Indian Institute of Technology, Bombay

Quantum Information

Report

Siddhant Midha

Summer 2022

Last updated on July 7, 2022

Contents

Note to the Reader

This report is intended to formalize concretely the concepts of quantum information. The main references are

- The Theory of Quantum Information by John Watrous.
- Quantum Computation and Quantum Information by M. Nielson and I. Chuang.
- From Classical to Quantum Shannon Theory by Mark Wilde.

Due to a stubborn affection for the dirac notation, I have used the same whenever denoting vectors in a space. For giving additionaly physics intuition, I have also added boxes titled *Physics Fact*, which attempt to draw the reader's attention to connect the dots.

§1. Mathematical Prerequisites

§§1.1. Linear Algebra

§§§1.1.1. Spaces and Composition of Spaces

The book defines complex euclidean spaces in a way unfamiliar to a standard user of linear algebra. So we shall make some effort highlighting the differences. An alphabet (usually denoted by capital greek letters Σ, Γ etc.) is a set of symbols (usually denoted by small letters a, b, c etc.).

Definition 1.1.1 (Complex Euclidean Space)**.** For any alphabet Σ, we define the space associated with it, $\mathbb{C}^\Sigma$ as follows,

$$\mathbb{C}^\Sigma := \{f | f : \Sigma \to \mathbb{C} \text{ is a function}\}$$

A vector space defined in this way will be called a complex euclidean space.

For some $u \in \mathbb{C}^\Sigma$, the 'indexing' is done as $u(a)$ for some $a \in \Sigma$. Addition and scalar multiplication are defined as usual. Such spaces will be denoted by scripted capital letters such as $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ etc. Vectors in these spaces will be denoted as x, y, z, u etc. The standard basis of $\mathbb{C}^\Sigma$ is the orthonormal basis given by $\{e_a | a \in \Sigma\}$ where $e_a(b) = \delta_{ab} \forall a, b \in \Sigma$.

In QI, we often deal with multiple systems (or *registers*[1]), and a need to combine vector spaces arises. Thus we make some definitiions.

Definition 1.1.2 (Direct Sums)**.** Let $\mathcal{X}_i = \mathbb{C}^{\Sigma_i}$ for $i = 1, 2 \dots n$. First define,

$$\Sigma_1 \sqcup \dots \Sigma_n := \bigcup_{k=1,2\dots n} \{(k, a) | a \in \Sigma_k\}$$

This is called the *disjoint union.* Now, wefine the space $\mathcal{X}_1 \oplus \mathcal{X}_2 \dots \mathcal{X}_n$ as

$$\mathcal{X}_1 \oplus \mathcal{X}_2 \dots \mathcal{X}_n := \mathbb{C}^{\Sigma_1 \sqcup \dots \Sigma_n}$$

Intuitively, a vector in the direct sum of two spaces is formed by stacking vectors from the two spaces. Hence, $u_1 \oplus \dots u_n$ may be viewed as

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

The following identities hold,

- $(u_1 \oplus \dots u_n) + (v_1 \oplus \dots v_n) := (u_1 + v_1) \oplus \dots (u_n + v_n)$.
- $\alpha(u_1 \oplus \dots u_n) := (\alpha u_1) \oplus \dots (\alpha u_n)$.
- $\langle u_1 \oplus \dots u_n, v_1 \oplus \dots v_n \rangle := \langle u_1, v_1 \rangle + \dots \langle u_n, v_n \rangle$.

Now, we describe an even more important way of combining vector spaces, called the tensor product.

Definition 1.1.3 (Direct Sums)**.** Let $\mathcal{X}_i = \mathbb{C}^{\Sigma_i}$ for $i = 1, 2 \dots n$. Now, we define the space $\mathcal{X}_1 \otimes \mathcal{X}_2 \dots \otimes \mathcal{X}_n$ as

$$\mathcal{X}_1 \otimes \mathcal{X}_2 \dots \otimes \mathcal{X}_n := \mathbb{C}^{\Sigma_1 \times \dots \times \Sigma_n}$$

where × denotes the cartesian product.

[1] as we shall later see

Intuitively, a tensor product of two vectors is formed by multiplying each element of the first vector by the second. As an example, take $u = (a, b)^T$ and $v = (c, d)^T$. We have,

$$u \otimes v = \begin{pmatrix} av \\ bv \end{pmatrix} = \begin{pmatrix} ac \\ ad \\ bc \\ bd \end{pmatrix}$$

It is left to the reader to correlate this with the definition.

The following indentities hold,

- $(u_1 \otimes \cdots \otimes u_n)(a_1, a_2, \ldots, a_n) := u_1(a_1)u_2(a_2)\ldots u_n(a_n)$
- $u_1 \otimes \cdots \otimes u_{k-1} \otimes (au_k + bv_k) \otimes u_{k+1} \otimes \cdots \otimes u_n := a(u_1 \otimes \cdots \otimes u_{k-1} \otimes u_k \otimes u_{k+1} \otimes \cdots \otimes u_n) + b(u_1 \otimes \cdots \otimes u_{k-1} \otimes v_k \otimes u_{k+1} \otimes \cdots \otimes u_n)$
- $\langle u_1 \otimes \cdots \otimes u_n, v_1 \otimes \cdots \otimes v_n \rangle := \prod_{i=1}^{n} \langle u_i, v_i \rangle$

Vectors of the form $u_1 \otimes \cdots \otimes u_n$ are called *elementary tensors*. These span the space, but not every vector in is an elementary tensor[2].

§§§1.1.2. Linear Operators

- Given spaces $\mathcal{X}$ and $\mathcal{Y}$, we define

  $$L(\mathcal{X}, \mathcal{Y}) := \{A \mid A : \mathcal{X} \to \mathcal{Y} \text{ is a linear operators}\}$$

  It is easy to see that $L(\mathcal{X}, \mathcal{Y})$ forms a vector space with commonly defined addition and scalar multiplication. Additionally, we denote $L(\mathcal{X}, \mathcal{X})$ as $L(\mathcal{X})$.

- We can define an inner product on $L(\mathcal{X}, \mathcal{Y})$ as

  $$\langle A, B \rangle := Tr[A^* B]$$

- Matrices? We define a matrix over $\mathbb{C}$ as a mapping of the form $M : \Gamma \times \Sigma \to \mathbb{C}$ for some alphabets Σ and Γ. For $a \in \Gamma$ and $b \in \Sigma$, we call the value $M(a, b)$ the $(a, b)^{th}$ entry of M, where a is the row index and b is the column index. Addition and scalar or matrix multiplication are defined as usual.

- For $\mathcal{X} = \mathbb{C}^\Sigma$ and $\mathcal{Y} = \mathbb{C}^\Gamma$ there is a bijective map between $L(\mathcal{X}, \mathcal{Y})$ and the set of matrices $M : \Gamma \times \Sigma \to \mathbb{C}$. This is given by

  1. $M_A(a, b) = \langle e_a, Ae_b \rangle$ for $e_a \in \mathcal{Y}$ and $e_b \in \mathcal{X}$.
  2. $(A_M u)(a) = \sum_{b \in \Sigma} M(a, b)u(b)$ for all $a \in \Gamma$.

  Owing to this, we simply denote $A(a, b) = \langle e_a, Ae_b \rangle$ with some abuse of notation.

- For the space $L(\mathcal{X} = \mathbb{C}^\Sigma, \mathcal{Y} = \mathbb{C}^\Gamma)$, we have the standard basis given by E_{ab} for all $a \in \Gamma$ and $b \in \Sigma$. Where $E_{ab}(c, d) := \delta_{ac}\delta_{bd}$.

- Tensor product of operators. We shall describe these in a bit detail. Suppose $\mathcal{X}_i = \mathbb{C}^{\Sigma_i}$ and $\mathcal{Y}_i = \mathbb{C}^{\Gamma_i}$ for $i = 1, 2 \ldots n$. For $A_i \in L(\mathcal{X}_i, y_i)$, we have the operator

  $$A_1 \otimes \cdots \otimes A_n \in L(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n)$$

  defined as

  $$(A_1 \otimes \cdots \otimes A_n)(u_1 \otimes \cdots \otimes u_n) = (A_1 u_1) \otimes \cdots \otimes (A_n u_n)$$

  The following holds

[2] implications to follow in the later sections

1. $(A_1 \otimes \dots A_n)(a_1, a_2, \dots a_n) := A_1(a_1)A_2(a_2)\dots A_n(a_n)$
2. $A_1 \otimes \cdots \otimes A_{k-1} \otimes (aA_k + bB_k) \otimes A_{k+1} \otimes \cdots \otimes A_n := a(A_1 \otimes \cdots \otimes A_{k-1} \otimes (A_k) \otimes A_{k+1} \otimes \cdots \otimes A_n) + b(A_1 \otimes \cdots \otimes A_{k-1} \otimes (B_k) \otimes A_{k+1} \otimes \cdots \otimes A_n)$
3. If we have $C_i \in L(\mathcal{Y}_i, \mathcal{Z}_i)$,
$$(C_1 \otimes \cdots \otimes C_n)(A_1 \otimes \cdots \otimes A_n) := (C_1 A_1) \otimes \cdots \otimes (C_n A_n)$$
4. $(A_1 \otimes \cdots \otimes A_n)^T := (A_1^T \otimes \cdots \otimes A_n^T)$
5. $\overline{(A_1 \otimes \cdots \otimes A_n)} = (\bar{A}_1 \otimes \cdots \otimes \bar{A}_n)$
6. $(A_1 \otimes \cdots \otimes A_n)^* := (A_1^* \otimes \cdots \otimes A_n^*)$

- Some important classes of operators

1. Positive operators. A positive semidefinite is one which can be written as $X = Y^*Y$ for some operator Y. We define
$$Pos(\mathcal{X}) := \{Y^*Y | Y \in L(\mathcal{X})\}$$
2. Density Operators. Positive semidefinite operators having unit trace are called density operators. Also, we define
$$D(\mathcal{X}) := \{\rho \in Pos(\mathcal{X}) | Tr(\rho) = 1\}$$

§§§1.1.3. The Vectorization Map

This is a particulary important concept in quantum information. Definitely deserves its own subsubsection. There is a correspondence between the spaces $L(\mathcal{Y}, \mathcal{X})$ and $\mathcal{X} \otimes \mathcal{Y}$ for spaces $\mathcal{X} = \mathbb{C}^\Sigma$ and $\mathcal{Y} = \mathbb{C}^\Gamma$ given by the map
$$vec : L(\mathcal{Y}, \mathcal{X}) \to \mathcal{X} \otimes \mathcal{Y}$$
defined by
$$vec(E_{ab}) = e_a \otimes e_b$$
for all $a \in \Sigma$ and $b \in \Gamma$. This mapping is a linear bijection. It is also an isometry, because,
$$\langle A, B \rangle = \langle vec(A), vec(B) \rangle \text{ for all } A, B \in L(\mathcal{X}, \mathcal{Y})$$

Some useful identities,

- $Tr_Y(vec(A)vec(B)^*) = AB^*$.
- $Tr_X(vec(A)vec(B)^*) = A^T \bar{B}$.

§2. Basics of Classical Information

§§2.1. Introduction

In this section, we shall deal with the very basic notions of classical information. The setup is as follows. We have a *information source*, which is some entity that emits some sort of information. Vague? We can model this information source as a random variable which can take values in some set of symbols.

Definition 2.1.1 (Information Source)**.** [3] Formally, an information source is a triple (X, Σ, p_X) where X is a random variable which can take values $X = a$ for $a \in \Sigma$ with probability $p_X(a)$.

This definition is equivalent to a probabilistic classical register as defined in **??**. Readers are suggested to go through the same.

Now we ask the question - how does one measure information? One can think of measuring information in terms of measuring the amount of surprise that the information source can provide. Let X_1 and X_2 be two sources[4] with $\Sigma = \{0, 1\}$ for both and let $p_{X_1} \equiv (0.5, 0.5)$ and $p_{X_2} \equiv (1, 0)$ be their probability vectors. Which source encodes more surprise? Yes, the first one. Why?

Definition 2.1.2 (Information content)**.** Given the source (X, Σ, p_X), define a map $i : \Sigma \to \mathbb{R}$ as

$$\begin{aligned} i(a) &= \log \frac{1}{p_X(a)} \text{ if } p_X(a) \neq 0 \\ &= 0 \text{ otherwise} \end{aligned}$$

The quantity $i(a)$ is said to be the information content in the symbol a (associated with this source).

If we see a source emitting a symbol which had less probability of being emitted, we'd be surprised!

Now, what is the surprise associated with the source X? We would define it as $i(X)$![5]. Which is again a random variable. Now we make a very important definition.

Definition 2.1.3 (Shannon Entropy)**.** The shannon entropy associated with the source X is defined as

$$H(X) = \mathbb{E}[i(X)] = -\sum_{a \in \Sigma} p_X(a) \log p_X(a)$$

[3] This definition is coined by yours truly.

[4] an abuse of notation, which unfortunately will be repeated

[5] there is no weirdness here, this is a transformation of random variables

§§2.2. Encodings

Consider a source X with $\Sigma = \{a, b, c, d\}$ and $p_X \equiv (1/2, 1/8/1/4, 1/8)$. We are tasked with coming up with an encoding scheme which uses the least number of average bits. [6] One is the straightforward way

$$a \to 00, b \to 01, c \to 10, d \to 11$$

Another way[7] is

$$a \to 0, b \to 110, c \to 10, d \to 111$$

These binary strings are called *codewords.* Check that the average number of bits needed for the second case is lesser than of the first case!

Can we always associate some codewords to all the symbols efficiently? Shannon suggested the idea of *block encoding.*

Informally, a block encoding involves letting the source emit a large number of symbols $x^n = x_1, x_2 \dots x_n$ for $n > > |\Sigma|$ in a memory-less fashion. More formally, we have a set of independent and identically distributed random variables $(X_1, X_2 \dots X_n)$. Thus,

$$p_{X^n}(x^n) = \Pi_{i=1}^n p_X(x_i)$$

If we denote $N(a|x^{)}$ as the number of times $a \in \Sigma$ occurred in x^n, then we have

$$p_{X^n}(x^n) = \Pi_{i=1}^n p_X(x_i) = \Pi_{a \in \Sigma} p_X(a)^{N(a|x^n)}$$

Definition 2.2.1 (Sample Entropy)**.** Given a sample $x^n = (x_1, \dots, x_n) \in \Sigma^n$, define the sample entropy as

$$-\frac{1}{n} \log p_{X^n}(x^n)$$

Now again, we define another random variable to be the sample entropy of the random vector X^n as

$$-\frac{1}{n} \log p_{X^n}(X^n)$$

Further,

$$\begin{aligned} -\frac{1}{n} \log p_{X^n}(X^n) &= -\frac{1}{n} \log \Pi_{a \in \Sigma} p_X(a)^{N(a|X^n)} \\ &= -\sum_{a \in \Sigma} \frac{N(a|X^n)}{n} \log p_X(a) \end{aligned}$$

As n becomes large, this random variable converges to $H(X)$. Thus, it is highly likely that the source emits a sequence with sample entropy close to the true entropy. This motivates us to define the following.

Definition 2.2.2 (Typical Sequences)**.** A sequences $x^n \in \Sigma^n$ is called a *typical sequence* if its sample entropy is 'close' to the true entropy. The set of all such sequences is called the typical set.

Then, we have the following property.

Lemma 2.2.1 (Asymptotic Equipartition Property)**.**

1. The probability that an emitted sequence is typical becomes large as n increases.
2. The size of the typical set is $\approx 2^{nH(X)}$.
3. The probability of a particular typical sequence is roughly uniform, $\approx 2^{-nH(X)}$.

[6] Shall not elaborate on why we need to do this.

[7] While transmission, a *bitstream* is sent, so keep in mind that it should be uniquely decodable.

§§2.3. Transmission over a channel

The problem setting is as follows. We have a sender, an encoding scheme, a (possibly noisy) channel, a decoding scheme and then a receiver. The sender sends messages from the message set $[M]$ defined as

$$[M] := \{1, 2 \ldots m\}$$

We model the noisy channel as a conditional probability distribution $p_{Y|X}(y|x)$. And as before, we shall transmit a message m using a block encoding $x^n \equiv x_1 x_2 \ldots x_n$ and then receive $y^n \equiv y_1 y_2 \ldots y_n$ on the other side of the channel. The IID assumption again holds, so we have

$$p_{Y^n|X^n}(y^n|x^n) = \Pi_{i=1}^{n} p_{Y|X}(y_i|x_i)$$

We also define the rate of a coding scheme as follows:

$$\text{rate} \equiv \frac{\text{no. of message bits}}{no. of channel uses}$$

Note that transmitting x^n means using the channel n times. So our rate is

$$R = \frac{\log_2 M}{n}$$

With this, we define the *capacity* of a cahnnel to be the highest rate at which it can communicate information reliably.

§3. The Basics of Quantum Information

§§3.1. Registers and Classical States

From a physical standpoint, a register is any entity which has something called a *state*, which may change over time. From a computer science standpoint, it may be thought of as a data storage element. We shall, as expected, use a more abstract formulation.

Definition 3.1.1 (Registers)**.** A register X is either one of the following objects,

1. (Simple Register) An alphabet Σ.
2. (Compund Register) A tuple $(\mathsf{Y}_1,\, \mathsf{Y}_2\, ...\, \mathsf{Y}_n)$ such that $\mathsf{Y}_1,\, ...\, \mathsf{Y}_n$ are all registers for some $n \in \mathbb{N}$.

§§§3.1.1. Classical States

Now let us define the classical state set of a register.

Definition 3.1.2 (Classical State Set of a Register)**.** The classical state set of a register X is determined as follows

1. If X is simple, then the classical state is the associated alphabet Σ.
2. If not, then if $\Sigma_1, \dots \Sigma_n$ are the alphabets of the subregisters, then the classical state of X is $\Sigma = \Sigma_1 \times \cdots \times \Sigma_n$

Elements of the classical state sets are called the classical[8] states. These are the states a register can be in at different points of time. If $\mathsf{X} = (\mathsf{Y}_1,\, \mathsf{Y}_2\, ...\, \mathsf{Y}_n)$, and Σ_i is the state set of Y_i, then for a given state $(a_1, a_2 \dots a_n)$ of X, the state of Y_i is a_i. Conversely, it is easy to determine the state of X given the states of all Y_i.

In additional to classical states of a register X, we can also have *probabilistic* states. A probabilistic state of a register X refers to a probability distribution over the state set of X. Formally, it is a probability vector $p \in \mathcal{P}(\Sigma)$. We shall now see how this is generalized in the quantum setting.

§§§3.1.2. Quantum States

Associated to any register X with the state set Σ, we have a complex euclidean space $\mathcal{X} = \mathbb{C}^\Sigma$. It is good to note that if $\mathsf{X} = (\mathsf{Y}_1,\, \mathsf{Y}_2\, ...\, \mathsf{Y}_n)$, then $\mathcal{X} = \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n$. The reader is encouraged to verify this.

Definition 3.1.3 (Quantum States)**.** A *quantum state* of a register X is a density operator $\rho \in D(\mathcal{X})$.

We first note that $D(\mathcal{X})$ is a convex set. Motivated by this, given a $\rho_a \in D(\mathcal{X}) \forall a \in \Gamma$ and a probability vector $p \in \mathcal{P}(\Gamma)$ for some alphabet Γ, we have that

$$\rho = \sum_{a \in \Gamma} p(a) \rho_a \in D(\mathcal{X})$$

The operator ρ describes a *mixture* given by $\{\rho_a | a \in \Gamma\}$ as per the vector p. This means that a random selection of $a \in \Gamma$ according to the distribution p results in the state of the register X being ρ_a. This can be more formally described as the following.

[8] 'classical' because 'state' will usually refer to a quantum state

Definition 3.1.4 (Ensemble of States)**.** For some alphabet Γ and a complex euclidean space $\mathcal{X}$, define a map

$$\eta : \Gamma \to Pos(\mathcal{X})$$

with the constraint

$$Tr(\sum_{a\in\Gamma} \eta(a)) = 1$$

The map η specifies as ensemble with the corresponding states as

$$\rho_a = \frac{\eta(a)}{Tr(\eta(a))}$$

and the probability vector $p = \{Tr(\eta(a))|a \in \Gamma\}$

Now, let us comment on the nature of these quantum states. We begin with the notion of *pure states.*

Definition 3.1.5 (Pure States)**.** A state $\rho \in D(\mathcal{X})$ is said to be a pure state if it has rank one. Equivalently, ρ is a pure state if there exits a unit vector $u \in \mathcal{X}$ such that $\rho = uu^*$.

A pure state is, with some abuse of notation, often referred to as the vector u. Next up we have *flat states.*

Definition 3.1.6 (Flat States)**.** A quantum state $\rho \in D(\mathcal{X})$ is said to be a flat state if it holds that

$$\rho = \frac{\Pi}{Tr(\Pi)}$$

for some non zero projection $\Pi \in Proj(\mathcal{X})$.

Now, with the definition of quantum states out of the way, we can explain a much intuitive idea - that probabilistic classical states are a restriction of quantum states. For some register X with the state set Σ and a probabilistic state $p \in \mathcal{P}(\Sigma)$. We can associate the density operator E_{aa} with the register being in the state a. Hence, we have

$$p \equiv \sum_{a\in\Sigma} p(a)E_{aa}$$

Hence a classical register can be thought of as a general quantum register whose density operator is restricted to be diagonal. Next up, are product states.

Definition 3.1.7 (Product States)**.** Given a register $\mathsf{X} = (\mathsf{Y}_1,\ldots\mathsf{Y}_n)$, a state $\rho \in D(\mathcal{X})$ is said to be a product state if

$$\rho = \sigma_1 \otimes \cdots \otimes \sigma_n$$

such that $\sigma_i \in D(\mathcal{Y}_i)$.

Intuitively, a product state indicates independence between the subregisters. If the state of X is not a product state, it indicates towards some correlation betweent the subregisters [9].

§§§3.1.3. Reductions and Purification

To motivate this title, we can pose one question - for some compound register, what if we remove one of the subregisters? How do we describe the state of the leftover regiser then? This is perhaps easy to do for a classical state. If the state before removal is $(a_1, a_2 \ldots a_n)$, then after removing, say the first subregister we simply are in the state $(a_2 \ldots a_n)$. What follows below is *the* way to do this for a quantum state.

Consider $\mathsf{X} = (\mathsf{Y}_1,\ldots\mathsf{Y}_{k-1},\mathsf{Y}_k,\mathsf{Y}_{k+1}\ \mathsf{Y}_n)$. We can create a new register $(\mathsf{Y}_1,\ldots\mathsf{Y}_{k-1},\mathsf{Y}_{k+1}\ldots\mathsf{Y}_n)$ by removing Y_k from X. The state of the leftover system $\rho[\neg k]$ is defined as

$$\rho[\neg k] := Tr_k(\rho)$$

where $Tr_k \in T(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n, \mathcal{Y}_1 \otimes \ldots \mathcal{Y}_{k-1} \otimes \mathcal{Y}_{k+1} \otimes \cdots \otimes \mathcal{Y}_n)$ is the *partial trace* map. We have,

[9] again, this notion will be made precise in one of the later sections

1. $Tr_k(Y_1 \otimes \cdots \otimes Y_n) = Tr(Y_k)(Y_1 \otimes Y_2 \otimes \ldots Y_{k-1} \otimes Y_{k+1} \otimes \cdots \otimes Y_n)$
2. $Tr_k = I_1 \otimes I_2 \otimes \ldots I_{k-1} \otimes Tr_k \otimes I_{k+1} \otimes \cdots \otimes I_n$

Example: Let $\mathsf{X} = (\mathsf{Y},\mathsf{Z})$ be a compound register such that the state sets of the subregisters are both Σ. For the state

$$\rho = \frac{1}{|\Sigma|} \sum_{a,b \in \Sigma} E_{ab} \otimes E_{ab}$$

We have

$$\rho[\not{Z}] = \frac{1}{|\Sigma|} \sum_{a,b \in \Sigma} E_{ab} Tr(E_{ab}) = \frac{1}{|\Sigma|} I_{\mathcal{Y}}$$

Such a state is analogous to uniform distribution, and represents the most chaotic state a register could have. In fact, one can show that the state ρ is a pure state (left to the reader). Thus, this example makes concrete the idea of correlations between subregisters.

Thus, by applying this method iteratively, one can specify the state of a register after removal of any[10] number of subregisters.

As we discussed reductions, a natural idea is extensions. That is, a register X in the state ρ can be viewed as a subregister of some 'higher' register (X,Z) in the state σ such that

$$\rho = Tr_Z(\sigma)$$

σ is called an extension of ρ. Now, the idea of purifying a state is simply to find such a Z such that σ is pure. More formally, we have

Definition 3.1.8 (Purification)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $P \in \text{Pos}(\mathcal{X})$ be a positive semidefinite operator, and let $|u\rangle \in \mathcal{X} \otimes \mathcal{Y}$ be a vector. The vector $ketu$ is said to be a purification of P if

$$\text{Tr}_{\mathcal{Y}} \left(|u\rangle \langle u| \right) = P$$

Recall that for spaces $\mathcal{X}$ and $\mathcal{Y}$, the vec mapping is a bijection from $L(\mathcal{Y}, \mathcal{X}) \to \mathcal{X} \otimes \mathcal{Y}$. Every vector $|u\rangle \in \mathcal{X} \otimes \mathcal{Y}$ can be written as $|u\rangle = vec(A)$ for some $A \in L(\mathcal{Y}, \mathcal{X})$. Also,

$$Tr_{\mathcal{Y}}(|u\rangle \langle u|) = Tr_{\mathcal{Y}}(vec(A)vec(A)^*) = AA^*$$

So in order to purify some positive P[11] it should be of the form AA^*. The following theorem makes this idea concrete.

Theorem 3.1.1 (The Purification Theorem)**.** Let $\mathcal{X}, \mathcal{Y}$ be spaces. And let $P \in Pos(\mathcal{X})$ be a positive semidefinite opeartor. There exists a vector $u \in \mathcal{X} \otimes \mathcal{Y}$ such that $Tr_{\mathcal{Y}}(|u\rangle \langle u|) = P$ **iff** $dim(\mathcal{Y}) \geq rank(P)$.

If the statement of the theorem is read carefully, one notices the 'there exists a $|u\rangle$'. So can there be more than one such $|u\rangle$?

Theorem 3.1.2 (Unitary Equivalence of Purifications)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex euclidean spaces, and let $|u\rangle, |v\rangle \in \mathcal{X} \otimes \mathcal{Y}$ be vectors such that

$$Tr_{\mathcal{Y}}(|u\rangle \langle u|) = Tr_{\mathcal{Y}}(|v\rangle \langle v|)$$

Then, there exists a unitary $U \in U(\mathcal{Y})$ such that $|v\rangle = (I_{\mathcal{X}} \otimes U) |u\rangle$.

[10] the resulting object must be a valid register
[11] we drop the trace condition without any loss of analytical work

§§3.2. Quantum Channels

Definition 3.2.1. Quantum Channels A *quantum channel* (or just channel) is a linear map

$$\Phi : L(\mathcal{X}) \to L(\mathcal{Y})$$

that is, $\Phi \in T(\mathcal{X}, \mathcal{Y})$ which satisfies

1. Φ is completely positive.
2. Φ is trace preserving.

The set of all such channels is called $C(\mathcal{X}, \mathcal{Y})$. As usual, we denote $C(\mathcal{X}, \mathcal{X})$ as $C(\mathcal{X})$.

Intuitively, for registers X and Y, one may view a channel $\Phi \in C(\mathcal{X}, \mathcal{Y})$ as the process of destroying X and creating Y, a transformation of X into Y. When $\mathsf{X} = \mathsf{Y}$, one may simply view it as changing the state of X.

Physics Fact

In quantum mechanics we usually deal with unitary evolution of quantum states. Since density operators are more general than pure states, naturally we expect they can undergo transformations which are not necessarily unitary! We shall see this soon.

Example 3.2.1.

- (Unitary Channels) The map $\Phi \in C(\mathcal{X})$ given by $\Phi(X) = UXU^*$ for some $U \in U(\mathcal{X})$ is called a unitary channel.
- (Replacement Channels) The map $\Phi \in C(\mathcal{X}, \mathcal{Y})$ given by $\Phi(X) = Tr(X)\sigma$ for some $\sigma \in D(\mathcal{Y})$ is called a replacement channel.

As we have dealt with product states, we also have product channels. A channel $\Phi \in C(\mathcal{X}_1 \otimes \mathcal{X}_2 \ldots \mathcal{X}_n, \mathcal{Y}_1 \otimes \mathcal{Y}_2 \ldots \mathcal{Y}_n)$ is called a product channel if

$$\Phi = \phi_1 \otimes \phi_2 \ldots \phi_n$$

for $\phi_i \in C(\mathcal{X}_i, \mathcal{Y}_i)$. Intuitively, a channel being a product channel represents an independent application of channels to a sequence of registers.

The next subsection goes out to fans of representation theory.

§§§3.2.1. Representing Quantum Channels

In certain situations, we might want to have a concrete *represenetation* of a quantum channel rather than treat it as an abstract map. We discuss three representations here.

1. *The Choi Representation*
   For spaces $\mathcal{X}, \mathcal{Y}$, define $J : T(\mathcal{X}, \mathcal{Y}) \to L(\mathcal{Y} \otimes \mathcal{X})$ as

   $$J(\Phi) = (\Phi \otimes \mathbb{1}_{L(\mathcal{X})})(vec(\mathbb{1}_{\mathcal{X}})vec(\mathbb{1}_{\mathcal{X}})^*) \text{ for all } \Phi \in T(\mathcal{X}, \mathcal{Y})$$

   If $\mathcal{X} = \mathbb{C}^\Sigma$ we have

   $$J(\Phi) = \sum_{a,b \in \Sigma} \Phi(E_{ab}) \otimes E_{ab}$$

   $J(\Phi)$ is called the *Choi Representation* of Φ. The mapping J is a bijection. The rank of $J(\Phi)$ is called the *Choi Rank* of Φ.

2. *Kraus Operators*
   For spaces $\mathcal{X}$ and $\mathcal{Y}$, and for some alphabet Σ if we have collections of operators $A_a, B_a \in L(\mathcal{X}, \mathcal{Y})$ for $a \in \Sigma$, we may define
   $$\Phi(X) := \sum_{a \in \Sigma} A_a X B_a^*$$
   We usually deal with the case when $A_a = B_a$ for all $a \in \Sigma$. We will see that such representations exist for CPTP channels, but they are *not* unique.

3. *Stinespring Representations*
   For spaces $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$, and operators $A, B \in L(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ we can define $\Phi \in T(\mathcal{X}, \mathcal{Y})$ as
   $$\Phi(X) = Tr_Z(AXB^*) \text{ for all } X \in L(\mathcal{X})$$
   Similar to kraus representations, we commonly encounter $A = B$ and such representations exist for CPTP maps but are not unique.

Now, let us see some relations between these representations.

Lemma 3.2.1 (Relations)**.** Let spaces $\mathcal{X}, \mathcal{Y}$ and $A_a, B_a \in L(\mathcal{X}, \mathcal{Y})$ for $a \in \Sigma$, and $\Phi \in T(\mathcal{X}, \mathcal{Y})$. The following statements are equivalent.

1. The Choi Representation is
   $$J(\Phi) = \sum_{a \in \Sigma} vec(A_a) vec(B_a)^*$$
2. The Kraus Representation is
   $$\Phi(X) = \sum_{a \in \Sigma} A_a X B_a^*$$
3. For $\mathcal{Z} = \mathbb{C}^\Sigma$ and $A, B \in L(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ defined as $A := \sum_a A_a \otimes e_a$ and $B := \sum_a B_a \otimes e_a$ we have the stinespring representation as
   $$\Phi(X) = Tr_Z(AXB^*)$$

With this, we have an important corollary.

Corollary 3.2.1. For spaces $\mathcal{X}, \mathcal{Y}$ and a nonzero map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ let $r = rank(J(\Phi))$ be its choi rank. Then,

1. There exists a kraus representation of Φ with $|\Sigma| = r$.
2. There exists a stinespring representation of Φ with $dim(\mathcal{Z}) = r$.

Now, let us see a useful result which charecterizes CPTP maps.

Theorem 3.2.1 (CPTP Maps)**.** Let $\Phi \in T(\mathcal{X}, \mathcal{Y})$ a map. The following are equivalent.

1. Φ is a quantum channel (CPTP map).
2. $J(\Phi) \in Pos(\mathcal{Y} \otimes \mathcal{X})$ and $Tr_Y(J(\Phi)) = \mathbb{1}_X$.
3. There exists alphabet Σ and collection $\{A_a | a \in \Sigma, A_a \in L(\mathcal{X}, \mathcal{Y})\}$ such that $\sum_{a \in \Sigma} A_a^* A_a = \mathbb{1}_X$ and
   $$\Phi(X) = \sum_{a \in \Sigma} A_a X A_a^*$$
   Further, this holds for $|\Sigma| = rank(J(\Phi))$.

4. There exists an isometry $A \in U(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ for osme space $\mathcal{Z}$ such that

$$\Phi(X) = Tr_Z(AXA^*)$$

Further, this holds for $dim(\mathcal{Z}) = rank(J(\Phi))$.

Let us see some more examples of channels with their representations.

- Completely depolarizing channel: $\Omega \in C(\mathcal{X})$ for $\mathcal{X} = \mathbb{C}^\Sigma$ is defined as

$$\Omega(X) := Tr(X)\omega$$

where

$$\omega = \frac{\mathbb{1}_X}{dim(\mathcal{X})}$$

is the completely mixed state. That is, the completely depolarizing channel turns every state into the completely mixed state. We see that

1. The Choi Representation is

$$\begin{aligned} J(\Omega) &= \sum_{a,b\in\Sigma} \Omega(E_{ab}) \otimes E_{ab} \\ &= \sum_{a,b\in\Sigma} \delta_{ab}\omega \otimes E_{ab} \\ &= \frac{\mathbb{1}_X \otimes \mathbb{1}_X}{dim(\mathcal{X})} \end{aligned}$$

2. The Kraus Representation can be seen by noting that

$$\begin{aligned} J(\Omega) &= \frac{1}{dim(\mathcal{X})} \sum_{ab} |a\rangle \langle a| \otimes |b\rangle \langle b| \\ &= \frac{1}{dim(\mathcal{X})} \sum_{ab} |a\rangle |b\rangle \otimes \langle a| \langle b| \end{aligned}$$

Now by the relation lemma, and the fact that $vec(|a\rangle \langle b|) = |a\rangle |b\rangle$, see that the kraus operators are $E_{ab}/dim(\mathcal{X})$ for $a, b \in \Sigma$.

- The completely dephasing channel: Let $\mathcal{X} = \mathbb{C}^\Sigma$ and define $\Delta \in T(\mathcal{X})$ as

$$\Delta(X) := \sum_{a\in\Sigma} X(a,a)E_{aa}$$

That is, this channel kills off all the non diagonal entries of the density operator, leaving us in an essentially classical (probabilistic) state. Moreover, it is the identity on diagonal density operators.

1. The Choi Representation is given by

$$\begin{aligned} J(\Delta) &= \sum_{a,b\in\Sigma} \Delta(E_{ab}) \otimes E_{ab} \\ &= \sum_{a\in\Sigma} E_{aa} \otimes E_{aa} \end{aligned}$$

2. (One of) The Kraus Representation is simply given by

$$\Delta(X) = \sum_{a\in\Sigma} E_{aa} X E_{aa}^*$$

3. Stinespring: If we define $A := \sum_{a\in\Sigma}(e_a \otimes e_a)e_a^*$ we have the stinespring representation as

$$\Delta(X) = Tr_Z(AXA^*)$$

for $\mathcal{Z} = \mathbb{C}^\Sigma$.

§§3.3. Measurements

From an introductary knowledge of quantum mechanics, we know that measurements are associated with a 'collapse of the wave function'. This involves reading out a classical value, known as the measurement outcome - some eigenvalue, and we consider the wavefunction to collapse to a eigenfunction with that eigenvalue. Here, we shall define measurements in two forms,

1. The first way will only concern with the outputs of the measurement.
2. The second way will also talk about the state of the register after measurement.

§§§3.3.1. The First Way

Definition 3.3.1. A measurmenet is a function of the form

$$\mu : \Sigma \to Pos(\mathcal{X})$$

for some choice of alphabet Σ and a space $\mathcal{X}$, with the constraint,

$$\sum_{a \in \Sigma} \mu(a) = \mathbb{1}_{\mathcal{X}}$$

A few points,

- The set Σ is the set of measurement outcomes.
- Each $\mu(a)$ is a *measurement operator* corresponding to the outcome a.
- When this measurement happens, two things occur for a register in the state $\rho \in D(\mathcal{X})$
  1. An element a of Σ is selected with the probability $p(a) = Tr(\mu(a)\rho)$[12], and is thrown out as the result of the measurement.
  2. The register ceases to exist, and we cannot do any further work with it.

Physics Fact

> One can associate this formalism with the POVM measurement set up. That is, we have a set of measurement operators $P_a \equiv \mu(a)$, and we can only talk about the measurement outcomes and the measurement statistics, and not about the post-measurement states. We shall alternate between the two notations.

An alternate, but *equivalent*[13] way to describe this formalism is by identifying measurements as a channel. Let us define a useful kind of channel, called a quantum-classical channel.

Definition 3.3.2 (Quantum-Classical Channel)**.** Let $\Phi \in C(\mathcal{X}, \mathcal{Y})$ be a channel. It is said to be Q2C if

$$\Phi = \Delta\Phi$$

for $\Delta \in C(\mathcal{Y})$ denoting the completely dephasing channel in $\mathcal{Y}$.

It is easy to see that every Q2C channel outputs a diagonal density matrix. This should be fairly intuitive in establishing the equivalence between Q2C channels and measurements as defined earlier. Formally, we have the following theorem.

[12] The reader is encouraged to verify that this is a valid probability distribution.
[13] to be proved!

Theorem 3.3.1 (Equivalence)**.** Let $\mathcal{X}$ be a space and Σ be an alphabet and let $\mathcal{Y} = \mathbb{C}^{\Sigma}$. The following facts hold,

1. For every Q2C channel $\Phi \in C(\mathcal{X}, \mathcal{Y})$ there exists a unique measurment $\mu : \Sigma \to Pos(\mathcal{X})$ such that

$$\Phi(\rho) = \sum_{a \in \Sigma} E_{aa} Tr(\mu(a)\rho)$$

2. For every measurment $\mu : \Sigma \to Pos(\mathcal{X})$ the mapping $\Phi \in T(\mathcal{X}, \mathcal{Y})$ defined by the above equation is a Q2C channel.

Let us now discuss some particular types of measurements.

- Essentially, every quantity we defined had a form wherein it factored out, so to speak. So, given a compound register $\mathsf{X} = (\mathsf{Y}_1, \dots \mathsf{Y}_n)$, if one has measurements on each of the subregisters as

$$\mu_i : \Sigma_i \to Pos(\mathcal{Y}_i)$$

then we may define a composite measurement as

$$\mu : \Sigma_1 \times \Sigma_2 \dots \Sigma_n \to Pos(\mathcal{X})$$

as

$$\mu(a_1, \dots a_n) = \mu_1(a_1) \otimes \dots \otimes \mu_n(a_n)$$

Such a measurement is called a *product measurement.*

- Again, we have a notion of doing something with only a subset of registers of a compound register. So, we talk about partial measurements. Given $\mathsf{X} = (\mathsf{Y}_1, \dots \mathsf{Y}_n)$, and a measurement

$$\mu_k : \Sigma \to Pos(\mathcal{Y}_k)$$

only on one register, for some $k \in [n]$. We demand two things of this measurement-

1. An outcome $a \in \Sigma$.
2. Specification of the register X/Y_k.[14]

The Q2C way of describing helps us here. We can apply

$$\Phi(Y) := \sum_{a \in \Sigma} Tr(\mu_k(a)Y) E_{aa} \forall Y \in L(\mathcal{Y}_k)$$

to the register Y_k, and then performing a permutation of registers, that is

$$(\mathsf{Y}_1, \dots, \mathsf{Y}_k, \dots \mathsf{Y}_n) \to (\mathsf{Y}_1, \dots, \mathsf{Z}, \dots \mathsf{Y}_n) \to (\mathsf{Z}, \mathsf{Y}_1, \dots \mathsf{Y}_n)$$

Thus we get

$$\sum_{a \in \Sigma} E_{aa} \otimes Tr_{\mathsf{Y}_k}((\mathbb{1} \otimes \dots \mu_k(a) \otimes \dots \mathbb{1})\rho)$$

If we let $\eta(a) := Tr_{\mathsf{Y}_k}(\mathbb{1} \otimes \dots \mu_k(a) \otimes \dots \mathbb{1})\rho$, each outcome a occurs with the probability

$$p(a) = Tr(\eta(a)) = Tr(\mu(a)\rho[Y_k]$$

leaving the rest of the registers in the state

$$\frac{\eta(a)}{Tr(\eta(a))}$$

[14] forgive the abuse of set notation

Let us now talk about projective measurements. Simply, a projective measurement is one for which the operators $\mu(a)$ are projections. That is $\mu(a)^2 = \mu(a)$ for all a. This we denote as $\mu(a) \in Proj(\mathcal{X})$. The following lemma establishes an important fact, which we shall relate to with the next *physics fact*.

Lemma 3.3.1. For alphabet Σ, and space $\mathcal{X}$ let $\mu : \Sigma \to Pos(\mathcal{X})$ be a projective measurement. The set

$$\{\mu(a)|a \in \Sigma\}$$

is an orthogonal set.

Physics Fact

In a standard treatment of quantum mechanics, we deal with *measuring observables*. These observables are hermitian operators, and hence have a spectral decomposition as

$$O = \sum_m mP_m$$

[15] where, m's are the eigenvalues (the observed values) and P_m's are the projectors onto the corresponding eigenspaces. Here, these P_m take the form of the projective measurement, and upon measuring m a state $|\psi\rangle$ collapses to $P_m |\psi\rangle$ (upto normalization). Also, in case we have full non-degeneracy we have $P_m = |m\rangle \langle m|$ with $|m\rangle$ being the eigenket of O with eigenvalue m.

Next up, we will answer the troubled physicists who have never dealt with any measurements except in the traditional projective sense. [16]

Theorem 3.3.2 (Naimark's Theorem)**.** Let $\mathcal{X}$ be a space and Σ be an alphabet. Let $\mu : \Sigma \to Pos(\mathcal{X})$ be a measurment and let $\mathcal{Y} = \mathbb{C}^{\Sigma}$. There exists an isometry $A \in U(\mathcal{X}, \mathcal{X} \otimes \mathcal{Y})$ such that

$$\mu(a) = A^*(\mathbb{1}_X \otimes E_{aa})A$$

for $a \in \Sigma$.

With this we have the following corollary.

Corollary 3.3.1. Let $\mathcal{X}$ be a space and Σ be an alphabet. Let $\mu : \Sigma \to Pos(\mathcal{X})$ be a measurment and let $\mathcal{Y} = \mathbb{C}^{\Sigma}$. Further, let $|u\rangle \in \mathcal{Y}$ be a unit ket. There exists a projective measurement $\nu : \Sigma \to Pos(\mathcal{X} \otimes \mathcal{Y})$ such that

$$Tr(\nu(a)(X \otimes |u\rangle \langle u|)) = Tr(\mu(a)X)$$

for all $\mathcal{X} \in L(\mathcal{X})$.

Physics Fact

Ah yes, reassurance. Why? The preceding corollary establishes that our measurement formalism (that is, POVMs) are equivalent to projective measurements at the cost of extending the register to a larger one. This motivates the following:

1. *General*[17] Measurements on a system are projective measurements on the system coupled with a large enough (as quantified by the corollary) environment.

2. Quantum Correlations are a thing. Our formalism is useful. If we accept that physical measurements correspond to hermitian observables (that is, projective measurements), then if we perform such a physical measurement on a system coupled with an environment, then the statistics of the system alone can not be described by a physical measurement. Thus the need for our (or POVM) formalism is justified.

[16] Some might say, they are *born* with it.

Now that we are familiar with the concepts regarding measurements, let us ask a question.

Can we infer a state entirely from measurement statistic? What kind of a measurement should this be?

This translates to the map from density operator space to the space of probability vectors being injective, and is certainly an interesting thought. Let us define the following, and see why it answers this question.

Definition 3.3.3 (Information Complete Measurements)**.** A measurement $\Sigma \to Pos(\mathcal{X})$ for some space $\mathcal{X}$ is said to be information complete if it holds that

$$span\{\mu(a)|a \in \Sigma\} = L(\mathcal{X})$$

That is, from the probability vector of such a measurement defined by

$$p(a) \equiv Tr(\mu(a)\rho)$$

we can infer ρ![18]

Lemma 3.3.2. Let $\Sigma = \{1, 2 \ldots n\}$, let $\mathcal{X}$ be a space and let $\{A_a|a \in \Sigma\} \subset L(\mathcal{X})$ be a collection of operators for which

$$span\{A_a|a \in \Sigma\} = L(\mathcal{X})$$

Then, the mapping $\phi : L(\mathcal{X}) \to \mathbb{C}^{\Sigma}$ defined by

$$\phi(X) := (Tr(A_1X), Tr(A_2X) \ldots Tr(A_nX)^T$$

is injective.

GIVE EXAMPLE!

§§§3.3.2. The Second Way

Let us define the non-destructive measurements.

Definition 3.3.4. A non destructive measurement on the space $\mathcal{X}$ is described by an alphabet Σ and a collection

$$\mathcal{M} = \{M_a|a \in \Sigma\} \subset L(\mathcal{X})$$

satisfying $\sum_{a \in \Sigma} M_a^* M_a = \mathbb{1}$. Two things happen when $\mathcal{M}$[19] is applied to a register in the state ρ

1. A symbol $a \in \Sigma$ is picked with probability

$$p(a) = Tr(M_a \rho M_a^*)$$

2. Conditioned on the a selected, the state of the register is changed to

$$\frac{M_a \rho M_a^*}{Tr(M_a \rho M_a^*)}$$

One can also talk about this procedure in an even more general sense with the CP maps

$$\{\Phi_a|a \in \Sigma\} \subset CP(\mathcal{X}, \mathcal{Y})$$

with the constraint

$$\sum_{a \in \Sigma} \Phi_a \in C(\mathcal{X}, \mathcal{Y})$$

Again, when applied to some register in the state $\rho \in D(\mathcal{X})$, two things happen

[18] in the sense that no other state can lead to that statistics vector.

[19] People often misconceptualize what 'applying a measurement' means: It should not be confused with evolutions from operators from $\mathcal{M}$, it is the application of $\mathcal{M}$ (which includes hermitian observables) as a whole.

- An element $a \in \Sigma$ is selected at random with probability

$$p(a) = Tr(\Phi_a(\rho))$$

- Conditioned on that selection, the register is transformed into a new register in the state

$$\frac{\Phi_a(\rho)}{Tr(\Phi_a(\rho))}$$

This formalism is called a *quantum instrument.* It includes the general measurement formalism with the following definition

$$\Phi_a(X) = M_a X M_a^*$$

Exercise 3.3.1. Show that the map $\Phi \in C(\mathcal{X}, \mathcal{Z} \otimes \mathcal{Y})$ defined as

$$\Phi(X) := \sum_{a \in \Sigma} E_{aa} \otimes \Phi_a(X)$$

followed by a measurement on the Z register can implement any quantum instrument as defined above.