise A.5.2 — Suppose the $n \times n$ matrix A has a row of zeroes. Show A is not invertible. What if A has a column of zeroes?

Lemma A.5.3 — An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n.

Proof. Suppose A is row equivalent to I_n, so there are elementary matrices $E_k, \ldots, E_1$ such that $E_k \cdots E_1 A = I_n$. Since each of the E_i's in invertible, we can multiply the inverses in the reverse order to see

$$A = E_1^{-1} \cdots E_k^{-1} E_k \cdots E_1 A = E_1^{-1} \cdots E_k^{-1} I_N = E_1^{-1} \cdots E_k^{-1}.$$

But we have just demonstrated that A is a product of elementary matrices which are invertible, so A is invertible by Exercise A.3.8.

Now suppose A is not row equivalent to I_n, so that $\mathrm{rref}(A)$ must have a row of zeroes. Then there is clearly a row of $\mathrm{rref}(A)$ without a pivot, and since A is square, there is also a column of $\mathrm{rref}(A)$ without a pivot. By the Linear Independence Test A.4.3, the columns of A are not linearly independent. This means that A is not invertible by Exercise A.5.1. ∎

Example A.5.4 — We row reduce the following matrix to determine if it is invertible:

$$A := \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - 4R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3 - 7R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix}$$

$$\xrightarrow{R_2 \leftarrow \frac{-1}{3} R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -6 & -12 \end{bmatrix}$$
$$\xrightarrow{R_3 \leftarrow R_3 + 6R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

We see that this row echelon form of A does not have a pivot in the final column, so A cannot be row equivalent to I_3. We conclude that A is not invertible.

We only apply the following algorithm to square matrices, as being square is necessary for being invertible.

Algorithm A.5.5 (Inverse) — Suppose A is an $n \times n$ matrix.

Step 1: Form the matrix $B := \begin{bmatrix} A & I_n \end{bmatrix}$.

Step 2: Row reduce B to obtain $\operatorname{rref}(B)$.

Step 3: A is invertible if and only if $\operatorname{rref}(B)$ has a pivot in each of the first n columns. In this case,
$$\operatorname{rref}(B) = \begin{bmatrix} I_n & A^{-1} \end{bmatrix}.$$

Proof. Adding I_n as additional columns to A to form B and row reducing will not change the location of the pivots in $\operatorname{rref}(B)$, so
$$\operatorname{rref}(B) = \begin{bmatrix} \operatorname{rref}(A) & B' \end{bmatrix}.$$
for some other $n \times n$ matrix B'. By Lemma A.5.3, A is invertible if and only if $\operatorname{rref}(A) = I_n$, which gives the first statement in Step 3.

Now suppose A is invertible. Then there are elementary matrices $E_k, \ldots, E_1$ such that $E_k \cdots E_1 A = I_n$. This means $A = E_1^{-1} \cdots E_k^{-1}$, so $A^{-1} = E_k \cdots E_1$ by Exercise A.3.8. Finally, we observe
$$E_k \cdots E_1 B := E_k \cdots E_1 \begin{bmatrix} A & I_n \end{bmatrix} = \begin{bmatrix} E_k \cdots E_1 A & E_k \cdots E_1 I_n \end{bmatrix} = \begin{bmatrix} I_n & E_k \cdots E_1 \end{bmatrix} = \operatorname{rref}(B),$$
as claimed. ∎

Example A.5.6 — We apply the Inverse Algorithm A.5.5 to compute the inverse of the following 3×3 matrix on the left hand side of the pipe (if it exists):
$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_3 \leftarrow R_3 - R_1} \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & -2 & 1 & -1 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{R_3 \leftarrow R_3 + 2R_2} \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & -1 & 2 & 1 \end{array}\right]$$

$$\xrightarrow{R_3 \leftarrow -R_3} \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & -1 \end{array}\right] \qquad \text{(ref)}$$

$$\xrightarrow{R_1 \leftarrow R_1 - 2R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & -2 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & -1 \end{array}\right]$$

$$\xrightarrow{R_1 \leftarrow R_1 - 2R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 2 & 2 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & -1 \end{array}\right]$$

$$\xrightarrow{R_2 \leftarrow R_2 + R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 2 & 2 \\ 0 & 1 & 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 1 & -2 & -1 \end{array}\right] \qquad \text{(rref)}$$

Indeed, our initial 3×3 matrix is invertible. We can now verify we have the correct inverse by direct multiplication:

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 & 2 \\ 1 & -1 & -1 \\ 1 & -2 & -1 \end{bmatrix} = I_3,$$

and similarly for multiplication in the other direction.

As an application, we can prove the Spanning Test, which is analogous to the Linear Independence Test A.4.3.

Algorithm A.5.7 (Spanning Test) — The following algorithm determines whether a set of vectors $|\psi_1\rangle, \ldots, |\psi_k\rangle \in \mathbb{C}^n$ span $\mathbb{C}^n$.

Step 1: Form the matrix $A := \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_k\rangle \end{bmatrix}$.

Step 2: Row reduce the matrix to obtain $\text{ref}(A)$.

Step 3: $|\psi_1\rangle, \ldots, |\psi_k\rangle$ span $\mathbb{C}^n$ if and only if $\text{ref}(A)$ has a pivot in every row.

Proof. Suppose $\text{ref}(A)$ has a pivot in every row. Then $n \leq k$, and by reordering and relabelling $|\psi_1\rangle, \ldots, |\psi_k\rangle$, we may assume all the pivots appear in the first n columns of $\text{ref}(A)$. We now claim $|\psi_1\rangle, \ldots, |\psi_n\rangle$ span $\mathbb{C}^n$. Indeed, let $|\eta\rangle \in \mathbb{C}^n$ and row reduce the matrix

$$B := \begin{bmatrix} A & |\eta\rangle \end{bmatrix} = \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_n\rangle & |\eta\rangle \end{bmatrix}.$$

Adding additional columns to the right does not change the location of the pivots of $\text{ref}(B)$, so there is still a pivot in every row of B, these pivots appear in the first n columns of B, and there is no pivot in the final column of $\text{ref}(B)$. By Lemma A.4.2(1), $|\eta\rangle$ is in the span of $|\psi_1\rangle, \ldots, |\psi_n\rangle$ as claimed.

Suppose now $\text{ref}(A)$ has a row without a pivot. Again, we may reorder and relabel $|\psi_1\rangle, \ldots, |\psi_k\rangle$ so that all pivots appear in the first $j < n$ columns of $\text{ref}(A)$, and no pivots occur in the the last $k - j$ columns. By Facts A.4.5(2), the $|\psi_i\rangle$ for $i > j$ are all in the span of $|\psi_1\rangle, \ldots, |\psi_j\rangle$, so deleting them does not affect the span. We now row reduce the matrix

$$C := \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_j\rangle & |e_1\rangle & \cdots & |e_n\rangle \end{bmatrix} = \begin{bmatrix} A' & I_n \end{bmatrix} \qquad \text{where} \qquad A' := \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_j\rangle \end{bmatrix}.$$

We claim that $\text{rref}(C)$ has a pivot in every row. Indeed, since $\text{rref}(C)$ is obtained by left multiplying by elementary matrices $E_k, \ldots, E_1$, we obtain

$$\text{rref}(C) = E_k \cdots E_1 C = \begin{bmatrix} E_k \cdots E_1 A' & E_k \cdots E_1 \end{bmatrix}.$$

Since $E_k \cdots E_1$ is invertible, it cannot have a row of all zeroes by Exercise A.5.2. This means every column of $\text{rref}(C)$ must have a pivot in every row as claimed. By the Linear Independence Test A.4.3, this means there is an $|e_i\rangle$ such that $|\psi_1\rangle, \ldots, |\psi_j\rangle, |e_i\rangle$ is linearly independent, i.e., $|e_i\rangle$ is not in the span of $|\psi_1\rangle, \ldots, |\psi_j\rangle$. ■

The following essential corollary now follows immediately.

Corollary A.5.8 — No set of $n - 1$ vectors can span $\mathbb{C}^n$.

Proof. If we row reduce $A := \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_{n-1}\rangle \end{bmatrix}$, there are at most $n - 1$ pivots, so there will be a row without a pivot. ■

We end this section by collecting many equivalent statements about invertibility.

Theorem A.5.9 ($n \times n$ Square Matrix Theorem) — For an $n \times n$ matrix A, the following are equivalent.

(S1) A is invertible.

(S2) A is row equivalent to I_n.

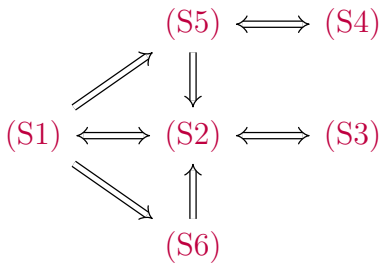
(S3) A is a product of elementary matrices.

(S4) $A|\psi\rangle = 0$ implies that $|\psi\rangle = 0$, i.e., $\ker(A) = \{0\}$.

(S5) The columns of A are linearly independent.

(S6) The columns of A span $\mathbb{C}^n$.

Proof. Below, we prove the following implications.



(S1)⇔(S2): This is exactly Lemma A.5.3.

(S2)⇔(S3): If A is row equivalent to I_n, then there are elementary matrices $E_k, \ldots, E_1$ such that $E_k \cdots E_1 A = I_n$. Then multiplying by the inverses on the left, we have $A = E_1^{-1} \cdots E_k^{-1}$. But since each E_i is elementary, E_i^{-1} is also elementary of the same type, so A is a product of elementary matrices.

Reading this proof in reverse gives the converse implication. Indeed, if $A = E_1 \cdots E_k$ is a product of elementary matrices, then $E_k^{-1} \cdots E_1^{-1} A = I_n$. Since each E_i^{-1} is elementary as well, we have demonstrated that A is row equivalent to I_n.

(S4)⇔(S5): This follows verbatim from Proposition A.2.7. (Note that these implications hold even when A is not square.)

(S1)⇒(S5): Since A is invertible, $A^{-1}A = I_n$. By Exercise A.5.1, the columns of A are linearly independent.

(S1)⇒(S6): Since A is invertible, $AA^{-1} = I_n$. By Exercise A.5.1, the columns of A span $\mathbb{C}^n$.

(S5)⇒(S2): If the columns of A are linearly independent, then $\text{rref}(A)$ has a pivot in every column by the Linear Independence Test A.4.3. Thus $\text{rref}(A) = I_n$.

(S6)⇒(S2): If the columns of A span $\mathbb{C}^n$, then $\text{rref}(A)$ has a pivot in every row by the Spanning Test A.5.7. Thus $\text{rref}(A) = I_n$. ∎

Remark A.5.10 — By Exercise A.3.9, A^T is invertible if and only if A^T is invertible, in which case $(A^T)^{-1} = (A^{-1})^T$. Since the columns of A^T are exactly the rows of A, we can add the following two equivalent conditions to Theorem A.5.9.

(S7) The rows of A are linearly independent.

(S8) The rows of A span $\mathbb{C}^n$.

A.6 Bases and the Extension and Contraction Theorems

For this section, we suppose that $V \subseteq \mathbb{C}^n$ is a subspace.

Definition A.6.1 — A *basis* for V is a subset $B \subset V$ which both is linearly independent and spans V.

The existence of bases will follow from the Extension Theorem A.7.4 below.
By combining Corollaries A.4.4 and A.5.8, we get the following essential theorem.

Theorem A.6.2 — Every basis of $\mathbb{C}^n$ has exactly n elements.

Proof. Suppose B is a basis of $\mathbb{C}^n$. By Corollary A.4.4, B has at most n elements. But by Corollary A.5.8, B must have at least n elements. We conclude that B has exactly n elements. ∎

Exercise A.6.3 — Let $S = \{|\psi_1\rangle, \ldots, |\psi_k\rangle\} \subset V$ be a spanning set for V. This means that every vector in V can be written as a linear combination of the elements of S. Show that the following are equivalent.

- S is also linearly independent, so that S is a basis.
- Every vector in V can be *uniquely* expressed as a linear combination of the elements of S.
- The zero vector $0 \in V$ can be *uniquely* expressed as a linear combination of the elements of S.

One should think of linearly independent sets as *small* and spanning sets as *big*, and the sets exactly in the middle are bases.

Exercise A.6.4 — Let $S_1 \subset S_2 \subset V$.

(1) If S_2 is linearly independent, then so is S_1.

(2) If S_1 is linearly dependent, then so is S_2.

(3) If S_1 spans V, then S_2 spans V.

(4) If S_2 does not span V, then neither does S_1.

Algorithm A.6.5 (Column Space Basis) — Given an $m \times n$ matrix, the following algorithm produces a basis for the column space.

Step 1: Row reduce A to obtain $\mathrm{ref}(A)$.

Step 2: The columns of A which have a pivot in $\mathrm{ref}(A)$ form a basis for the column space.

Proof. By the Linear Independence Test A.4.3, the chosen columns of A are linearly independent. By Facts A.4.5(2), all other columns are in the span of the chosen columns, so the chosen columns also span the column space. Thus the chosen columns are a basis for the column space. ∎

Example A.6.6 — We compute a basis for the column space of the following matrix using the Column Space Basis Algorithm A.6.5:

$$\begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 2 & 2 & -1 & 5 & 4 \\ 3 & 3 & -1 & 8 & 5 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & -1 & -1 & 2 \\ 3 & 3 & -1 & 8 & 5 \end{bmatrix}$$
$$\xrightarrow{R_3 \leftarrow R_3 - 3R_1} \begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & -1 & -1 & 2 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}$$
$$\xrightarrow{R_2 \leftarrow -R_2} \begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}$$
$$\xrightarrow{R_3 \leftarrow R_3 + R_2} \begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Since there are pivots in the first and third columns, a basis for the column space is given by

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} \right\}.$$

Remark A.6.7 — Suppose $V \subseteq \mathbb{C}^n$ is a subspace, and suppose $S = \{|\psi_1\rangle, \ldots, |\psi_k\rangle\} \subset V$ is a spanning set. There is a subset $B \subset S$ such that B is a basis of V. Indeed, this follows by applying the Column Space Basis Algorithm A.6.5 to

$$A = \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_k\rangle \end{bmatrix}.$$

Algorithm A.6.8 (Null Space Basis) **—** Given an $m \times n$ matrix, the following algorithm produces a basis for the null space.

Step 1: Row reduce A to obtain $\mathrm{rref}(A)$.

Step 2: Look at the equation

$$\operatorname{rref}(A) \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{A.6.9}$$

which gives a system of m linear equations.

Step 3: If column i of $\operatorname{rref}(A)$ has a pivot, solve for x_i in terms of $x_{i+1}, \ldots, x_m$.

Step 4: The x_i's corresponding to the columns without pivots are *free variables*. Parametrize the solutions $|x\rangle$ to (A.6.9) in terms of the free variables to get a basis for the null space.

Proof. This algorithm gives a direct parametrization of all vectors $|\psi\rangle$ for which $\operatorname{rref}(A)|\psi\rangle = 0$. Since $A|\psi\rangle = 0$ if and only if $\operatorname{rref}(A)|\psi\rangle = 0$ by Lemma A.4.1, the result is a direct parametrization of all vectors $|\psi\rangle$ for which $A|\psi\rangle = 0$. This means the obtained set of vectors will span the null space. To see linear independence, observe that the vector parametrized by the free variable x_i will automatically have a 1 in the i-th entry and zeroes below the i-th entry. This is directly because the i-th column is a linear combination of the previous columns by Facts A.4.5(2). Thus the vectors obtained by this algorithm automatically have a strict staircase pattern of zeroes at the bottom as in Example A.2.2, and are thus linearly independent. ∎

Example A.6.10 — We apply the Null Space Basis Algorithm A.6.8 to the same matrix from Example A.6.6:

$$\begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 2 & 2 & -1 & 5 & 4 \\ 3 & 3 & -1 & 8 & 5 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We get the following system of linear equations:

$$\begin{bmatrix} 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 + 3x_4 + x_5 \\ x_3 + x_4 - 2x_5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

We solve for x_1 in terms of x_2, x_4, x_5 and x_3 in terms of x_4, x_5:

$$x_1 = -x_2 - 3x_4 - x_5 \qquad\qquad x_3 = -x_4 - x_5.$$

Thus x_2, x_4, x_5 are our free variables. We now parametrize the solution space in terms

of the free variables:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -x_2 - 3x_4 - x_5 \\ x_2 \\ -x_4 - x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}.$$

We conclude that a basis for the null space is given by

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Algorithm A.6.11 — The following algorithm extends a linearly independent set of vectors $|\psi_1\rangle, \ldots, |\psi_k\rangle$ of $\mathbb{C}^n$ to a basis.

Step 1: Form the matrix $A := \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_k\rangle & I_n \end{bmatrix}$.

Step 2: Row reduce A to obtain $\text{ref}(A)$.

Step 3: The columns of A which have pivots in $\text{rref}(A)$ give a basis for $\mathbb{C}^n$ containing $|\psi_1\rangle, \ldots, |\psi_k\rangle$.

Proof. Since $|\psi_1\rangle, \ldots, |\psi_k\rangle$ are linearly independent and in the first k columns of A, there will be a pivot in the first k columns by the Linear Independence Test A.4.3. But since the columns of A clearly span $\mathbb{C}^n$, $\text{ref}(A)$ will have a pivot in every row, of which there are n.

Now take $|\psi_1\rangle, \ldots, |\psi_k\rangle$ together with the columns of I_n which have a pivot in $\text{ref}(A)$. That is, we take the n columns of A which have pivots in $\text{ref}(A)$ to form a new matrix

$$B = \begin{bmatrix} |\psi_1\rangle & \cdots & |\psi_k\rangle & |e_{i_1}\rangle & \cdots & |e_{i_{n-k}}\rangle \end{bmatrix}.$$

We claim these n vectors form a basis for $\mathbb{C}^n$. Indeed, if we row reduce B, we see $\text{ref}(B)$ has a pivot in every column and row. Hence these n vectors are linearly independent by the Linear Independence Test A.4.3, and they span $\mathbb{C}^n$ by the Spanning Test A.5.7. ∎

Example A.6.12 — We extend

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

to a basis for $\mathbb{C}^4$. We row reduce

$$\left[\begin{array}{cc|cccc} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{cc|cccc} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \end{array}\right]$$

$$\longrightarrow \left[\begin{array}{cc|cccc} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{array}\right]$$

$$\longrightarrow \left[\begin{array}{cc|cccc} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{array}\right]$$

$$\longrightarrow \left[\begin{array}{cc|cccc} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{array}\right]$$

$$\longrightarrow \left[\begin{array}{cc|cccc} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right]$$

Since there are pivots in the columns corresponding to $|e_2\rangle$ and $|e_3\rangle$, we conclude that

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

is a basis for $\mathbb{C}^4$.

A.7 The Extension and Contraction Theorems

The Extension and Contraction Theorems play spanning sets off linearly independent sets for strong results in dual ways, and the structures of the proofs are very similar.

Lemma A.7.1 (Extension Lemma) — Suppose $S = \{|\psi_1\rangle, \ldots, |\psi_k\rangle\}$ is a linearly independent subset of V, and suppose $|\eta\rangle \notin \text{span}(S)$. Then $S \cup |\eta\rangle$ is linearly independent.

Proof. Since S is linearly independent, we know that

$$\sum_{i=1}^{k} c_i|\psi_i\rangle = 0 \qquad \Longrightarrow \qquad c_i = 0 \text{ for all } i. \tag{A.7.2}$$

Now suppose

$$d|\eta\rangle + \sum_{i=1}^{k} c_i|\psi_i\rangle = 0. \tag{A.7.3}$$

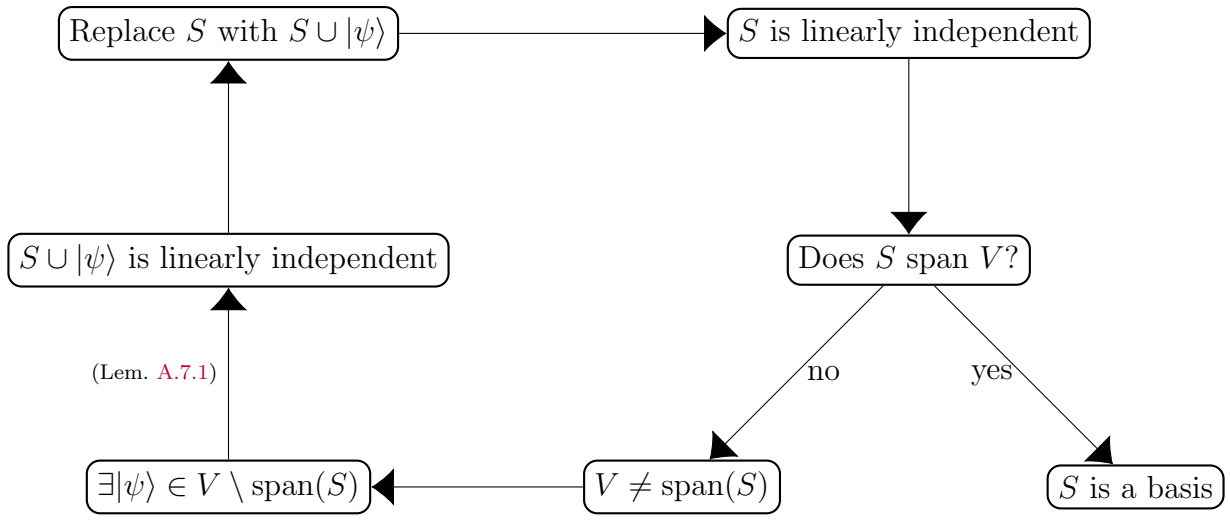
If $d \neq 0$, then

$$|\eta\rangle = \frac{-1}{d}\sum_{i=1}^{k} c_i|\psi_i\rangle = \sum_{i=1}^{k} \frac{-c_i}{d}|\psi_i\rangle \in \text{span}(S),$$

a contradiction. Hence $d = 0$, and plugging this in to (A.7.3), we can conclude $c_i = 0$ for all i by (A.7.2). Hence all the scalars in equation (A.7.3) must be zero, so $S \cup |\eta\rangle$ is linearly independent. ∎

Theorem A.7.4 (Extension) — Let $S \subset V$ be a linearly independent subset. Then there is a finite set $B \supset S$ such that B is a basis for V.

Proof. The proof is the following algorithm, represented as a flow chart:



Since S is linearly independent, if S spans V, then S is a basis for V. If S does not span V, then there is a $|\psi\rangle \in V \setminus \text{span}(S)$, and $S \cup |\psi\rangle$ is linearly independent by Lemma A.7.1. Thus we may replace S with $S \cup |\psi\rangle$ and repeat the above procedure with a new linearly independent set with strictly more elements. This algorithm must eventually terminate, as we cannot have more than $n + 1$ linearly independent vectors in $\mathbb{C}^n$ by Corollary A.4.4. ∎

Corollary A.7.5 (Existence of Bases) — V has a basis.

Proof. Use the Extension Theorem A.7.4 on the empty set, which is linearly independent by definition. ∎

Corollary A.7.6 — Let W be a subspace of V, and suppose S is a basis of W. There is a basis B for V with $S \subset B$.

Proof. Use the Extension Theorem A.7.4 on S viewed as a subset of V to obtain B. ∎

The rest of this section is redundant if we only consider subspaces of $\mathbb{C}^n$, as one can just use the Column Space Basis Algorithm A.6.5 to contract a spanning set to a basis. Please feel free to skip this part in a first reading of these notes. However, the proofs below apply to a subspace of *any* finite dimensional vector space. We include them here for symmetry with Extension Theorem A.7.4 results above.

Lemma A.7.7 (Contraction Lemma) — Suppose $S = \{|\psi_1\rangle, \ldots, |\psi_k\rangle\}$ spans V. If $|\psi_j\rangle \in \text{span}(S \setminus |\psi_j\rangle)$, then $S \setminus |\psi_j\rangle$ spans V.

Proof. Since $|\psi_j\rangle \in \text{span}(S \setminus |\psi_j\rangle)$, $|\psi_j\rangle$ is a linear combination of the $|\psi_i\rangle$'s where the coefficient of $|\psi_j\rangle$ is zero:

$$|\psi_j\rangle = \sum_{i \neq j} c_i |\psi_i\rangle.$$

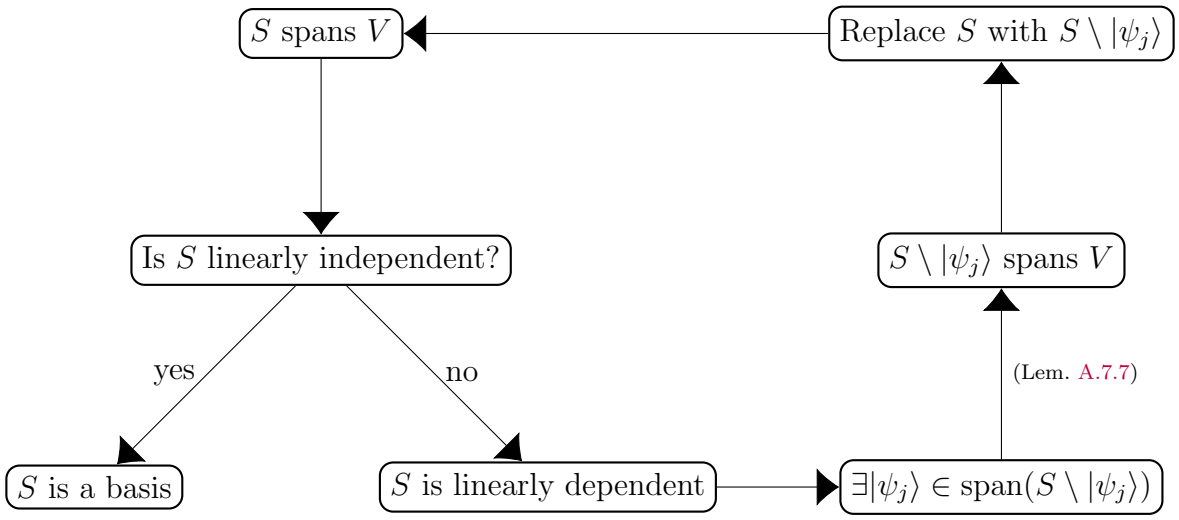
Now if $|\eta\rangle \in V$, then there are $d_1, \ldots, d_n \in \mathbb{C}$ such that

$$\begin{aligned} |\eta\rangle &= \sum_{i=1}^{n} d_i |\psi_i\rangle = d_j |\psi_j\rangle + \sum_{i \neq j} d_i |\psi_i\rangle = d_j \sum_{i \neq j} c_i |\psi_i\rangle + \sum_{i \neq j} d_i |\psi_i\rangle \\ &= \sum_{i \neq j} (d_j c_i + d_i) |\psi_i\rangle \in \text{span}(S \setminus |\psi_j\rangle). \end{aligned}$$

Hence $S \setminus |\psi_j\rangle$ spans V. ∎

Theorem A.7.8 (Contraction) — Let $S = \{|\psi_1\rangle, \ldots, |\psi_k\rangle\} \subset V$ be a spanning set. There is a subset $B \subset S$ such that B is a basis of V.

Proof. The proof is the following algorithm, represented as a flow chart:



Since S spans V, if S is also linearly independent, then S is a basis. If S is not linearly independent, then S is linearly dependent. Then there is a $|\psi_j\rangle \in S$ such that $|\psi_j\rangle$ is a linear combination of the other elements of S. By Lemma A.7.7, the set $S \setminus |\psi_j\rangle$ still spans V, so we may replace S with $S \setminus |\psi_j\rangle$ and repeat the above procedure with a new finite set with strictly fewer elements. This algorithm must eventually terminate, since the empty set is linearly independent by definition. ∎

A.8 Dimension

We now extend Theorem A.6.2 to arbitrary subspaces to show that every basis of $V \subseteq \mathbb{C}^n$ always has the same number of elements.

Theorem A.8.1 — Let $S, T \subset V$. If S is linearly independent and T spans V, then $|S| \leq |T|$.

Proof. Let $B = \{|\psi_1\rangle, \ldots, |\psi_k\rangle\}$ be a basis for V, which exists by Corollary A.7.5. There is a linear map $\mathbb{C}^k \to V$ given by

$$\begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix} \longmapsto a_1|\psi_1\rangle + a_2|\psi_2\rangle + \cdots + a_k|\psi_k\rangle$$

Since B spans V, this map is onto, as every $|\psi\rangle \in V$ is a linear combination of $|\psi_1\rangle, \ldots, |\psi_k\rangle$. Since B is linearly independent, this map is one-to-one, as every $|\psi\rangle$ can be written uniquely as a linear combination of $|\psi_1\rangle, \ldots, |\psi_k\rangle$. Since this map is both one-to-one and onto, it is

invertible. Its inverse is called the *coordinate map* $[\,\cdot\,]_B : V \to \mathbb{C}^k$, and it is given by

$$a_1|\psi_1\rangle + a_2|\psi_2\rangle + \cdots + a_k|\psi_k\rangle \longmapsto \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}.$$

We now consider the *images* of S and T in $\mathbb{C}^k$:

$$[S]_B = \{[|\psi\rangle]_B | |\psi\rangle \in S\} \qquad \text{and} \qquad [T]_B = \{[|\psi\rangle]_B | |\psi\rangle \in T\}.$$

Since S is linearly independent and $[\,\cdot\,]_B$ is invertible, $[S]_B$ is linearly independent. By Corollary A.4.4, $|S| = |[S]_B| \leq k$. Since T spans V and $[\,\cdot\,]_B$ is invertible, $[T]_B$ spans $\mathbb{C}^k$. By Corollary A.5.8, $|T| = |[T]_B| \geq k$. Hence $|S| \leq k \leq |T|$. ■

Corollary A.8.2 — Every basis for V has the same number of elements.

Proof. Suppose B_1, B_2 are two bases for V. Since B_1 is linearly independent and B_2 spans V, by Theorem A.8.1, $|B_1| \leq |B_2|$. But since B_2 is linearly independent and B_1 spans V, by Theorem A.8.1, $|B_2| \leq |B_1|$. We conclude that $|B_1| = |B_2|$. ■

Exercise A.8.3 — Suppose $S \subset V$ has $\dim(V)$ elements. Prove that the following are equivalent.

(1) S is linearly independent,

(2) S is a basis for V, and

(3) S spans V.

That is, maximal linear independent sets are bases are minimal spanning sets.

Definition A.8.4 — We define the *dimension* of V, denoted $\dim(V)$, to be the number of elements in a basis for V, which is well-defined by Corollaries A.7.5 and A.8.2.

Definition A.8.5 — Let A be an $m \times n$ matrix. The *rank* of A is the dimension of $\mathrm{im}(A)$. The *nullity* of A is the dimension of $\ker(A)$.

Theorem A.8.6 (Rank-Nullity) — Let A be an $m \times n$ matrix. The rank of A plus the nullity of A is equal to n.

Proof. By the Column Space Basis Algorithm A.6.5, the rank of A is the number of columns in $\text{rref}(A)$ with a pivot. By the Null Space Basis Algorithm A.6.8, the nullity of A is the number of columns of $\text{rref}(A)$ without a pivot. The sum of these is the number of columns of A. ∎

A.9 Orthonormality

In this final section, we explain how to obtain an orthonormal set from a linearly independent set which preserves the spans of initial chains of vectors.

Algorithm A.9.1 (Gram-Schmidt Orthonormalization) — The following inductive algorithm takes a linearly independent subset $|\psi_1\rangle, \ldots, |\psi_k\rangle \in \mathbb{C}^n$ and produces an orthogonal set $|\eta_1\rangle, \ldots, |\eta_k\rangle \in \mathbb{C}^n$ such that for every $1 \leq j \leq k$,

$$\text{span}\{|\psi_1\rangle, \ldots, |\psi_j\rangle\} = \text{span}\{|\eta_1\rangle, \ldots, |\eta_j\rangle\}.$$

Further normalizing each $|\eta_i\rangle$ to have norm 1 gives an orthonormal set.

Step 1: Set $|\eta_1\rangle := |\psi_1\rangle$

Step 2: Suppose we have already produced $|\eta_1\rangle, \ldots, |\eta_j\rangle$. Set

$$|\eta_{j+1}\rangle := |\psi_{j+1}\rangle - \frac{\langle\eta_1|\psi_{j+1}\rangle}{\langle\eta_1|\eta_1\rangle}|\eta_1\rangle - \frac{\langle\eta_2|\psi_{j+1}\rangle}{\langle\eta_2|\eta_2\rangle}|\eta_2\rangle - \cdots - \frac{\langle\eta_j|\psi_{j+1}\rangle}{\langle\eta_j|\eta_j\rangle}|\eta_j\rangle.$$

Step 3: When $j + 1 = n$, we are finished.

Proof. Observe that at each step, $|\eta_{j+1}\rangle$ is non-zero, as

$$|\psi_{j+1}\rangle \notin \text{span}\{|\psi_1\rangle, \ldots, |\psi_j\rangle\} = \text{span}\{|\eta_1\rangle, \ldots, |\eta_j\rangle\}$$

and orthogonal to $|\eta_1\rangle, \cdots, |\eta_j\rangle$ as for all $1 \leq i \leq j$,

$$\begin{aligned}
\langle\eta_i|\eta_{j+1}\rangle &= \langle\eta_i|\psi_{j+1}\rangle - \frac{\langle\eta_1|\psi_{j+1}\rangle}{\langle\eta_1|\eta_1\rangle}\langle\eta_i|\eta_1\rangle - \frac{\langle\eta_2|\psi_{j+1}\rangle}{\langle\eta_2|\eta_2\rangle}\langle\eta_i|\eta_2\rangle - \cdots - \frac{\langle\eta_j|\psi_{j+1}\rangle}{\langle\eta_j|\eta_j\rangle}\langle\eta_i|\eta_j\rangle \\
&= \langle\eta_i|\psi_{j+1}\rangle - \frac{\langle\eta_i|\psi_{j+1}\rangle}{\langle\eta_i|\eta_i\rangle}\langle\eta_i|\eta_i\rangle \\
&= \langle\eta_i|\psi_{j+1}\rangle - \langle\eta_i|\psi_{j+1}\rangle = 0.
\end{aligned}$$

Moreover, we have both $|\psi_{j+1}\rangle \in \text{span}\{|\eta_1\rangle, \ldots, |\eta_j\rangle, |\eta_{j+1}\rangle\}$ and

$$|\eta_{j+1}\rangle \in \text{span}\{|\eta_1\rangle, \ldots, |\eta_j\rangle, |\psi_{j+1}\rangle\} = \text{span}\{|\psi_1\rangle, \ldots, |\psi_j\rangle, |\psi_{j+1}\rangle\},$$

so at each step the $|\psi_i\rangle$ and the $|\eta_i\rangle$ have the same span. ■

Example A.9.2 — We apply the Gram-Schmidt Orthonormalization Algorithm A.9.1 to obtain an orthonormal set with the same span as the linearly independent set

$$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

First, we set

$$|\eta_1\rangle := \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Second, we have

$$|\eta_2\rangle := \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{\begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Third, we set

$$|\eta_2\rangle := \begin{bmatrix}0\\1\\1\\0\end{bmatrix} - \frac{\begin{bmatrix}1&1&0&0\end{bmatrix}\begin{bmatrix}0\\1\\1\\0\end{bmatrix}}{\begin{bmatrix}1&1&0&0\end{bmatrix}\begin{bmatrix}1\\1\\0\\0\end{bmatrix}}\begin{bmatrix}1\\1\\0\\0\end{bmatrix} - \frac{\begin{bmatrix}0&0&1&1\end{bmatrix}\begin{bmatrix}0\\1\\1\\0\end{bmatrix}}{\begin{bmatrix}0&0&1&1\end{bmatrix}\begin{bmatrix}0\\0\\1\\1\end{bmatrix}}\begin{bmatrix}0\\0\\1\\1\end{bmatrix}$$

$$= \begin{bmatrix}0\\1\\1\\0\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\1\\0\\0\end{bmatrix} - \frac{1}{2}\begin{bmatrix}0\\0\\1\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}-1\\1\\1\\-1\end{bmatrix}.$$

One can directly verify that $|\eta_1\rangle, |\eta_2\rangle, |\eta_3\rangle$ are orthogonal. Finally, to get an ONB, we must normalize:

$$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}0\\0\\1\\1\end{bmatrix}, \frac{1}{2}\begin{bmatrix}-1\\1\\1\\-1\end{bmatrix}.$$

Corollary A.9.3 — Every subspace $V \subseteq \mathbb{C}^N$ has an ONB.

Proof. Use Corollary A.7.5 to obtain a basis for V. We may then feed this basis into the Gram-Schmidt Algorithm A.9.1 to obtain an ONB. ∎

Corollary A.9.4 — Let W be a subspace of V, and suppose S is an ONB of W. There is an ONB B for V with $S \subset B$.

Proof. Use the Extension Theorem A.7.4 on S viewed as a subset of V to obtain B. If S is an ONB, we can put the elements in S before those in B, and then run Gram-Schmidt Algorithm A.9.1 to obtain an ONB. Since S was before B, S is left unchanged by the algorithm. ∎

References

[BB84] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, 10-12 December*, pages 175–179, 1984.

[BBC+93] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, Mar 1993. DOI:10.1103/PhysRevLett.70.1895.

[Bel64] J. S. Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1:195–200, Nov 1964. DOI:10.1103/PhysicsPhysiqueFizika.1.195.

[BW92] Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on einstein-podolsky-rosen states. *Phys. Rev. Lett.*, 69:2881–2884, Nov 1992. DOI:10.1103/PhysRevLett.69.2881.

[CG23] Bob Coecke and Stefano Gogioso. *Quantum in pictures*. Quantinuum Ltd., 2023.

[CK17] Bob Coecke and Aleks Kissinger. *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge University Press, 2017. DOI:10.1017/9781316219317.

[EK66] Samuel Eilenberg and G. M. Kelly. A generalization of the functorial calculus. *J. Algebra*, 3:366–375, 1966. MR190204 DOI:10.1016/0021-8693(66)90006-8.

[Eke91] Artur K. Ekert. Quantum cryptography based on bell's theorem. *Phys. Rev. Lett.*, 67:661–663, Aug 1991. DOI:10.1103/PhysRevLett.67.661.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, May 1935. DOI:10.1103/PhysRev.47.777.

[FC72] Stuart J. Freedman and John F. Clauser. Experimental test of local hidden-variable theories. *Phys. Rev. Lett.*, 28:938–941, Apr 1972. DOI:10.1103/PhysRevLett.28.938.

[GHSZ90] Daniel M. Greenberger, Michael A. Horne, Abner Shimony, and Anton Zeilinger. Bell's theorem without inequalities. *American Journal of Physics*, 58(12):1131–1143, 12 1990. DOI:10.1119/1.16243.

[GS22] Walther Gerlach and Otto Stern. Das magnetische moment des silberatoms. *Zeitschrift für Physik*, 9(1):353–355, 1922. DOI:10.1007/BF01326984.

[HV19] Chris Heunen and Jamie Vicary. *Categories for quantum theory*, volume 28 of *Oxford Graduate Texts in Mathematics*. Oxford University Press, Oxford, 2019. An introduction, MR3971584 DOI:10.1093/oso/9780198739623.001.0001.

[Jon85] Vaughan F. R. Jones. A polynomial invariant for knots via von Neumann algebras. *Bulletin of the American Mathematical Society*, 12(1):103–111, 1985. MR0766964.

[JS91] André Joyal and Ross Street. The geometry of tensor calculus. I. *Adv. Math.*, 88(1):55–112, 1991. MR1113284 DOI:10.1016/0001-8708(91)90003-P.

[Kau87] Louis H. Kauffman. State models and the Jones polynomial. *Topology*, 26(3):395–407, 1987. MR899057, DOI:10.1016/0040-9383(87)90009-7.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, 2000. MR1796805.

[Pen71] Roger Penrose. Applications of negative dimensional tensors. In *Combinatorial Mathematics and its Applications (Proc. Conf., Oxford, 1969)*, pages 221–244. Academic Press, London, 1971.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948. MR26286 DOI:10.1002/j.1538-7305.1948.tb01338.x.

[TL71] Harold N. V. Temperley and Elliott H. Lieb. Relations between the "percolation" and "colouring" problem and other graph-theoretical problems associated with regular planar lattices: some exact results for the "percolation" problem. *Proc. Roy. Soc. London Ser. A*, 322(1549):251–280, 1971. MR0498284.

[ZCZW19] Bei Zeng, Xie Chen, Duan-Lu Zhou, and Xiao-Gang Wen. *Quantum information meets quantum matter*. Quantum Science and Technology. Springer, New York, 2019. From quantum entanglement to topological phases of many-body systems, With a foreword by John Preskill, MR3929747 DOI:10.1007/978-1-4939-9084-9 arXiv:1508.02595.