University of Houston

Classical and Quantum Information Theory

Math 6397

Li Gao

Khalid Hourani

Contents

1 Overview

Information theory studies the processing, quantification, storage, and communication of information.

- 1948 — Claude Shannon defines *Shannon Entropy* in “The Mathematical Theory of Communication.” Answers questions:
  1. What is information?
  2. How do we quantify information?
  3. How do we transmit information?
- 2001 — Shannon Award is created, with Shannon the first recipient.
- 1900 — Max Plank describes Black-body Radiation
- 1920s — Heisenberg, Bohr, and Schrödinger, Matrix Mechanics
- 1930s — Hilbert, Dirac, Von Neumann describe the Hilbert Space, Mathematical foundation of Quantum Mechanics, and Von Neumann Entropy
- Interaction: Quantum Information
- 1950s – 1970s — Mathematical Quantities of Information
- 1970s
  - Information Transmission by Coherent Laser
  - Alexander Holevo — Holevo Bound
    - 1998 — Holevo et al show bound is tight (receive 2017 Shannon Award)
- 1980s — Richard Feynman: Computing with Quantum Mechanical Model
- 1990s — Peter Schor: Quantum Algorithm for Prime Factorization
  - In general, the only known algorithm for determining the prime factors of a number is naïve factorization. For example, given $n = 4801 \times 35317 = 169556917$, to retrieve the factors 4801 and 35317 requires substantially more time than to simply construct the number via multiplication.
- let’s finish the rest of the trivia chapter later

2 Probability Theory

A discrete probability space $(\Omega, \mathbb{P})$ is given by

- a finite or countably infinite set Ω
  - e.g. $\{a, b, c, d\}$, $\mathbb{N} = \{0, 1, 2, \dots\}$
- a probability mass function $\mathbb{P} : \Omega \to [0, 1]$, such that
  (1) For all $\omega \in \Omega$, $\mathbb{P}(\omega) \geq 0$
  (2) $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$

  For $\omega \in \Omega$, $\mathbb{P}(\omega)$ is the probability that ω "occurs"

Definition 2.1 ▶ Event

Given a probability space $(\Omega, \mathbb{P})$, an *event* E is a subset $E \subseteq \Omega$, with corresponding probability

$$\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\omega)$$

The function $\mathbb{P} : \Omega \to [0, 1]$ induces a *probability distribution*,

$$\mathbb{P} : 2^{\Omega} \to [0, 1]$$

also denoted by $\mathbb{P}$, with properties:

(1) if $A \cap B = \emptyset$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

(2) $\mathbb{P}(\Omega) = 1$

As an abuse of notation, we write $\mathbb{P}(\omega)$ and $\mathbb{P}(\{\omega\})$ interchangeably.

Example 2.1 ▶ Rolling a fair die

TBD

Definition 2.2 ▶ Conditional Probability

Let $A, B \subseteq \Omega$. The *conditional probability* of A given B, denoted by $\mathbb{P}(A \mid B)$, is defined

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Example 2.2 ▶ Fair Die Revisited

TBD

Theorem 2.1 ▶ Bayes' Rule

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\,\mathbb{P}(B)}{\mathbb{P}(A)}$$

Proof. By definition,

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

hence

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(B \mid A)\,\mathbb{P}(A) \\ &= \mathbb{P}(A \mid B)\,\mathbb{P}(B)\end{aligned}$$

from which the result follows. □

Example 2.3 ▶ Flipping a fair coin twice

TBD

Definition 2.3 ▶ Random Variable

A *Random Variable* X is a function

$$X : \Omega \to \mathcal{X}$$

from probability space $(\Omega, \mathbb{P})$ to a target space $\mathcal{X}$. We say X is discrete if $\mathcal{X}$ is discrete and call

$$\mathcal{X} = \{x_1, x_2, \dots\}$$

the *alphabet* of X.

Notice that X induces a distribution on $\mathcal{X}$. For any $x \in \mathcal{X}$

$$\mathbb{P}_X(x) = \mathbb{P}(\{\omega \mid X(\omega) = x\})$$

In many cases, $(X.\mathbb{P}_x)$ captures all information needed from random variable X. We write $X \sim \mathbb{P}_x$ to indicate that X has distribution $\mathbb{P}_x$ on $\mathcal{X}$.

Example 2.4 ▶ 52 Card Deck

TBD

Definition 2.4 ▶ Joint Distribution

Let $X : \Omega \to \mathcal{X}$, $Y : \Omega \to \mathcal{Y}$ be two random variables. The *joint distribution* on $\mathcal{X} \times \mathcal{Y}$ is given by

$$\mathbb{P}_{XY}(X = x, Y = y) = \mathbb{P}(\{X(\omega) = x, Y(\omega) = y\})$$

For subsets $E_1 \subseteq \mathcal{X}$, $E_1 \subseteq \mathcal{Y}$

$$\mathbb{P}_{XY}(X \in E_1, Y \in E_2) = \mathbb{P}(\{X(\omega) \in E_1, Y(\omega) \in E_2\})$$

Notice that $\mathbb{P}_{XY}$ is a distribution on the product space $(\mathcal{X} \times \mathcal{Y}, \mathbb{P}_{XY})$.

Example 2.5 ▶ Fair Die Joint Distribution

TBD

Example 2.6 ▶ Flipping a fair coin twice joint distribution

TBD

Definition 2.5 ▶ Independent Random Variables

Two random variables X and Y are *independent* if, for any x, y

$$\mathbb{P}_{XY}(X = x, Y = y) = \mathbb{P}_X(X = x)\,\mathbb{P}_Y(Y = y)$$

Equivalently, if for any subsets E_1 and E_2

$$\mathbb{P}_{XY}(X \in E_1, Y \in E_2) = \mathbb{P}_X(X \in E_1)\,\mathbb{P}_Y(Y \in E_2)$$

Definition 2.6 ▶ Product Probability

Given two probability spaces $(\Omega_1, \mathbb{P}_1)$, $(\Omega_2, \mathbb{P}_2)$

$$\mathbb{P}_1 \times \mathbb{P}_2(E_1 \times E_2) = \mathbb{P}_1(E_1)\,\mathbb{P}_2(E_2)$$

is the product probability on $\Omega_1 \times \Omega_2$.

Thus, we have the property that X and Y are independent random variables if and only if $\mathbb{P}_{XY} = \mathbb{P}_X \times \mathbb{P}_Y$.

Example 2.7 ▶ Rank and Suit of a card

TBD

Definition 2.7 ▶ Real Random Variable

A *Real Random Variable* is a function

$$X : \Omega \to \mathbb{R}$$

For example, the height of a randomly sampled person, the value of a die, and the rank of a playing card (where Ace is 1, Jack is 11, Queen is 12, and King is 13) are all real random variables. On the other hand, the suit of a playing card is *not* a real random variable.

In the discrete case, if $X : \Omega \to \mathcal{X}$ is a random variable, then

$$\mathbb{P}_X : X \to [0, 1]$$

is a real random variable.

Definition 2.8 ▶ Conditional Distribution

Given two random variables X and Y, the conditional distribution is the real random variable given by

$$\mathbb{P}_{X|Y} : \mathcal{X} \times \mathcal{Y} \to [0, 1]$$

where

$$\mathbb{P}_{X|Y}(x \mid y) = \mathbb{P}(X = x \mid Y = y)$$

Given two real random variables X and Y, we can define

- $X + Y$
- $X \cdot Y$
- $f(X)$ (where $f : \mathbb{R} \to \mathbb{R}$)

as new random variables.

Definition 2.9 ▶ Expectation and Variance

The *expected value* (or expectation or mean) of a real random variable X is defined as the real number

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x\,\mathbb{P}_X(X = x) = \sum_{\omega \in \Omega} X(\omega)\,\mathbb{P}(X = \omega)$$

The *variance* is defined as

$$\mathrm{Var}[X] = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$$

Example 2.8 ▶ Expected Value and Variance of a Fair Die

TBD

Theorem 2.2 ▶ Linearity of Expectation

Let X and Y be real random variables and $a, b \in \mathbb{R}$. Then

$$\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$$

Proof. By definition,

$$\begin{aligned}
\mathbb{E}[aX + bY] &= \sum_{\omega\in\Omega} (aX(\omega) + bY(\omega))\,\mathbb{P}(\omega) \\
&= \sum_{\omega\in\Omega} aX(\omega)\,\mathbb{P}(\omega) + \sum_{\omega\in\Omega} bY(\omega)\,\mathbb{P}(\omega) \\
&= a\sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\omega) + b\sum_{\omega\in\Omega} Y(\omega)\,\mathbb{P}(\omega) \\
&= a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]
\end{aligned}$$

□

Corollary 2.3

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

Proof. By definition,

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[X^2 - 2X\,\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[2X\,\mathbb{E}[X]] + \mathbb{E}\left[\mathbb{E}[X]^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\,\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2
\end{aligned}$$

□

If X and Y are independent, we have the following

Theorem 2.4

Let X and Y be independent real random variables. Then

(1) $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

(2) $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$

Proof. First, Item (1):

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} xy\,\mathbb{P}_{XY}(X = x, Y = y) \\
&= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} xy\,\mathbb{P}_X(X = x)\,\mathbb{P}_Y(Y = y) \text{ since } X \text{ and } Y \text{ are independent} \\
&= \sum_{x\in\mathcal{X}} x\,\mathbb{P}_X(X = x) \sum_{y\in\mathcal{Y}} y\,\mathbb{P}_Y(Y = y) \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]
\end{aligned}$$

Now,

$$\begin{aligned}
\mathrm{Var}[X + Y] &= \mathbb{E}\left[(X + Y)^2\right] - \mathbb{E}[X + Y]^2 \\
&= \mathbb{E}\left[X^2 + 2XY + Y^2\right] - (\mathbb{E}[X] + \mathit{expectation}Y)^2 \\
&= \mathbb{E}\left[X^2\right] + 2\,\mathbb{E}[XY] + \mathbb{E}\left[Y^2\right] - \mathbb{E}[X]^2 - 2\,\mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[Y]^2 \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathbb{E}[XY] - 2\,\mathbb{E}[X]\,\mathbb{E}[Y] \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] \text{ by Item (1)}
\end{aligned}$$

□

Definition 2.10

A sequence of random variables X_1, X_2, ..., X_n is independent and identically distributed from $\mathbb{P}_X$ (i.i.d $\sim \mathbb{P}_X$) if

(1) for all i, $X_i \sim \mathbb{P}_x$

(2) X_1, X_2, ..., X_n are mutually independent, i.e., for any $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$

$$\mathbb{P}(X_{i_1} X_{i_2} \ldots X_{i_k}) = \mathbb{P}(X_{i_1})\,\mathbb{P}(X_{i_2}) \ldots \mathbb{P}(X_{i_k})$$

Theorem 2.5 ▶ The Weak Law of Large Numbers (WLLN)

Let X_n be an infinite i.i.d. sequence drawn from $\mathbb{P}_X$. Write

$$\hat{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

and suppose $\mathrm{Var}[X]$ and $\mathbb{E}[X]$ are both finite. Then, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\hat{X}_n - \mathbb{E}[X]\right| < \varepsilon\right) = 1$$

We first show the following two lemmas.

Lemma 2.6 ▶ Markov's Inequality

Let X be any non-negative random variable and $a > 0$. Then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Proof. Define the indicator random variable

$$1_{X\geq a} = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{if } X < a \end{cases}$$

and notice that $\mathbb{E}[1_{X\geq a}] = \mathbb{P}(X \geq a)$. Clearly, $X \geq a 1_{X\geq a}$, hence

$$\mathbb{E}[X] \geq a\,\mathbb{E}[1_{X\geq a}] = a\,\mathbb{P}(X \geq a)$$

from which the result follows. $\square$

Lemma 2.7 ▶ Chebyshev's Inequality

Let X be any random variable with finite variance. Then

$$\mathbb{P}\big(|X - \mathbb{E}[X]| \geq \varepsilon^2\big) \leq \frac{\mathrm{Var}[X]}{\varepsilon}$$

for any $\varepsilon > 0$.

Proof. Set $Y = (X - \mathbb{E}[X])^2$ and notice that $\mathbb{E}[Y] = \mathrm{Var}[X]$. Then,

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) &= \mathbb{P}\big(Y \geq \varepsilon^2\big) \\ &\leq \frac{\mathbb{E}[Y]}{\varepsilon^2} \text{ by Markov's Inequality} \\ &= \frac{\mathrm{Var}[X]}{\varepsilon^2} \end{aligned}$$ $\square$

Now, we prove Theorem 2.5.

Proof. First, notice that

$$\begin{aligned}\mathbb{E}\left[\hat{X}_n\right] &= \mathbb{E}\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] \\ &= \frac{1}{n}\cdot n\,\mathbb{E}[X] \text{ by Linearity of Expectation} \\ &= \mathbb{E}[X]\end{aligned}$$

and

$$\begin{aligned}\operatorname{Var}\left[\hat{X}_n\right] &= \operatorname{Var}\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] \\ &= \frac{1}{n^2}(\operatorname{Var}[X_1] + \operatorname{Var}[X_2] + \cdots + \operatorname{Var}[X_n]) \\ &= \frac{1}{n^2}\cdot n\operatorname{Var}[X] \\ &= \frac{1}{n}\operatorname{Var}[X]\end{aligned}$$

then, by Chebyshev's Inequality,

$$\begin{aligned}\mathbb{P}\left(\left|\hat{X}_n - \mathbb{E}[X]\right| \geq \varepsilon\right) &\leq \frac{\operatorname{Var}\hat{X}_n}{\varepsilon^2} \\ &= \frac{\operatorname{Var}[X]}{n\varepsilon^2} \to 0 \text{ as } n \to \infty\end{aligned}$$

hence

$$\mathbb{P}\left(\left|\hat{X}_n - \mathbb{E}[X]\right| < \varepsilon\right) = 1 - \mathbb{P}\left(\left|\hat{X}_n - \mathbb{E}[X]\right| \geq \varepsilon\right) \to 1 \text{ as } n \to \infty$$

□

Example 2.9 ▶ Bernoulli Random Variable

TBD

Definition 2.11 ▶ Vector Valued Random Variable

Let

$$X = (X_1, X_2, \ldots, X_n) : \Omega \to \mathbb{R}^n$$

3 Entropy

Definition 3.1 ▶ Entropy

Let $\mathbb{P}$ be a probability distribution on a discrete space Ω. The Shannon Entropy (hereby simply Entropy) of $\mathbb{P}$ is defined

$$H(\mathbb{P}) = \sum_{\omega \in \Omega} \mathbb{P}(\omega) \log \frac{1}{\mathbb{P}(\omega)}$$

If X is a discrete random variable, we define

$$\begin{aligned} H(X) &= H(\mathbb{P}_X) \\ &= \sum_{x \in X} \mathbb{P}_X(x) \log \frac{1}{\mathbb{P}_X(x)} \\ &= \mathbb{E}\left[\log \frac{1}{\mathbb{P}_X(X)}\right] \end{aligned}$$

noting that $\log \frac{1}{\mathbb{P}_X(X)}$ is a real random variable.

We can think of $\log \frac{1}{\mathbb{P}_X(x)}$ as the level of "surprise" that $X = x$ occurs and $H(X)$ as the uncertainty or randomness of $\mathbb{P}_X$.

Note that, in Definition 3.1, log refers to $\log_2$, and $\log_2(X)$ is the number of bits of X. Additionally, since a byte is 8 bits, $\log_{256}(X)$ is the number of bytes of X. Additionally, we define $0 \log \frac{1}{0} = 0$, which can be motivated by the fact that

$$\lim_{x \to 0^+} x \log \frac{1}{x} = 0$$

Example 3.1 ▶ Bernoulli Distribution

The Bernoulli Distribution is the discrete random variable

$$\begin{aligned} \mathbb{P}(X = 1) &= p \\ \mathbb{P}(X = 0) &= 1 - p \end{aligned}$$

and has entropy

$$H(X) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}$$

Definition 3.2 ▶ Binary Entropy

The binary entropy of p, $h(p)$, is the entropy of the Bernoulli Distribution with parameter p, i.e.,

$$h(p) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}$$

Notice that $h(0) = h(1) = 0$ and $h\left(\frac{1}{2}\right) = 1$. More generally, the graph of $h(p)$ is given in Figure 1.

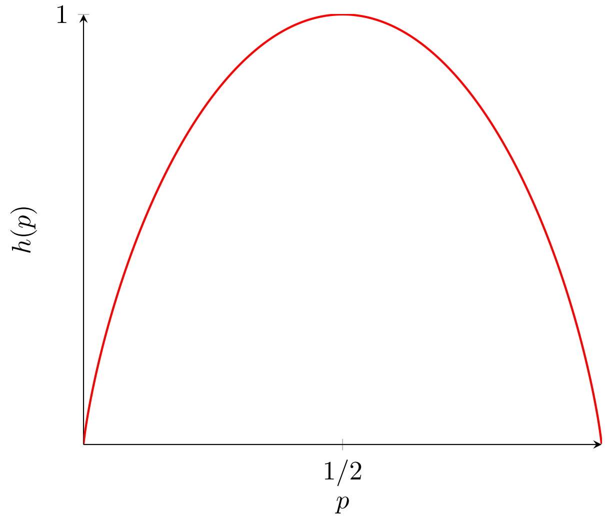


Figure 1: Binary Entropy as a function of p. Notice that the entropy is maximized when $p = 1/2$ and 0 when $p = 0$ or $p = 1$. When $p = 0$ or $p = 1$, the Bernoulli Distribution is non-random, and thus there is no uncertainty.

Figure 2: Drawing of ant nest used to empirically verify ...

Example 3.2 ▶ Geometric Distribution

The Geometric Distribution is the positive, integer-valued random variable that describes the number of Bernoulli trials performed until a success. That is,

$$\mathbb{P}(X = k) = p(1-p)^{k-1}$$

is the probability that it will require k trials until a success.
The entropy of the Geometric Distribution is given by

$$\begin{aligned}
H(X) &= \sum_{k=1}^{\infty} p(1-p)^k \log \frac{1}{p(1-p)^k} \\
&= \sum_{k=1}^{\infty} p(1-p)^k \left(\log \frac{1}{p} + k \log \frac{1}{1-p} \right) \\
&= p \log \frac{1}{p} \sum_{k=1}^{\infty} (1-p)^k + p \log \frac{1}{1-p} \sum_{k=1}^{\infty} k(1-p)^k \\
&= p \frac{1}{p} \log \frac{1}{p} + p \log \frac{1}{1-p} \frac{1-p}{p^2} \\
&= \frac{1}{p} \left(p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \right) \\
&= \frac{h(p)}{p} \to 0 \text{ as } p \to 0^+
\end{aligned}$$

Example 3.3 ▶ Distribution with ∞ Entropy

TBD

An empirical justification for the use of $\log_2$.

Definition 3.3 ▶ Convexity

Let $V \cong \mathbb{R}^n$ be a vector space. A subset $S \subseteq V$ is convex if, for any pair $\mathbf{x}, \mathbf{y} \in S$

$$\lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in S \text{ for all } \lambda \in [0,1]$$

Example 3.4

The following are convex

(1) $\mathbb{R}^n$

(2)

(3)

Definition 3.4 ▶ Convex Function

A function $f : S \to \mathbb{R}$ is

(i) convex if $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in S$ and $\lambda \in [0,1]$

(ii) *strictly* convex if $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in S$ and $\lambda \in [0,1]$

Definition 3.5 ▶ Concave Function

A function $f : S \to \mathbb{R}$ is (strictly) concave if $-f$ is (strictly) convex.

Example 3.5

Notice

(1) The function $x \to x \log x$ is strictly convex

(2) The function $x \to \log x$ is strictly concave

(3) The function $X \to \mathbb{E}[X]$ is convex (but not strictly)

Theorem 3.1 ▶ Jensen's Inequality

Let X be a real vector valued random variable. Then, if f is any convex function,

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

If f is strictly convex, then $f(\mathbb{E}[X]) = \mathbb{E}[f(X)]$ if and only if $X = \mathbb{E}[X]$, i.e., X is a constant random variable.

Proof. Since f is convex,

$$\begin{aligned} f(\mathbb{E}[X]) &= f\left(\sum_{x \in X} x \, \mathbb{P}(X = x)\right) \\ &\leq \sum_{x \in X} f(x) \, \mathbb{P}(X = x) \text{ since } f \text{ is convex and } \mathbb{P}(X = x) \in [0,1] \\ &= \mathbb{E}[f(X)] \end{aligned}$$

□

Theorem 3.2 ▶ Properties of Entropy

The Entropy function satisfies

(1) $H(X) \geq 0$ with equality if and only if X is constant

(2) if $\mathcal{X}$ is finite, then $H(X) \leq \log|\mathcal{X}|$ with equality if and only if $\mathbb{P}_X$ is uniform on $\mathcal{X}$

(3) For any injective f, $H(X) = H(f(X))$

(4) $\mathbb{P} \to H(\mathbb{P})$ is strictly concave

Proof.

(1) $H(X) = \mathbb{E}\left[\log \frac{1}{\mathbb{P}_X}\right] \geq 0$ with equality if and only if $\log \frac{1}{\mathbb{P}_X} = 0$, which occurs only when $\mathbb{P}_X \equiv 1$.

(2) If $\mathcal{X}$ is finite, then

$$\begin{aligned} H(X) &= \mathbb{E}\left[\log \frac{1}{\mathbb{P}_X}\right] \\ &\leq \log \mathbb{E}\left[\frac{1}{\mathbb{P}_X}\right] \\ &= \log \sum_{x \in X} \mathbb{P}(x) \frac{1}{\mathbb{P}(X)} \\ &= \log|\mathcal{X}| \end{aligned}$$

with equality if and only if $\log \frac{1}{\mathbb{P}_x}$ is constant, which forces $\mathbb{P}(X) = \frac{1}{|\mathcal{X}|}$.

(3) If f is injective, then $\mathbb{P}_{f(X)}(f(x)) = \mathbb{P}_X(x)$, and the result follows.

(4) Take $\lambda \in [0,1]$ and write $f(x) = x \log \frac{1}{x}$, then

$$\begin{aligned} H(\lambda \mathbb{P}_1 + (1-\lambda)\mathbb{P}_2) &= \sum_{\omega \in \Omega} f(\lambda\, \mathbb{P}_1(\omega) + (1-\lambda)\, \mathbb{P}_2(\omega)) \\ &\geq \sum_{\omega \in \Omega} \lambda f(\mathbb{P}_1(\omega)) + (1-\lambda) f(\mathbb{P}_2(\omega)) \\ &= \lambda \sum_{\omega \in \Omega} f(\mathbb{P}_1(\omega)) + (1-\lambda) \sum_{\omega \in \Omega} f(\mathbb{P}_2(\omega)) \\ &= \lambda H(\mathbb{P}_1) + (1-\lambda) H(\mathbb{P}_2) \end{aligned}$$

□

4 Conditional Entropy

Definition 4.1 ▶ Joint Entropy

Given random variables X and Y, the *Joint Entropy*, $H(XY)$, is defined

$$H(XY) = \mathbb{E}\left[\log \frac{1}{\mathbb{P}_{XY}}\right] = \sum_{x \in X} \sum_{y \in Y} \mathbb{P}_{XY}(X = x, Y = y) \log \frac{1}{\mathbb{P}_{XY}(X = x, Y = y)}$$

Definition 4.2 ▶ Conditional Entropy

Let X and Y be random variables. Then

$$H(X \mid Y) = \underset{y \sim \mathbb{P}_Y}{\mathbb{E}}\left[H\left(\mathbb{P}_{X|Y=y}\right)\right] = \mathbb{E}\left[\log \frac{1}{\mathbb{P}_{X|Y}}\right]$$

This can be thought of as the expected uncertainty $H\left(\mathbb{P}_{X|Y=y}\right)$ over $y \sim \mathbb{P}_Y$.

Definition 4.3 ▶ Conditional Probability Notation

Some notation:

(1) $\mathbb{P}_{X|Y=y}$ is a distribution on X, with

$$\begin{aligned}\mathbb{P}_{X|Y=y}(x) &= \mathbb{P}(X = x \mid Y = y) \\ &= \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}\end{aligned}$$

(2) $\mathbb{P}_{X|Y}$ is a random variable on $\mathcal{X} \times \mathcal{Y}$ with

$$\mathbb{P}_{X|Y}(x, y) = \mathbb{P}(X = x \mid Y = y)$$

Example 4.1 ▶ Joint and Conditional Entropy of a Fair Die

TBD

Theorem 4.1 ▶ Properties of Conditional Entropy

Let X and Y be random variables. Then

(1) $H(X \mid Y) \leq H(X)$ with equality if and only if X and Y are independent

(2) $H(XY) = H(Y) + H(X \mid Y) \leq H(Y) + H(X)$ with equality if and only if X and Y are independent

(3) $H(XY) \geq \max\{H(X), H(Y)\}$

Proof.

(1)

$$\begin{aligned}H(X \mid Y) &= \underset{y \sim \mathbb{P}_Y}{\mathbb{E}}\left[H\left(\mathbb{P}_{X|Y=y}\right)\right] \\ &\leq H\left(\underset{y \sim \mathbb{P}_Y}{\mathbb{E}}\left[\mathbb{P}_{X|Y=y}\right]\right) \\ &= H(\mathbb{P}_X) \\ &= H(X)\end{aligned}$$

(2)

(3) $H(XY) = H(X) + H(Y \mid X) \geq H(X)$ The same argument shows $H(XY) \geq H(Y)$, hence it must be greater than or equal to the maximum of the two.

$\square$

Corollary 4.2

For any function f

(1) $H(X) = H(Xf(X))$

(2) $H(f(X) \mid X) = 0$

(3) $H(X) \geq H(f(X))$ with equality if and only if f is injective

Proof. $\square$

Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
000	000	Null	032	020	Space	064	040	@	096	060	‘
001	001	Start of Heading	033	021	!	065	041	A	097	061	a
002	002	Start of Text	034	022	”	066	042	B	098	062	b
003	003	End of Text	035	023	#	067	043	C	099	063	c
004	004	End of Transmission	036	024	$	068	044	D	100	064	d
005	005	Enquiry	037	025	%	069	045	E	101	065	e
006	006	Acknowledgement	038	026	&	070	046	F	102	066	f
007	007	Bell	039	027	’	071	047	G	103	067	g
008	008	Backspace	040	028	(	072	048	H	104	068	h
009	009	Horizontal Tab	041	029	)	073	049	I	105	069	i
010	00a	Line Feed	042	02a	*	074	04a	J	106	06a	j
011	00b	Vertical Tab	043	02b	+	075	04b	K	107	06b	k
012	00c	Form Feed	044	02c	,	076	04c	L	108	06c	l
013	00d	Carriage Return	045	02d	-	077	04d	M	109	06d	m
014	00e	Shift Out	046	02e	.	078	04e	N	110	06e	n
015	00f	Shift In	047	02f	/	079	04f	O	111	06f	o
016	010	Data Link Escape	048	030	0	080	050	P	112	070	p
017	011	Device Control 1	049	031	1	081	051	Q	113	071	q
018	012	Device Control 2	050	032	2	082	052	R	114	072	r
019	013	Device Control 3	051	033	3	083	053	S	115	073	s
020	014	Device Control 4	052	034	4	084	054	T	116	074	t
021	015	Negative	053	035	5	085	055	U	117	075	u
022	016	Synchronous Idle	054	036	6	086	056	V	118	076	v
023	017	End of Trans. Block	055	037	7	087	057	W	119	077	w
024	018	Cancel	056	038	8	088	058	X	120	078	x
025	019	End of Medium	057	039	9	089	059	Y	121	079	y
026	01a	Substitute	058	03a	:	090	05a	Z	122	07a	z
027	01b	Escape	059	03b	;	091	05b	[	123	07b	{
028	01c	File Separator	060	03c	<	092	05c	\	124	07c	|
029	01d	Group Separator	061	03d	=	093	05d	]	125	07d	}
030	01e	Record Separator	062	03e	>	094	05e	ˆ	126	07e	~
031	01f	Unit Separator	063	03f	?	095	05f	_	127	07f	

Figure 3: ASCII code for characters.

5 Lossless Data Compression

"Today is a Wednesday" is a sequence of letters, which can be converted into bytes (8 bits) via Figure 3.

One may ask if this is optimal. In fact, if using only English words, then we need only

$$2^5 = 32 < 26 \times 2 = 52 < 64 = 2^6$$

6 bits.

Definition 5.1 ▶ Lossless Compression

Let $\mathcal{X}$ denote some alphabet, and let f and g be functions:

$$\mathcal{X} \xrightarrow[\text{compressor}]{f} \{0,1\}^* \xrightarrow[\text{decompressor}]{g} \mathcal{X}$$

where $\{0,1\}^*$ is the set of all binary strings (including the empty string)[1]. The functions f and g are also often called the *encoder* and *decoder*, respectively.
We say that f and g form a *lossless compression scheme* if $g \circ f \equiv I_{\mathcal{X}}$ is the identity function on the alphabet. For each $x \in \mathcal{X}$, we call $f(x)$ the *code word* or *encoding* of x and refer to the set $f(X)$ as the *code book*.

Definition 5.2 ▶ Length of a Code Word

The length of a code word $\omega \in \{0,1\}^*$ is the number of bits in ω and is denoted $\ell(\omega)$. Note that

$$\ell : \{0,1\}^* \to \mathbb{N}$$

Notice that, given an alphabet $\mathcal{X}$, the maximal length of a compression must be $\log|\mathcal{X}|$, by the pigeonhole principle: enumerate the alphabet of $\mathcal{X}$, say $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$. Then we can map

$$\begin{aligned} x_1 &\to \emptyset \\ x_2 &\to 0 \\ x_3 &\to 1 \\ x_4 &\to 00 \\ &\vdots \end{aligned}$$

and clearly our maximum length is $\log|\mathcal{X}|$. However, given a distribution (a list of frequencies of the occurrences of the alphabet), we can reduce the *expected* code word length.

Example 5.1

Say $\mathcal{X} = \{a, b, c, d\}$. Of course, we can map

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

and our expected codeword length will obviously be 2. On the other hand, say our alphabet has the following frequencies

Character	Frequency
a	1/2
b	1/8
c	1/4
d	1/8

We can map

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

(called a variable length encoding, in constrast to the fixed length encoding given above), and see that our expected codeword length is

$$\frac{1}{2} \cdot 1 + \frac{1}{8} \cdot 3 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 = \frac{7}{4} < 2$$

i.e., we can do better than 2 bits per character.

In general, we determine the frequency of letters empirically, which we then model by a probability distribution. Our objective is to minimize both $\sup \ell(f(\mathcal{X})$ and $\mathbb{E}[\ell(f(\mathcal{X}))]$. In fact, there is an optimal compressor f^* which minimizes both. The core idea is to assign shorter code words to more frequently occurring characters.

[1]The * is called the Kleene Star.

Theorem 5.1 ▶ Optimal Compressor

Let $\mathcal{X}$ be an alphabet. Without loss of generality, say

$$\mathcal{X} = \{1, 2, \ldots, n\}$$

and that $\mathbb{P}_X(i) \geq \mathbb{P}_X(i+1)$, i.e., that the frequencies are in decreasing order. Then

(1) $\ell(f^*(i)) = \lfloor \log i \rfloor$

(2) For all k and any encoder f,

$$\mathbb{P}(\ell(f(x)) \leq k) \leq \mathbb{P}(\ell(f^*(x)) \leq k)$$

Example 5.2 ▶ Telegraph and Morse Code

TBD

First, we show the following lemma

Lemma 5.2

Let Z be a positive integer valued random variable with finite expectation. Then $H(Z) \leq \mathbb{E}[Z] H\left(\frac{1}{\mathbb{E}[Z]}\right)$.

Proof. Let Q_p denote the geometric distribution with parameter p, i.e., Q_p is the distribution with positive integer random variable X given by

$$\mathbb{P}(X = i) = p(1-p)^{i-1}$$

and recall that

$$H(Q_p) = \frac{h(p)}{p}$$

The relative entropy from Q to P, $D(P \parallel Q)$, is given by

$$D(P \parallel Q) = \sum_{\omega \in \Omega} P(\omega) \log \frac{P(\omega)}{Q(\omega)}$$

Notice that

$$\begin{aligned}
D(P \parallel Q) &= \sum_{\omega \in \Omega} P(\omega) \log \frac{P(\omega)}{Q(\omega)} \\
&= \sum_{\omega \in \Omega} P(\omega) \log P(\omega) - \sum_{\omega \in \Omega} P(\omega) \log Q(\omega) \\
&= \sum_{\omega \in \Omega} P(\omega) \log \frac{1}{Q(\omega)} - \sum_{\omega \in \Omega} P(\omega) \log \frac{1}{P(\omega)} \\
&= H(P, Q) - H(P)
\end{aligned}$$

where

$$H(P, Q) = \sum_{\omega \in \Omega} P(\omega) \log \frac{1}{Q(\omega)}$$

is the cross-entropy of P and Q. Further, $D(P \parallel Q) \geq 0$: since log is concave, we must have

$$
\begin{aligned}
D(P \parallel Q) &= \sum_{\omega\in\Omega} P(\omega)\log\frac{P(\omega)}{Q(\omega)} \\
&= -\sum_{\omega\in\Omega} P(\omega)\log\frac{Q(\omega)}{P(\omega)} \\
&\geq -\log\left(\sum_{\omega\in\Omega} P(\omega)\frac{Q(\omega)}{P(\omega)}\right) \\
&= -\log\left(\sum_{\omega\in\Omega} Q(\omega)\right) \\
&\geq -\log 1 \\
&= 0
\end{aligned}
$$

Now, set $p = 1/\mathbb{E}[Z]$[2] and notice that

$$
\begin{aligned}
H(Z, Q_p) &= \sum_{z=1}^{\infty} P(z)\log\frac{1}{p(1-p)^z} \\
&= \sum_{z=1}^{\infty} P(z)\left(\log\frac{1}{p} + z\log\frac{1}{1-p}\right) \\
&= \log\frac{1}{p}\sum_{z=1}^{\infty} P(z) + \log\frac{1}{1-p}\sum_{z=1}^{\infty} zP(Z) \\
&= \log\frac{1}{p} + \log\frac{1}{1-p}\,\mathbb{E}[Z] \\
&= \log\frac{1}{p} + \frac{1}{p}\log\frac{1}{1-p} \\
&= H(Q_p) \\
&= \mathbb{E}[Z]h\left(\frac{1}{\mathbb{E}[Z]}\right)
\end{aligned}
$$

Now, since $D(P \parallel Q) \geq 0$, we conclude that

$$
\begin{aligned}
D(Z \parallel Q_p) &= H(Z, Q_p) - H(Z) \\
&= H(Q_p) - H(Z) \\
&= \mathbb{E}[Z]h\left(\frac{1}{\mathbb{E}[Z]}\right) - H(Z)
\end{aligned}
$$

is greater than or equal to 0, hence

$$
H(Z) \leq \mathbb{E}[Z]h\left(\frac{1}{\mathbb{E}[Z]}\right) \qquad \square
$$

Now, we prove Theorem 5.1.

Proof. $\square$

[2]Since $\mathbb{E}[Z] < \infty$ and Z is positive-valued, it is easy to see that

$$
\mathbb{E}[Z] = \sum_{z=1}^{\infty} z\,\mathbb{P}(Z = z) \geq \sum_{z=1}^{\infty} \mathbb{P}(Z = z) = 1
$$

hence $Q_{1/\mathbb{E}[Z]}$ is a well-defined distribution.

Theorem 5.3 ▶ Optimal Average Code Length

Given $\mathcal{X}$ and $\mathbb{P}_X(1) \geq \mathbb{P}_X(2) \geq \dots$, then

(1) $\mathbb{E}[\ell(f^*(x))] = \sum_{k=0\infty} \mathbb{P}(X \geq 2^k)$

(2) $H(X) - \log(eH(X) + e) \leq \mathbb{E}[\ell(f^*(x))] \leq H(X)$

Proof. Throughout, let $\ell^*(X) = \ell(f^*(X))$. Item (1) is straightforward:

$$\begin{aligned}
\mathbb{E}[\ell^*(X)] &= \mathbb{E}[\lfloor \log X \rfloor] \\
&= \sum_{k=1}^{\infty} \mathbb{P}(\lfloor \log X \rfloor \geq k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(\log X \geq k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}\big(X \geq 2^k\big)
\end{aligned}$$

Now, notice that

$$f^*(X_n) = (f^*(X_1), f^*(X_2), \dots, f^*(X_n))$$

hence

$$\ell^*(X_n) = \sum_{i=1}^{n} \ell^*(X_i)$$

Additionally, we have $\mathbb{P}_X(m) \leq \frac{1}{m}$, since

$$\begin{aligned}
m\,\mathbb{P}_X(m) &\leq \sum_{i=1}^{m} \mathbb{P}_X(i) \\
&\leq 1
\end{aligned}$$

hence

$$\ell^*(m) = \lfloor \log m \rfloor \leq \log \frac{1}{\mathbb{P}_X(m)}$$

Thus, we conclude that

$$\mathbb{E}[\ell^*(X)] \leq \mathbb{E}\left[\log \frac{1}{\mathbb{P}_X(x)}\right] = H(X)$$

For the other side of the inequality, apply Lemma 5.2.

$$\begin{aligned}
H(X) &= H(X \mid L) + H(L) \\
&= \sum \mathbb{P}_L(k) H(X \mid L = k) + h\left(\frac{1}{H(\mathbb{E}[L])}\right)(1 + \mathbb{E}[L]) \\
&\leq \sum \mathbb{P}_L(k) \log 2^k + h\left(\frac{1}{H(\mathbb{E}[L])}\right)(1 + \mathbb{E}[L]) \\
&= \mathbb{E}[L] + \log(1 + \mathbb{E}[L]) + \mathbb{E}[L] \log\left(1 + \frac{1}{\mathbb{E}[L]}\right) \\
&\leq \mathbb{E}[L] + \log(e(1 + H(X)))
\end{aligned}$$

□

Corollary 5.4

If $X = S^n$ is an iid sequence, then

$$nH(S) - \log n + \mathcal{O}(1) \leq \mathbb{E}[\ell^*(S^n)] \leq nH(S)$$

hence

$$\lim_{n \to \infty} \frac{\mathbb{E}[\ell^*(S^n)]}{n} = H(S)$$

i.e., the expected length per message approaches the entropy.

In 2011, Szpankowski and Verdú showed

$$\mathbb{E}[\ell^*(S^n)] = nH(S) - \frac{1}{2}\log n + \mathcal{O}(1)$$

Additionally, by the weak law of large numbers,

$$\frac{\ell^*(S^n)}{n} \to H(S) \text{ in probability}$$

6 Preliminary on Linear Algebra

Definition 6.1

Complex Vector Space A *complex vector space* is a set V together with two binary operations,

$$+ : V \times V \to V$$
$$\cdot : \mathbb{C} \times V \to V$$

such that, for any $\mathbf{u}, \mathbf{v}$ in V and any complex α, β,

1) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (associativity)
2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (commutativity)
3) There exists an element $\mathbf{0} \in V$, called the *zero vector*, such that $\mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$
4) For every element $\mathbf{v} \in V$, there is an element $-\mathbf{v} \in V$, called the *additive inverse* such that $\mathbf{v} + -\mathbf{v} = \mathbf{0}$
5) $\alpha \cdot (\beta \cdot \mathbf{v}) = (\alpha\beta) \cdot \mathbf{v}$
6) $1 \cdot \mathbf{v} = \mathbf{v}$
7) $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$
8) $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$

We call $\alpha\mathbf{u} + \beta\mathbf{v}$ a *linear combination* of $\mathbf{u}$ and $\mathbf{v}$. Additionally, writing a vector in bold, as in $\mathbf{v}$, is often omitted, with the fact that v is a vector determined from context. Similarly, the $\cdot$ is not generally written, so $a \cdot \mathbf{v} = a\mathbf{v}$.

Any n-dimensional vector can be viewed as a column vector, as in

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

and addition and scalar multiplication as

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ n \end{pmatrix} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix}$$
$$\alpha\mathbf{v} = \alpha \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{pmatrix}$$

Definition 6.2 ▶ Span

The *span* of a set of vectors $W \subseteq V$, denoted $\text{span}(W)$, is the set of all linear combinations of vectors in W.

Definition 6.3 ▶ Spanning Set

A subset W of V is a spanning set if $\text{span}(W) = V$, i.e., every vector in V can be written as a linear combination of vectors in W.

Definition 6.4 ▶ Linear Independence

A set of vectors W is *linearly independent* if

$$\alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2 + \cdots = \mathbf{0}$$

implies $\alpha_1 = \alpha_2 = \cdots = 0$.

Definition 6.5 ▶ Basis

A *basis* of a vector space is a linearly independent spanning set.

Definition 6.6 ▶ Homomorphism

A function $\phi : V \to W$ is a vector-space homomorphism if, for any α, β in $\mathbb{C}$ and $\mathbf{u}$, $\mathbf{v}$ in V

$$\phi(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha\phi(\mathbf{u}) + \beta\phi(\mathbf{v})$$

A bijective homomorphism is called an *isomorphism*. If an isomorphism exists between vector spaces V and W, we say that V and W are *isomorphic*.

The *standard basis* is given by

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 2 \\ \vdots \\ 0 \end{pmatrix} \quad \ldots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

Notice that any vector can be written as a linear combination of the standard basis. That is

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + \cdots + u_n\mathbf{e}_n$$

Definition 6.7 ▶ Vector Space of Functions

Let I be a finite set. Then

$$\mathbb{C}^I = \{u : I \to \mathbb{C}\}$$

where $u + v$ is the function given by

$$(u + v)(a) = u(a) + v(a)$$

and αu the function

$$(\alpha u)(a) = \alpha \cdot u(a)$$

Theorem 6.1

The vector space $\mathbb{C}^I$ is isomorphic to $\mathbb{C}^n$ if and only if $|I| = n$.

Proof. Write $I = \{x_1, x_2, \ldots, x_n\}$ and define $e_i : I \to \mathbb{C}$ by

$$e_i(y) = \delta_{x_i y} = \begin{cases} 1 & \text{if } y = x_i \\ 0 & \text{otherwise} \end{cases}$$

It suffices to show that $\{e_i\}$ forms a basis — that this set spans is simple: suppose u is defined by

$$u(x_i) = a_i$$

then

$$u = a_1e_1 + a_2e_2 + \cdots + a_ne_n$$

since each $e_i(x_i) = 1$ and $e_j(x_i) = 0$ for $j \neq i$.

To see that this set is also linearly independent, suppose that

$$a_1 e_1 + a_2 e_2 + \cdots + a_n e_n = 0$$

where here 0 is the function that identically maps I to 0. Then

$$(a_1 e_1 + a_2 e_2 + \cdots + a_n e_n)(x_1) = a_1 e_1(x_1) = 0$$

which forces $a_1 = 0$. Similar reasoning shows that all a_i must be 0, as desired. □

Example 6.1

Let $\mathcal{X} = \{0,1\}^2 = \{00, 01, 10, 11\}$. Then

$$u \in \mathbb{C}^{\mathcal{X}} \leftrightarrow \begin{pmatrix} u(00) \\ u(01) \\ u(10) \\ u(11) \end{pmatrix} \in \mathbb{C}^4$$

Definition 6.8

The inner product on $\mathbb{C}^I$ is the function defined by

$$\langle \cdot, \cdot \rangle : \mathbb{C}^I \times \mathbb{C}^I \to \mathbb{C}$$
$$\langle u, v \rangle = \sum_{i \in I} \overline{u(i)} v(i)$$

where $\overline{a + bi} = a - bi$ for $a, b \in \mathbb{R}$.
The inner product satisfies:

(1) Linearity in second input: $\langle u, \alpha v + \beta w \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle$

(2) Conjugate symmetry: $\langle u, v \rangle = \overline{\langle v, u \rangle}$

(3) Positivity: $\langle u, u \rangle \geq 0$ with equality if and only if $u = 0$

Note that items 1 and 2 above imply anti-linearity in the first input:

$$\langle \alpha u + \beta v, w \rangle = \overline{\alpha} \langle u, w \rangle + \overline{\beta} \langle v, w \rangle$$

Definition 6.9 ▶ Norm and Distance

The *norm* of a vector is given by

$$\|v\| = \sqrt{\langle v, v \rangle}$$

The distance between vectors u and v is defined

$$d(u, v) = \|u - v\| = \sqrt{\sum_{i=1}^{n} |u(i) - v(i)|^2}$$

Theorem 6.2 ▶ Properties of the norm

For all u, v in V and $\alpha \in \mathbb{C}$, the norm satisfies:

1. Positivity: $\|u\| \geq 0$ with equality iff $u = 0$
2. $\|\alpha u\| = |\alpha| \|u\|$
3. $\|u + v\| \leq \|u\| + \|v\|$ (triangle inequality)

Theorem 6.3 ▶ The Cauchy-Schwarz Inequality

For any $u, v \in \mathbb{C}^n$,

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

with equality if and only if $u = \alpha v$ for some $\alpha \in \mathbb{C}$.

The Cauchy-Schwarz Inequality is equivalent to the triangle inequality.
Other excamples of norms include:

$$\|u\|_p = \left(\sum_{i \in I} |u(i)|^p \right)^{\frac{1}{p}}$$

$$\|u\|_\infty = \max_{i \in I} \{u(i)\}$$

Definition 6.10 ▶ Hilbert Space

A *Hilbert Space* is a vector space with the norm

$$\|u\| = \left(\sum_{i \in I} |u(i)|^2 \right)^{\frac{1}{2}} = \|u\|_2$$

Definition 6.11 ▶ Orthogonal Set

A set $\{u_1, u_2, \ldots, u_k\}$ is *orthogonal* if $\langle u_i, u_j \rangle = 0$ for all $i \neq j$. The set is *orthonormal* if all vectors are unit vectors, i.e., $\|u_i\| = 1$ for all i.
A set is an orthonormal basis if it is orthonormal and a basis.

For example, the standard basis e_1, e_2, ..., e_n is an orthonormal basis on $\mathbb{C}^n$. Similarly, where $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$, the set e_{x_1}, e_{x_2}, ..., e_{x_n} is an orthonormal basis on $\mathbb{C}^{\mathcal{X}}$.
The vector space $(\mathbb{C}^n, \langle \cdot, \cdot \rangle)$ is called the n-dimensional complex Hilbert (or Euclidean) Space.

Theorem 6.4

Every n-dimensional complex Hilbert Space is isomorphic to $(\mathbb{C}^n, \langle \cdot, \cdot \rangle)$.

Proof. Since the dimension is n, there is an orthonormal basis of n vectors, v_1, v_2, ..., v_n. The map

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \to \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_n e_n$$

is the desired isomorphism. $\square$

In general, a Hilbert Space is a vector space together with an inner product. A Hilbert space H exhibits "completeness", any series of vectors

$$\sum_{k=0}^{\infty} u_k$$

that converges, i.e., satisfies

$$\sum_{k=0}^{\infty} \|u_k\| < \infty$$

converges in H.

A Hilbert Space can have infinite dimension. For example, the space

$$L_2(\mathbb{R}) = \left\{ \underset{\text{measurable}}{f : \mathbb{R} \to \mathbb{C}} \mid \int |f(x)|^2 < \infty \right\}$$

with inner product

$$\langle f, g \rangle = \int \overline{f} g$$

is an infinite-dimensional Hilbert space.

Definition 6.12 ▶ Linear Operator

Let $\mathcal{V}$, $\mathcal{W}$ be complex vector spaces. A map $L : \mathcal{V} \to \mathcal{W}$ is linear if

$$L(\alpha u + \beta v) = \alpha L(u) + \beta L(v)$$

for all α, β in $\mathbb{C}$ and u, v in $\mathcal{V}$.

We denote by $L(\mathcal{V}, \mathcal{W})$ the set of *all* linear operators from $\mathcal{V}$ to $\mathcal{W}$.
$L(\mathcal{V}, \mathcal{W})$ is a complex vector space with

- addition: $(A + B)u = Au + Bu$
- Scalar multiplication: $(\alpha A)u = \alpha Au$

If $\dim \mathcal{V} = n$ and $\dim \mathcal{W} = m$ with basis

$$\begin{aligned} &\{v_1, v_2, \ldots, v_n\} \text{ of } \mathcal{V} \\ &\{w_1, w_2, \ldots, w_m\} \text{ of } \mathcal{W} \end{aligned}$$

and operator A, we can write

$$\begin{aligned} Av_1 &= a_{11}w_1 + a_{12}w_2 + \cdots + a_{1m}w_m \\ Av_2 &= a_{21}w_1 + a_{22}w_2 + \cdots + a_{2m}w_m \\ &\vdots \\ Av_n &= a_{n1}w_1 + a_{n2}w_2 + \cdots + a_{nm}w_m \end{aligned}$$

In the basis for $\mathcal{W}$, these are just the vectors

$$\begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1m} \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2m} \end{pmatrix} \cdots \begin{pmatrix} a_{n1} \\ a_{n2} \\ \vdots \\ a_{nm} \end{pmatrix}$$

$M = (a_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}}$ is an $m \times n$ comple matrix. The i, j-th entry is given by $\langle w_i, Av_j \rangle$. The space of all $n \times m$ complex matrices is denoted $\mathbb{M}_{n \times m}$ and is isomorphic to $L(\mathbb{C}^n, complexnums^m)$, the space of linear operators.

$$\begin{array}{cccc} \text{matrix} & M & \underset{\{w_j\} \text{ basis of } \mathbb{C}^m}{\overset{\{v_i\} \text{ basis of } \mathbb{C}^n}{\longleftrightarrow}} & \underset{Au}{A} \text{ linear operator} \\ & {\scriptstyle M}\begin{pmatrix} u(1) \\ u(2) \\ \vdots \\ u(n) \end{pmatrix} & & \end{array}$$

For $A \in L(\mathcal{V}, \mathcal{W})$, $B \in L(\mathcal{W}, \mathcal{Z})$, $AB \in L(\mathcal{V}, \mathcal{Z})$, where

$$\begin{aligned} A \circ B(u) &= A(B(u)) \\ A \circ B(\alpha u + \beta v) &= A(B(\alpha u + \beta v)) \\ &= A(\alpha Bu + \beta Bv) \\ &= \alpha ABu + \beta ABv \end{aligned}$$

thus, this map is linear. Note the "$\circ$" is often omitted.
$A \circ B \leftrightarrow MN$

$$(MN)_{ik} = \sum_{j=1}^{m} M_{ij} N_{jk} \underset{1 \le i \le d, \quad 1 \le k \le n}{\text{matrix multiplication}}$$

$$
\begin{tikzcd}
\mathcal{V} \arrow[r, "A"] \arrow[d, leftrightarrow, "\{v_i\}"'] & \mathcal{W} \arrow[r, "B"] \arrow[d, leftrightarrow, "\{w_i\}"] & \mathcal{Z} \arrow[d, leftrightarrow, "\{z_i\}"] \\
\mathbb{C}^n \arrow[r, "M"] & \mathbb{C}^m \arrow[r, "N"] & \mathbb{C}^d
\end{tikzcd}
$$

Basis for $\mathbb{M}_{n\times m}$.

$$E_{i,j}(k,\ell) = \begin{cases} 1 & \text{if } (k,\ell) = (i,j) \\ 0 & \text{otherwise} \end{cases}$$

$$M = \sum_{M(i,j)E_{i,j}}$$

The set $\{E_{ij}\}$ forms a basis for $\mathbb{M}_{n\times m}$.

Basis for $L(\mathcal{V},\mathcal{W})$.

For $v \in \mathcal{V}$, $w \in \mathcal{W}$, $E_{w,v}(u) = \langle v, u\rangle w$.

Given a basis $\{v_i\} \subseteq \mathcal{V}$, $\{w_j\} \subseteq \mathcal{W}$,

$$E_{w_j,w_i}(v_k) = \delta_{ik} w_j = \begin{cases} w_j & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}$$

The set

$$\left\{ E_{w_j,v_i} \mid \begin{array}{c} i = 1,2,\ldots,n \\ j = 1,2,\ldots,m \end{array} \right\}$$

forms a basis for $L(\mathcal{V},\mathcal{W})$.

Thus, $\dim \mathbb{M}_{n\times m} = nm$ and $\dim L(\mathcal{V},\mathcal{W}) = \dim \mathcal{V} \dim \mathcal{W}$.

Definition 6.13 ► Direct Sum of Vector Spaces

Given $\mathcal{V}_1 = \mathbb{C}^{\mathcal{X}_1}$, $\mathcal{V}_2 = \mathbb{C}^{\mathcal{X}_2}$, ..., $\mathcal{V}_n = \mathbb{C}^{\mathcal{X}_n}$, the direct sum $\mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus \mathcal{V}_n$ is defined

$$\mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus \mathcal{V}_n = \left\{ \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \mid v_i \in \mathcal{V}_i \right\}$$

and

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = v_1 \oplus v_2 \oplus \ldots \oplus v_n$$

with addition defined

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} v_1 + u_1 \\ v_2 + u_2 \\ \vdots \\ v_n + u_2 \end{pmatrix}$$

Example 6.2

$\mathbb{C}^2 \oplus \mathbb{C}^4 \oplus \mathbb{C}^3 = \mathbb{C}^{2+4+3} = \mathbb{C}^9$, e.g.

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \oplus \begin{pmatrix} 2 \\ 3 \\ 1 \\ 2 \end{pmatrix} \oplus \begin{pmatrix} -5 \\ 6 \\ 7 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 2 \\ 3 \\ 1 \\ 2 \end{pmatrix} \\ \begin{pmatrix} -5 \\ 6 \\ 7 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 3 \\ 1 \\ 2 \\ -5 \\ 6 \\ 7 \end{pmatrix}$$

Definition 6.14 ▶ Inner Product of Direct Sum

We define the inner product on $\mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus \mathcal{V}_n$ as

$$\left\langle \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \right\rangle = \sqrt{\sum_i \langle u_i, v_i \rangle}$$

If the set $\{e_k^i\}_{k=1}^{|\mathcal{V}_i|}$ is an orthonormal basis of $\mathcal{V}_i$ for all i, then $\{e_k^i\}_{k,i}$ is an orthonormal basis of $\mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \ldots \oplus V_n$. Thus

$$\mathbb{C}^{d_1} \oplus \mathbb{C}^{d_2} \oplus \ldots \oplus \mathbb{C}^{d_n} \cong \mathbb{C}^{d_1+d_2+\cdots+d_n}$$

Definition 6.15 ▶ Direct Sum of Linear Operators

Let

$$\begin{aligned} A_1 &\in L(\mathcal{V}_1, \mathcal{W}_1) \\ A_2 &\in L(\mathcal{V}_2, \mathcal{W}_2) \\ \vdots A_n & \qquad \in L(\mathcal{V}_n, \mathcal{W}_n) \end{aligned}$$

Then $A_1 \oplus A_2 \oplus \ldots \oplus A_n$ is the function in $L(V_n, V_m)$ given by

$$A_1 \oplus A_2 \oplus \ldots \oplus A_n \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} A_1 u_1 \\ A_2 u_2 \\ \vdots \\ A_n u_n \end{pmatrix}$$

If A_1 has matrix M_1, A_2 has matrix M_2, and so on, then

$$A_1 \oplus A_2 \oplus \ldots \oplus A_n \leftrightarrow \begin{bmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_n \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

Definition 6.16 ▶ Tensor Product

Given $\mathcal{V} = \mathbb{C}^{\mathcal{X}}$ and $\mathcal{W} = \mathbb{C}^{\mathcal{Y}}$, the tensor product $\mathcal{V} \otimes \mathcal{W}$ is defined

$$\mathcal{V} \otimes \mathcal{W} = \mathbb{C}^{\mathcal{X} \times \mathcal{Y}} = \{u : \mathcal{X} \times \mathcal{Y} \to \mathbb{C}\}$$

with

$$v \otimes w(a, b) = v(a)w(b)$$

and

$$\begin{aligned} \mathcal{V} \otimes \mathcal{W} &= \operatorname{span}\{v \otimes w \mid v \in \sqsubseteq, w \in \mathcal{W}\} \\ &= \left\{\sum \alpha_i v_i \otimes w_i \mid \alpha_i \in \mathbb{C}, v_i \in \mathcal{V}, w_i \in \mathcal{W}\right\} \\ &\text{all linear combinations of elementary tensors} \end{aligned}$$

Notice that

$$\begin{aligned} v_1 \otimes w + v_2 \otimes w &= (v_1 + v_2) \otimes w \\ v \otimes w_1 + v \otimes w_2 &= v \otimes (w_1 + w_2) \\ \alpha(v \otimes w) &= \alpha v \otimes w = v \otimes \alpha w \\ \langle v_1 \otimes w_1, v_2 \otimes w_2 \rangle &= \langle v_1, v_2 \rangle \langle w_1, w_2 \rangle \end{aligned}$$

If $\{v_i\} \subseteq \mathcal{V}$ and $\{w_i\} \subseteq \mathcal{W}$ are bases, then

$$\begin{aligned} \mathcal{V} \otimes \mathcal{W} &= (a_1 v_1 + a_2 e_2 + \cdots + a_n e_n) \otimes (b_1 w_1 + b_2 w_2 + \cdots + b_m 2_m) \\ &= \sum_{i,j} a_i b_j v_i \otimes w_j \end{aligned}$$

so $\{v_i \otimes w_j\}_{i,j}$ is a basis of $\mathcal{V} \otimes \mathcal{W}$, and $\dim(\mathcal{V} \otimes \mathcal{W}) = \dim(\mathcal{V}) \dim(\mathcal{W})$.
One can similarly define $\mathcal{V}_1 \otimes \mathcal{V}_2 \otimes \ldots \otimes \mathcal{V}_n$.
In general,

$$\begin{aligned} \mathcal{V} \oplus \mathcal{W} &= \mathcal{W} \oplus \mathcal{V} \\ \mathcal{V} \otimes \mathcal{W} &= \mathcal{W} \otimes \mathcal{V} \\ (\mathcal{V} \oplus \mathcal{W}) \otimes \mathcal{Z} &= \mathcal{V} \otimes \mathcal{Z} \oplus \mathcal{W} \otimes \mathcal{Z} \\ \mathcal{V} \otimes \mathcal{W} \otimes \mathcal{Z} &\cong (\mathcal{V} \otimes \mathcal{W}) \otimes \mathcal{Z} \\ \mathcal{V} \oplus \mathcal{W} \oplus \mathcal{Z} &\cong (\mathcal{V} \oplus \mathcal{W}) \oplus \mathcal{Z} \end{aligned}$$

Definition 6.17 ▶ Tensor Product Operator

For $A \in L(\mathcal{V}_1, \mathcal{W}_1)$ and $B \in L(\mathcal{V}_2, \mathcal{W}_2)$, the tensor product of A and B is defined

$$\begin{aligned} A \otimes B &\in L(\mathcal{V}_1 \otimes \mathcal{V}_2, \mathcal{W}_1 \otimes \mathcal{W}_2) \\ A \otimes B(v \otimes w) & Av \otimes Bw \end{aligned}$$

Equivalently,

$$M_{n_1 \times m_1} \otimes M_{n_2 \times m_2} = M_{n_1 n_2 \times m_1 m_2}$$

7 Spectral Theorem

Recall the correspondence

$$
\begin{array}{rl}
L(\mathbb{C}^n, \mathbb{C}^n) \overset{\text{basis}}{\longleftrightarrow} & \mathbb{M}_{m\times n} \\
A & (a_{ij})_{i,j} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \\
A+B & (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})_{i,j} \\
AB & (a_{ij})(b_{j\ell}) = \left(\sum a_{ij} b_{j\ell}\right)_{i,\ell} \\
L(\mathbb{C}^n, \mathbb{C}^n) = B(\mathbb{C}^n) \overset{\text{basis}}{\longleftrightarrow} & \mathbb{M}_{n\times n} = \mathbb{M}_n \\
\text{Operator on } \mathbb{C}^n & \text{square matrix}
\end{array}
$$

For $A, B \in B(\mathbb{C}^n) \cong \mathbb{M}_n$, we can define

1) Addition: $A + B$

   $A + B = B + A$ commutative

   $0u = \mathbf{0}$ 0 operator such that $A + 0 = 0 + A = A$

2) Multiplication: AB

   $AB \neq BA$ non-commutative, e.g.,
   $$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$$
   $$\begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$$

   $Iu = u$ identity operator such that $AI = AI = A$, in matrix form $I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$

3) Adjoint A^*

   $\langle A^* v, u\rangle = \langle v, Au\rangle$, in matrix form if $A = \begin{bmatrix} & a_{ij} & \end{bmatrix}$ then $A^* = \begin{bmatrix} & \overline{a_{ji}} & \end{bmatrix}$

   Thus $(A^*)^* = A$ and $A^* = \overline{A^T}$ is the conjugate transpose of A and $(AB)^* = B^* A^*$ and $(A+B)^* = B^* + A^* = A^* + B^*$, e.g. $\begin{bmatrix} 1+i & 2-i \\ 4 & 3 \end{bmatrix}^* = \begin{bmatrix} 1-i & 4 \\ 2+i & 3 \end{bmatrix}$

In the above, Items 1) and 2) form an *algebra* and Items 1) to 3) form a **-algebra*.
$B(\mathbb{C}^n)$ is the algebra of (linear) operators on $\mathbb{C}^n$
$\mathbb{M}_n$ is the algebra of $n \times n$ complex matrices
Other examples of algebras include function algebras, e.g., $\{u : \Omega \to \mathbb{C}\}$ (which is not commutative, i.e., $f \circ g \neq g \circ f$).
Operators / Matrices with special Properties

1. A is *self-adjoin* if $A = A^*$

   e.g. $A = \begin{bmatrix} 1 & 3+i \\ 3-i & 2 \end{bmatrix}$

   equivalently, if $\langle v, Av\rangle \in \mathbb{R}$ for all $v \in \mathbb{C}^n$, since

   $$\begin{aligned} \overline{\langle v, Av\rangle} &= \langle Av, v\rangle \\ &= \langle A^* v, v\rangle \\ &= \langle v, Av\rangle \end{aligned}$$

2. A is *positive* if $A = B^*B$ for some B

   e.g. $B = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$ and $B^*B = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}$

   equivalently, if $\langle v, Av\rangle \geq 0$ for all $v \in \mathbb{C}^n$, since

$$\begin{aligned}\langle v, Av\rangle &= \langle v, B^*Bv\rangle \\ &= \langle Bv, Bv\rangle \\ &\geq 0\end{aligned}$$

   e.g. $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, then

$$\begin{aligned}\left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{pmatrix} a \\ b \end{pmatrix}\right\rangle &= \left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} a \\ 2b \end{pmatrix}\right\rangle \\ &= \overline{a}a + 2\overline{b}b \\ &= |a|^2 + 2|b|^2 \\ &\geq 0\end{aligned}$$

   Similarly, if $A = \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}$ then

$$\begin{aligned}\left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}\begin{pmatrix} a \\ b \end{pmatrix}\right\rangle &= 5|a|^2 + \overline{a}b + a\overline{b} + |b|^2 \\ &= 4|a|^2 + \left(|a|^2 + conjab + a\overline{b} + |b|^2\right) \\ &= 4|a|^2 + |a+b|^2 \\ &\geq 0\end{aligned}$$

   Notice that $(B^*B)^* = B^*(B^*)^* = B^*B$, hence if $A = B^*B$ is positive, A is self-adjoint.

3. P is a projection if $P = P^* = P^2$

   e.g. $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ or $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ or $P(u) = \langle v, u\rangle v$ for some unit-vector v (this is called the rank-one projection).

   P is a projection if and only if, for all v, $\langle Pv, (1-P)v\rangle = 0$. $v = Pv + (I-P)v$, $PV \perp (1-P)v$.

   If $\mathcal{V} \subseteq \mathbb{C}^n$ is a subspace (i.e., for any $u, v \in \mathcal{V}$, $\alpha u + \beta v \in \mathcal{V}$), then there is a projection $P_\mathcal{V}$ such that $\min_{v \in P_\mathcal{V}} \|u - v\| = \|u - P_\mathcal{V} u\|$

   We say $\mathcal{V} \perp \mathcal{W}$ if for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$, $\langle v, w\rangle = 0$ (orthogonal). Equivalently, if $P_\mathcal{V} P_\mathcal{W} = 0$. $\mathcal{V} \leq \mathcal{W}$ (i.e., $\mathcal{V}$ is a subspace of $\mathcal{W}$) if and only if $P_\mathcal{V} P_\mathcal{W} = P_\mathcal{V}$.

4. U is unitary if $U^*U = U^* = I$. e.g., $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $Y = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$, $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, the Pauli Matrices, are unitary.

   U is unitary if and only if $\langle Uv, Uw\rangle = \langle v, w\rangle$, since

$$\langle Uv, Uw\rangle = \langle v, U^*Uw\rangle = \langle v, w\rangle$$

   Change of basis: given an orthonormal basis $\{v_i\} \subseteq \mathbb{C}^n$ then $\{Uv_i\}$ is also an orthonormal basis, since

$$\langle Uv_i, Uv_j\rangle = \langle v_i, v_j\rangle = 0$$

   In particular, given any two bases $\{v_i\}$ and $\{w_i\}$, there is a unique unitary operator U such that $Uv_i = w_i$ for all i.

   If U is unitary and maps the standard basis

$$e_i \to v_i$$

then U^* is unitary and maps

$$v_i \to e_i$$

with $U^* = U^{-1}$.

Definition 7.1 ▶ Diagonalizable Operator

An operator is *diagonalizable* if there exists an orthonormal basis $\{v_i\}$ such that

$$\langle v_j, Av_i \rangle = \lambda_i \delta_{ij} = \begin{cases} \lambda_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

A matrix $A = (a_{ij})$ is diagonalizable if there exists a unitary U such that

$$UAU^* = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

is a diagonal matrix.

Theorem 7.1 ▶ Spectral Theorem

A is diagonalizable if and only if $A^*A = A^*A$. In particular,

$$A = \sum \lambda_i E_i$$

where $\lambda_i \in \mathbb{C}$ and $\{E_i\}$ is a set of mutually orthogonal projections such that

$$\sum E_i = I$$

An operator A that satisfies $AA^* = A^*A$ is called a *normal* operator.

Definition 7.2 ▶ Eigenvalue

The value λ is an *eigenvalue* of an operator A if there exists a vector v such that $Av = \lambda v$. The *eigenspace* of an eigenvalue λ is

$$V_\lambda = \{v \mid Av = \lambda v\}$$

We define the finite set

$$\operatorname{spec}(A) = \{\lambda \mid \ \lambda \text{ is an eigenvalue of } A\}$$

Notice that E_λ is a projection onto V_λ.
If $A^*A = AA^*$, then $A = \sum_{\lambda \in \operatorname{spec}(A)} \lambda E_\lambda$, and

$$UAU^* = \begin{bmatrix} \lambda_1 E_1 & & & \\ & \lambda_2 E_2 & & \\ & & \ddots & \\ & & & \lambda_n E_N \end{bmatrix} = \begin{bmatrix} \ddots & & & & & & & & \\ & \lambda_1 & & & & & & & \\ & & \lambda_2 & & & & & & \\ & & & \ddots & & & & & \\ & & & & \lambda_2 & & & & \\ & & & & & \ddots & & & \\ & & & & & & \lambda_n & & \\ & & & & & & & \ddots & \\ & & & & & & & & \lambda_n \end{bmatrix}$$

Example 7.1

$$\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} = 1E_1 + (-2)E_2$$

has eigenvectors

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Similarly,

$$\begin{aligned}
\begin{bmatrix} 1 & -2i \\ 2i & -2 \end{bmatrix} &= \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{2} & 1 \\ 1 & 2 \end{bmatrix} \\
&= -3 \begin{bmatrix} \frac{1}{5} & \frac{-2i}{5} \\ \frac{2i}{5} & \frac{4}{5} \end{bmatrix} + 2 \begin{bmatrix} \frac{4}{5} & \frac{2i}{5} \\ \frac{-2i}{5} & \frac{4}{5} \end{bmatrix}
\end{aligned}$$

Note that a real matrix can have complex eigenvalues. For example

$$\operatorname{spec} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \{i, -i\}$$

7.1 Functional Calculus

For a normal operator A, we have

$$A = U^* D_u U = \sum \lambda_i E_i$$

where

$$D_u = \begin{bmatrix} u_1 & & & \\ & u2 & & \\ & & \ddots & \\ & & & u_n \end{bmatrix}$$

for $u_i \in \operatorname{spec}(A)$.

For a function $f : \operatorname{spec}(A) \to \mathbb{C}$, define

$$f(A) = U^* \begin{bmatrix} f(u_1) & & & \\ & f(u2) & & \\ & & \ddots & \\ & & & f(u_n) \end{bmatrix} U = \sum f(\lambda_i) E_i$$

This can be justified as follows. Notice that

$$\begin{aligned}
A^k &= \underbrace{A \cdot A \cdot \dots \cdot A}_{k \text{ times}} \\
&= \underbrace{(U^* D_u U)(U^* D_u U) \dots (U^* D_u U)}_{k \text{ times}} \\
&= U^* \underbrace{D_u (UU^*) D_u (UU^*) \dots (UU^*) D_u}_{k \text{ times}} U \\
&= U^* D_u^k U
\end{aligned}$$

Thus, for polynomial $f(x) = a_0x^k + a_1x^{k-1} + \cdots + a_{k-1}X + a_k$,

$$\begin{aligned}
f(A) &= a_0A^k + a_1A^{k-1} + \cdots + a_{k-1}A + a_k \\
&= a_0U^*D_u^kU + a_1U^*D_u^{k-1}U + \cdots + a_{k-1}U^*D_uU \\
&= U^*a_0D_u^kU + U^*a_1D_u^{k-1}U + \cdots + a_{k-1}U^*a_1D_uU \\
&= U^*\big(a_0D_u^k + a_1D_u^{k-1} + \cdots + a_{k-1}D_u + a_k\big)U \\
&= U^*\left(\begin{bmatrix} a_0u_1^k & & & \\ & a_0u_2^k & & \\ & & \ddots & \\ & & & a_0u_n^k \end{bmatrix} + \begin{bmatrix} a_1u_1^{k-1} & & & \\ & a_0u_2^{k-1} & & \\ & & \ddots & \\ & & & a_0u_n^{k-1} \end{bmatrix} + \cdots + \begin{bmatrix} a_0 & & & \\ & a_0 & & \\ & & \ddots & \\ & & & a_0 \end{bmatrix}\right)U \\
&= U^*\begin{bmatrix} f(u_1) & & & \\ & f(u2) & & \\ & & \ddots & \\ & & & f(u_n) \end{bmatrix}U
\end{aligned}$$

Now, for a general function $f : \mathrm{spec}(A) \to \mathbb{C}$, there exists a polynomial sequence P_k that converges to f uniformly. Then

$$P_k(A) \to f(A)$$

and

$$U^*\begin{bmatrix} P_k(u_1) & & & \\ & P_k(u2) & & \\ & & \ddots & \\ & & & P_k(u_n) \end{bmatrix}U \to U^*\begin{bmatrix} f(u_1) & & & \\ & f(u2) & & \\ & & \ddots & \\ & & & f(u_n) \end{bmatrix}U$$

Example 7.2

Let $f(x) = e^x$, then

$$(A) = e^A = \sum_{n=0}^{\infty} \frac{1}{n!}A^n$$

Taking $A = \begin{bmatrix} & t \\ t & \end{bmatrix}$,

$$e^A = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}$$

Similarly, taking $f(x) = \sqrt{x}$ and $A \geq 0$, $f(A) = \sqrt{A}$ is the unique positive operator such that $A^2 = A$. Taking $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, we have

$$\sqrt{A} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

We can also define $|A| = \sqrt{A^*A}$.

zath
For any $f, g : \mathrm{spec}(A) \to \mathbb{C}$, $f(A)g(A) = fg(A) = g(A)f(A)$.
We can similarly define $f(t) = \log(t)$ and $f(t) = t\log(t)$ on positive operators

$$\log A = U^* \begin{bmatrix} \log u_1 & & & \\ & \log u_2 & & \\ & & \ddots & \\ & & & \log u_n \end{bmatrix} U$$

$$A \log A = U^* \begin{bmatrix} u_1 \log u_1 & & & \\ & u_1 \log u_2 & & \\ & & \ddots & \\ & & & u_n \log u_n \end{bmatrix} U$$

In general $AB \neq BA$. However, if A and B are diagonal, then $AB = BA$. More specifically, $AB = BA$ if and only if there exists a unitary operator U such that $A = UD_1U^*$ and $B = UD_2U^*$ for some diagonal matrices D_1 and D_2.

8 Postulates of Quantum Mechanics

Direct's Bra-cket Notation.

Given a vector $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \in \mathbb{C}^n$, we write $|u\rangle$, prounounced "ket" u.

For $v \in (\mathbb{C})^* = L(\mathbb{C}^n, \mathbb{C})$, We denote by $\langle v|$, prounounced "bra" v,

$$v = \begin{pmatrix} \overline{v_1} & \overline{v_2} & \dots & \overline{v_n} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}^*$$

Notice that

$$\begin{aligned} \langle u||v\rangle &= \begin{pmatrix} \overline{u_1} & \overline{u_2} & \dots & \overline{u_n} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \\ &= \sum_{i=1}^{n} \overline{u_i} v_i \\ &= \langle u, v\rangle \end{aligned}$$

hence we denote it $\langle u|v\rangle$.

Similarly,

$$\underset{L(\mathbb{C},\mathbb{C}^n)}{|u\rangle^*} = \underset{L(\mathbb{C}^n,\mathbb{C})}{\langle u|}$$

Finally, write

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad |2\rangle = \begin{pmatrix} 1 \\ 2 \\ \vdots \\ 0 \end{pmatrix} \dots$$

How it works:

1) Inner Product: $\langle u|v\rangle$

2) Matrix Action: $\langle u|A|v\rangle = \langle u, Av\rangle$

3) Rank one operator: $E_{u,v} = \underset{\text{outer product}}{|u\rangle\langle v|}$ with

$$\begin{aligned} E_{u,v}(|w\rangle) &= (|u\rangle\langle v|)|w\rangle \\ &= |u\rangle\langle v|w\rangle \\ &= \langle v|w\rangle|v\rangle \end{aligned}$$

4) General Operator: $A = \sum a_i |u_i\rangle\langle v_i|$

   e.g.

$$\begin{aligned} A &= \begin{bmatrix} a_{ij} \end{bmatrix} \\ &= \sum_{ij} a_{ij} |i\rangle\langle j| \end{aligned}$$

   Notice that $\{|i\rangle\}$ are the elements of the standard basis.

5) Matrix multiplication:

$$
\begin{aligned}
A &= \sum a_i |u_i\rangle\langle v_i| \\
B &= \sum b_j |w_j\rangle\langle x_j| \\
AB &= \sum_{i,j} a_i b_j |u_i\rangle\langle v_i|w_j\rangle\langle x_j| \\
&= \sum_{i,j} a_i b_j \langle v_i|w_j\rangle |u_i\rangle\langle x_j|
\end{aligned}
$$

Notice that $\langle v_i|w_j\rangle \in \mathbb{C}$.

6) Diagonal operator: $A = \sum \lambda_i |v_i\rangle\langle v_i|$ with eigenvalues λ_i and eigenvectors $|v_i\rangle$ of A.
orthogonal decomposition

Recall the Spectral Theorem

1. A positive $\implies \lambda_i \geq 0$
2. A Hermitian $\implies \lambda_i \in \mathbb{R}$
3. A projection $\implies \lambda_i \in \{0, 1\}$
4. A unitary $\implies |\lambda_i| = 1$ for all i

Postulates are working assumptions under a framework or model (see Feynmann's vide on difference between mathematics and physics https://www.youtube.com/watch?v=obCjODeoLVw).

$$
\begin{tikzcd}
\text{Real World} \arrow[r, "\text{Postulates}"] & \text{Quantum Mechanics using Mathematics} \\
\text{Experiments/Theory} \arrow[u, phantom, ""] &
\end{tikzcd}
$$

Postulate 1. (State Space) Any isolated (quantum) system is associated a $\mathbb{C}$ Hilbert Space as its *state space*. The state of a system is given by a unit vector $|\phi\rangle$ with $\||\phi\rangle\| = 1$, called the *state vector*.

For example, $\mathbb{C}^2 = \{a|0\rangle + b|1\rangle\}$.

The basis $|0\rangle, |1\rangle$ is called the *logic* basis. Another basis is $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$, $|-\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$.

Superposition: $|\phi\rangle = a|0\rangle + b|1\rangle$, unit/state $\iff |a|^2 + |b|^2 = 1$.

$|\phi\rangle = e^{i\delta} \cos\frac{\theta}{2}|0\rangle + e^{i(\delta+\psi)} \sin\frac{\theta}{2}|1\rangle$ describes the Bloch Sphere, with $0 \leq \delta \leq 2\pi$, $0 \leq \theta \leq \pi$, and $0 \leq \psi \leq 2\pi$. e^{δ} is the global phase, not physically relevant. ψ is the relative phase, physical.

Example: 1-dim wave function

Postulate 2. (Measurement) Quantum Information Form — every quantum measurement is given by a collection $\{E_m\}$ of measurement operators satisfying the completeness equation $\sum E_m = I$ with $E_m \geq 0$. The set $\{E_m\}$ is called a Positive Operator-Valued Measurement (POVM).

The probability of seeing an outcome m is given by

$$\langle\phi|E_m|\phi\rangle$$

with

$$\sum_m \langle\phi|E_m|\phi\rangle = 1$$

for any ϕ. Notice this implies $\sum E_m = I$.

Pose measurement state:

$$\frac{E_m|\phi\rangle}{\|E_m|\phi\rangle\|} = fracE_m|\phi\rangle\sqrt{\langle\phi|E_m|\phi\rangle}$$

Example: Shr odinger's Cat

Set $|\phi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$, $E_0 = |0\rangle\langle 0|$, $E_1 = |1\rangle\langle 1|$. After measurement, $\frac{1}{2}$ prob, $E_0|\phi\rangle = \frac{1}{\sqrt{2}}|0\rangle$, $E_1|\phi\rangle = \frac{1}{\sqrt{2}}|1\rangle$.

A *projection value measurement* (PVM) is a set of mutually orthogonal projections $\{P_m\}$ such that $\sum P_m = 1$. For example, taking $\{|v_i\rangle\}$ to be some orthonormal basis, the set $\{E_i = |v_i\rangle\langle v_i|\}$ is a PVM.

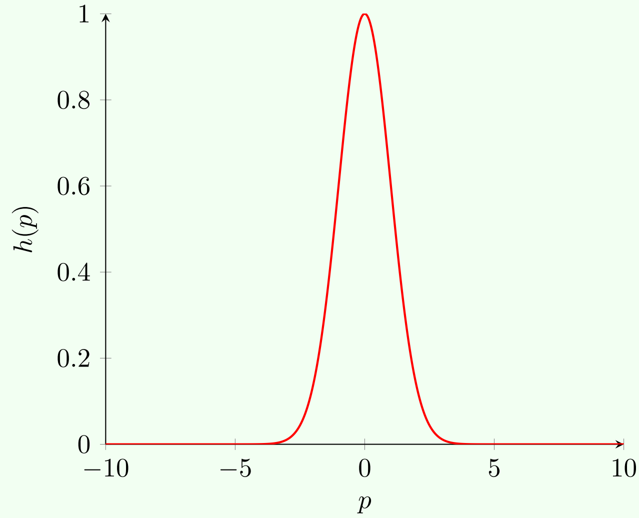
Quantum Mechanics version: every observable corresponds to a Hermitian operator (inf-dim). The only values that will be observed are eigenvalues.

Example 8.1 ▶ Position Operator

The position operator is defined

$$\phi(x) \to x\phi(x)$$

For example, $e^{-\frac{x^2}{2}} \to xe^{-\frac{x^2}{2}}$ has expected position $\langle\phi|X|phi\rangle = \int_{-\infty}^{\infty} X\big|\phi(x)^2\big|dx$



We can similarly define the momentum operator $P = -i\hbar\frac{\partial}{\partial x}$, where $\hbar$is the Planck Constant, and the kinetic energy operator

$$H = -\frac{\hbar^2}{2m}\frac{\partial}{\partial x^2}$$

where m is the mass.

In finite dimensional space, $A = A^*$ (Hermitian) implies $A = \sum \lambda_i E_I$ with $\{E_i\}$ a PVM. If $|\phi_i\rangle$ is an eigenvector of λ_i, then $\langle\phi_i|A|\phi_i\rangle = \lambda_i$, hence all λ_i are real.

Postulate 3. (Evolution) The evolution of a closed quantum system is described by a unitary, $|\phi\rangle \to U|\phi\rangle$.

For example,

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad X|0\rangle = |1\rangle \quad X|1\rangle = |1\rangle \text{ bit flip}$$

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad Z|0\rangle = |0\rangle \quad Z|1\rangle = -|1\rangle \text{ phase flip}$$

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H|0\rangle = |+\rangle \quad H|1\rangle = |-\rangle \text{ Hadamard gate}$$

Why unitary? $|\phi\rangle \to |\psi\rangle = U|\phi\rangle$ so $\|U|\phi\rangle\| = \||\phi\rangle\|$ for all $|\phi\rangle$.

Postulate 3.1 The time evolution of a closed system is given by Schr odinger's Equation

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

$\langle\psi|H|\psi\rangle$ is the expected energy.

1) $H(t)$ time dependent

2) $H(t) \equiv H$ time independent

$|\psi(t)\rangle = e^{iHt}|\psi(0)\rangle$

H hermitian $\implies$ e^{iHt} is unitary

For example: $H = \sum_E E|\phi_E\rangle\langle\phi_E|$ $\quad |\phi_e\rangle$ energy basis

$e^{iHt}|E\rangle = e^{iEt}|E\rangle \quad |\phi\rangle = \sum a_E|E\rangle \quad e^{iHt} = \sum a_E e^{iEt}|\phi_e\rangle$

Postulate 4. (Composite System) The state of a composite system is a tensor product of the state space

Definition 8.1 ▶ Tensor Product Space

Given $V = \mathbb{C}^{\mathcal{X}}$ and $W = \mathbb{C}^{\mathcal{Y}}$,

$$
\begin{aligned}
V \otimes W = \mathbb{C}^{\mathcal{X}\times\mathcal{Y}} &= \{\phi \mid \mathcal{X}\times\mathcal{Y} \to \mathbb{C}\} \\
\phi \otimes \psi(x,y) &= \phi(x)\psi(y) \\
&|\phi\rangle \otimes |\psi\rangle \\
V \otimes W &= \operatorname{span}\{|\phi\rangle \otimes |\psi\rangle \mid |\phi\rangle \in V, |\psi\rangle \in W\} \\
\alpha(|\phi\rangle \otimes |\psi\rangle) &= \alpha|\phi\rangle \otimes |\psi\rangle = |\psi\rangle \otimes \alpha|\psi\rangle \\
\alpha|\phi_1\rangle \otimes |\phi\rangle + \beta|\phi_2\rangle \otimes |\psi\rangle &= (\alpha|\phi_1\rangle + \beta|\phi_2\rangle) \otimes |\psi\rangle \\
(\bar{\phi_1} \otimes \bar{\phi_2})(|\psi_1\rangle \otimes |\psi_2\rangle) &= \langle\phi_1|\psi_1\rangle\langle\phi_2|\psi_2\rangle
\end{aligned}
$$

Given $\{|v_i\rangle\} \subseteq \mathcal{V}$, $\{|w_j\rangle\} \subseteq \mathcal{W}$ orthonormal bases, the set $\{|v_i\rangle \otimes |w_j\rangle \mid 1 \le i \le n, 1 \le j \le m\}$ is an orthonormal basis of $\mathcal{V} \otimes \mathcal{W}$. So $\dim(\mathcal{V} \otimes \mathcal{W}) = \dim\mathcal{V}\dim\mathcal{W}$

Example 8.2 ▶ Product State

Take $V_1 = V_2 = \mathbb{C}^2$. Then

$$
\begin{aligned}
|0\rangle \otimes |0\rangle &= |00\rangle \\
|0\rangle \otimes |1\rangle &= |01\rangle \\
|1\rangle \otimes |0\rangle &= |10\rangle \\
|1\rangle \otimes |1\rangle &= |11\rangle
\end{aligned}
$$

And

$$
\begin{aligned}
|\phi\rangle &= a|0\rangle + b|1\rangle \\
|\psi\rangle &= \alpha|0\rangle + \beta|1\rangle \\
|\phi\rangle \otimes |\psi\rangle &= a\alpha|00\rangle + a\beta|01\rangle + b\alpha|10\rangle + b\beta|11\rangle
\end{aligned}
$$

Example 8.3 ▶ Entangled State

$$
\begin{aligned}
|\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle) \\
&= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\
|\Phi^-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\
|\Psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \\
|\Psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)
\end{aligned}
$$

The vectors $|\Phi^+\rangle$, $|\Phi^-\rangle$, $|\Psi^+\rangle$, $|\Psi^-\rangle$ form an orthonormal basis of $\mathbb{C}^2 \times \mathbb{C}^2 \cong \mathbb{C}^4$. The state is called *entangled* because $|\Psi^+\rangle \neq |\phi\rangle \otimes |\psi\rangle$ for any $|\phi\rangle$, $|\psi\rangle$, i.e., it cannot be expressed as a product state.

Operations on Tensor Product System

$$A \in B(\mathbb{C}^n) \quad B \in B(\mathbb{C}^m)$$

Define $(A \otimes B)|\phi\rangle \otimes |\psi\rangle = A|\phi\rangle \otimes B|\psi\rangle$.

$$A \otimes B \in B(\mathbb{C}^n \otimes \mathbb{C}^m) = \operatorname{span}\{A \otimes B \mid A \in B(\mathbb{C}^n), B \in B(\mathbb{C}^m)\}$$

Definition 8.2 ▶ Product Measurement

$$A \in B(\mathbb{C}^n) \quad B \in B(\mathbb{C}^m)$$

$$\begin{aligned} (A \otimes B)^* = A^* \otimes B^* = A \otimes B &\text{ is Hermitian in } B(\mathbb{C}^n \otimes \mathbb{C}^m) \\ \langle\phi| \otimes \langle\psi|(A \otimes B)|\phi\rangle|\psi\rangle &= \langle\phi|A|\phi\rangle\langle\psi|B|\psi\rangle \end{aligned}$$

$\langle\phi|A|\phi\rangle$ is the expectation of observable A given ϕ and $\langle\psi|B|\psi\rangle$ is the expectation of observable B given ψ.

Given POVMs E_i and F_j, we have

$$\begin{aligned} \sum E_i &= I_1 \\ \sum F_j &= I_2 \\ \sum E_i \otimes F_j &= I_1 \otimes I_2 \end{aligned}$$

and $E_i \geq 0$, $F_j \geq 0$, and $E_i \otimes F_j \geq 0$.

Definition 8.3 ▶ Partial Measurement

Given $A \in B(\mathbb{C}^n)$,

$$\begin{aligned} \langle\phi| \otimes \langle\psi|A \otimes I|\phi\rangle \otimes |\psi\rangle &= \langle\phi|A|\phi\rangle\langle\psi|\psi\rangle \\ &= \langle\phi|A|\phi\rangle \end{aligned}$$

or

$$\sum_m E_m = I \quad \langle\phi|\langle\psi|(E_m \otimes I)|\psi\rangle|\phi\rangle = \langle\phi|E_m|\phi\rangle$$

$$\begin{aligned} \text{Product State} &\sim \text{independent measurement outcome} \\ \text{Entangled State} &\sim \text{correlated measurement outcome} \end{aligned}$$

Example 8.4

Given

$$
\begin{aligned}
E_0 &= |0\rangle\langle 0| \\
E_1 &= |1\rangle\langle 1| \\
E_+ &= |+\rangle\langle +| \\
E_- &= |-\rangle\langle -| \\
\left|\Phi^+\right\rangle &= \frac{|00\rangle + |11\rangle}{\sqrt{2}}
\end{aligned}
$$

Then

$$
\begin{aligned}
\left\langle \Phi^+ \middle| E_0 \otimes I \middle| \Phi^+ \right\rangle &= \frac{1}{2} = \left\langle \Phi^+ \middle| I \otimes E_0 \middle| \Phi^+ \right\rangle \\
\left\langle \Phi^+ \middle| E_1 \otimes I \middle| \Phi^+ \right\rangle &= \frac{1}{2} = \left\langle \Phi^+ \middle| I \otimes E_1 \middle| \Phi^+ \right\rangle \\
\left\langle \Phi^+ \middle| E_i \otimes E_j \otimes I \middle| \Phi^+ \right\rangle &= \begin{cases} \frac{1}{2} & \text{if } i = j \text{, not independent on } i, j \\ 0 & \text{if } i \neq j \end{cases} \\
\left\langle \Phi^+ \middle| E_i \otimes E_\pm \middle| \Phi^+ \right\rangle &= 0 \text{ for } i = 1, 2, \ldots \\
\left\langle \Phi^+ \middle| A \otimes I \middle| \Phi^+ \right\rangle &= \frac{1}{2}\langle 0|A|0\rangle + \frac{1}{2}\langle 1|A|1\rangle \\
&\neq \langle \phi|A|\phi\rangle \forall |\phi\rangle \in \mathbb{C}^n
\end{aligned}
$$

The observation of $|\Phi^+\rangle \in \mathbb{C}^n \otimes \mathbb{C}^n$ on the first system does not match any vector state. Does this violate postulate 1?

No, because a *closed system* has a vector state $|\phi\rangle$. The state of a general system is given by a *density operator*.

A vector state $|\phi\rangle \in \mathbb{C}^n$ is also called a *pure* state.

In general, a quantum system can have mixed state, given by $\{p_i, |\psi_i\rangle\}$, an *ensemble* or pure states. p_i is the probability the system is in $|\psi_i\rangle$. Then

$$
P = \sum p_i |\psi_i\rangle\langle\psi_i|
$$

is the *density operator*.

Example 8.5

If $P = |\phi\rangle\langle\phi|$ ...

Postulate 1 (State) The state of a quantum system is completely determined by a density operator P acting on the state space of the system

$$
P = \sum p_i |\psi_i\rangle\langle\psi_i|
$$

if the system is of probability p_i in the pure state $|\psi_i\rangle$.

Postulate 2 (Measurement) A POVM $\{E_m\}$ has probability of

$$
\mathbb{P}(m) = \mathrm{tr}(P E_m)
$$

to be outcome E_m.

(Observable) The expected value of an observable $A = A^*$ in a given state P is $\mathrm{tr}(PA)$.

Postulate 3 (Evolution) The unitary evolution of a closed quantum system is given by

$$
P \to U P U^*
$$

9 Density Matrix/Operator

Composite System $\mathbb{C}^m \otimes \mathbb{C}^n$, $|\phi\rangle \in \mathbb{C}^m \otimes \mathbb{C}^n$ a vector state. Do a partial measurement $A = A^* \in B(\mathbb{C}^m)$, $\langle\phi|A \otimes I|\phi\rangle$, e.g.,

$$\begin{aligned}
\left|\Phi^+\right\rangle &= \frac{|00\rangle + |11\rangle}{\sqrt{2}} \\
\left\langle\Phi^+\middle|A \otimes I\middle|\Phi^+\right\rangle &= \frac{\langle 0|A|0\rangle}{2} + \frac{\langle 1|A|1\rangle}{2} \\
&\stackrel{?}{=} \langle\phi|A|\phi\rangle \text{ for some } |\phi\rangle \in \mathbb{C}^2
\end{aligned}$$

No, for any $\alpha|1\rangle + \beta|0\rangle = |\phi\rangle$,

$$\begin{aligned}
\langle\phi|A|\phi\rangle &= |\alpha|^2\langle 0A|0\rangle| + |\beta|^2\langle 1|A|1\rangle + \alpha\overline{\beta}\langle 1|A|0\rangle + \beta\alpha\langle 0|A|1\rangle \\
&= |\alpha|^2 a_{00} + |\beta|^2 a_{01} + \alpha\overline{\beta}a_{10} + \beta\alpha a_{11}
\end{aligned}$$

Then what is the state for $\langle\Phi^+|A \otimes I|\Phi^+\rangle = \frac{1}{2}\langle 0|A|0\rangle + \frac{1}{2}\langle 1|A|1\rangle$?

In general, the state of a quantum system can be described by $\{p_i, |\psi_i\rangle\}$, an ensemble of pure states, where p_i is the probability system in $|\psi_i\rangle$, e.g.,

$$\left\langle\Phi^+\middle|A \otimes I\middle|\Phi^+\right\rangle = \frac{1}{2}\langle 0|A|0\rangle + \frac{1}{2}\langle 1|A|1\rangle \rightsquigarrow \left\{\left(\frac{1}{2}, |0\rangle\right), \left(\frac{1}{2}, |1\rangle\right)\right\}$$

A vector space $\{|\phi\rangle\}$ corresponds to the single ensemble $\{1, |\phi\rangle\}$.

Given an ensemble $\{p_i, |\phi_i\rangle\}$, then

$$\begin{aligned}
\text{expected value of } A = A^* &: \sum_i p_i\langle\phi_i|A|\phi_i\rangle \\
\text{POVM}\{E_m\} &: \sum_i p_i\langle\phi_i|E_m|\phi_i\rangle \text{ probability of outcome } m \\
\text{Unitary Transformation} &: \{p_i, |\phi_i\rangle\} \xrightarrow{U} \{p_i, U|\phi_i\rangle\}
\end{aligned}$$

However, in terms of measurement, the ensemble representation is not unique. For any obversable $A = A^*$,

$$\frac{1}{2}\langle +|A|+\rangle + \frac{1}{2}\langle -|A|-\rangle = \frac{1}{2}\langle 0|A|0\rangle + \frac{1}{2}\langle 1|A|1\rangle$$

So, from physics measurement, we can not distinguish

$$\left\{\left(\frac{1}{2}, |0\rangle\right), \left(\frac{1}{2}, |1\rangle\right)\right\}\left\{\left(\frac{1}{2}, |+\rangle\right), \left(\frac{1}{2}, |-\rangle\right)\right\}$$

What is really unique here is the notation of state (in mathematics).

$$\phi : B(\mathbb{C}^n) \to \mathbb{C}, \phi(A) = \left\langle\Phi^+\middle|A \otimes I\middle|\Phi^+\right\rangle$$

ϕ is linear, so $\phi \in L(B(\mathbb{C}^n), \mathbb{C}) = B(\mathbb{C})^*$.

Recall that, for a vector space $\mathcal{V}$, the dual space is $\mathcal{V}^* = L(\mathcal{V}, \mathbb{C})$.

Example 9.1

Take

$$\mathcal{V} = \mathbb{C}^n = \left\{\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \mid u_i \in \mathbb{C}\right\}$$

Then

$$\mathcal{V}^* \cong \mathbb{C}^n = \left\{\begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} \mid v_i \in \mathbb{C}\right\}$$

$$\begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} : \mathbb{C}^n \to \mathbb{C}$$

$$\begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \sum_{i=1}^{n} v_i u_i \text{ linear functional}$$

Given $e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 2 \\ \vdots \\ 0 \end{pmatrix}$, ..., $e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$ basis, there exists $\{e_i^*\} \subseteq \mathcal{V}^*$, a dual basis, such that

$$e_i^*(e_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Indeed, $e_i^* = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$.

Theorem 9.1

If $\dim \mathcal{V} < \infty$, then $\mathcal{V} \cong \mathcal{V}^*$.

Now, consider $\mathcal{V} = B(\mathbb{C}^n) \cong \mathbb{M}_n$. What is $\mathcal{V}^*$?
Recall the trace functional: $\operatorname{tr}((a_{ij})) = \sum aii$ or, equivalently, $\operatorname{tr}(A) = \sum_i \langle e_i | A | e_i \rangle$
The trace is

1) Independent of Basis
2) Satisfies the *tracial property*

Theorem 9.2 ► Tracial Property

For all A, B in $\mathbb{M}_n(\mathbb{C})$ and unitary U

1) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$
2) $\operatorname{tr}(U^* A U) = \operatorname{tr}(A)$
3) $\operatorname{tr}(A) = \sum_i \langle \phi_i | A | \phi_i \rangle$ for any orthonormal basis $\{|\phi_i\rangle\}$.

Proof.

1) Write $A = (a_{ij})$, $B = (b_{ij})$. Then

$$AB = \left(\sum_k a_{ik} b_{kj} \right)_{ij}$$

$$BA = \left(\sum_\ell b_{i\ell} a_{\ell j} \right)_{ij}$$

□

Lemma 9.3

$\mathbb{M}_n \cong \mathbb{M}_n^*$ by the following bijection: $f \in \mathbb{M}_n^* \leftrightarrow$ operator X_f such that $f(A) = \operatorname{tr}(AX)$

To see this in an elementary way, consider basis

$$\{E_{ij} = |i\rangle\langle j|\} \subseteq \mathbb{M}_n$$

and dual basis

$$\{f_{ij}(e_{k\ell})\} = \delta_{(ij)(k\ell)}$$

then

$$f_{ij} \leftrightarrow E_{ji} \in \mathbb{M}_n$$

i.e., there is a one-to-one correspondence between f_{ij} and E_{ji}. Then

$$\begin{aligned} f_{ij}(A) &= \operatorname{tr}(A\langle j||i\rangle) \\ &= \operatorname{tr}\Big(\sum a_{k\ell}|k\rangle\langle\ell|(|j\rangle\langle i|)\Big) \\ &= a_{ij} \end{aligned}$$

Thus, $\operatorname{span}\{E_{ji}\} = \mathbb{M}_n = \mathbb{M}_n^*$.
Now, for $|\phi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^n$, $\phi(A) = \langle\phi|A \otimes I|\phi\rangle$ corresponds to an operator P such that $\phi(A) = \operatorname{tr}(AP)$.
The operator ϕ has the following properties:

1) if $A \geq 0$ and $A \otimes I \geq 0$, then $\phi(A) \geq 0$
2) $\phi(I) = \langle\phi|I \otimes I|\phi\rangle = \langle\phi|\phi\rangle = 1$

A linear function that satisfies Items 1) and 2) is called a seate (sp?).
What property should the operator P have?

1) $\operatorname{tr}(PA) = \langle\phi|A \otimes I|\phi\rangle \geq 0 \implies P \geq 0$ (choose $A = |h\rangle\langle h|$)
2) $\operatorname{tr}(PI) = \operatorname{tr}(P) = 1$

$P \geq 0 \implies P = \sum P_i|\psi_i\rangle\langle\psi_i|$, $P_i \geq 0$ (by orthogonal decomposition)
$\operatorname{tr}(P) = 1 \implies \sum P_i = 1$ (by basis independence of trace)
A state of a quantum system $\mathbb{C}^n$ can be equivalently described by one of the following

1. An ensemble of pure states $\{p_i, |\psi_i\rangle\}$ with $\sum p_i = 1$, $p_i \geq 0$, $|\psi_i\rangle \in \mathbb{C}^n$
2. A density operator $P \in B(\mathbb{C}^n)$ such that $P \geq 0$ and $\operatorname{tr}(P) = 1$
3. A linear functional $\phi : B(\mathbb{C}^n) \to \mathbb{C}$ with $\phi(1) = 1$ and $\phi(A) \geq 0$ if $A \geq 0$
4. A state vector $|\phi\rangle \in \mathbb{C}^n \times \mathbb{C}^m$ for some m

Examples:

1. Pure state = vector state: $|\phi\rangle \leftrightarrow \phi(A) = \langle\phi|A|\phi\rangle \leftrightarrow P = |\phi\rangle\langle\phi|$ density operator
2. A mixed state $P\ sump_i|\phi_i\rangle\langle\phi_i|$ with $\{|\phi_i\rangle\}$ orthogonal and $\sum p_i = 1$, $p_i \geq 0$
   Mixed state are convex combination of pure state. If $p_i = 1$ and $p_j = 0$ for all $j \neq i$, then $P = |\phi_i\rangle\langle\phi_i|$ pure state.
   For $\mathbb{C}^n$, $p_i = \frac{1}{n}$, $P = \sum \frac{1}{n}|\phi_i\rangle\langle\phi_i| = \frac{1}{n}I$, completely mixed state (like uniform distribution $\begin{pmatrix}\frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n}\end{pmatrix}$)
3. Hat state: P projection, then

$$P = \frac{P}{\operatorname{tr}(P)} = \frac{1}{\operatorname{tr}(P)} \sum_{i=1}^{\operatorname{tr}(P)} |\phi_i\rangle\langle\phi_i|$$

   for orthonormal basis $\{|\phi_i\rangle\}$
4. Ensemble of pure states

$$\{p_i, |\phi_i\rangle\} \to P = \sum p_i|\phi_i\rangle\langle\phi_i|$$

   e.g. $\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2} = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -|$

Definition 9.1 ▶ Product State

Let $\rho \in D(\mathbb{C}^n)$ and $\sigma \in D(\mathbb{C}^m)$, then $\rho \otimes \sigma \in D(\mathbb{C}^{n\times m})$.

Since $\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B)$, $\operatorname{tr}(\rho \otimes \sigma) = \operatorname{tr}(\rho)\operatorname{tr}(\sigma) = 1$. Then $\rho \geq 0$, $\sigma \geq 0$ implies $\rho \otimes \sigma \geq 0$.

9.1 Join States/Density Operator

Denote $H_A = \mathbb{C}^n$ and $H_B = \mathbb{C}^m$. A density operator $\rho_{ab} \in B(H_A \otimes H_B)$ is called a joint density operator (states over a joint system) AB.

Example 9.2

With $\rho \in D(H_A)$ and $\sigma \in D(H_B)$

1. $\rho \otimes \sigma$, with

$$\begin{aligned}\rho &= \frac{1}{4}|0\rangle\langle 0| + \frac{3}{4}|1\rangle\langle 1| \\ &= \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{3}{4} \end{bmatrix} \\ \sigma &= \frac{1}{3}|+\rangle\langle +| + \frac{2}{3}|-\rangle\langle -1| \\ &= \begin{bmatrix} \frac{1}{2} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{2} \end{bmatrix}\end{aligned}$$

then

$$\rho \otimes \sigma = \begin{bmatrix} \frac{1}{4} \cdot \sigma & 0 \cdot \sigma \\ 0 \cdot \sigma & \frac{3}{4} \cdot \sigma \end{bmatrix} = \begin{bmatrix} \frac{1}{8} & -\frac{1}{24} & 0 & 0 \\ -\frac{1}{24} & \frac{1}{8} & 0 & 0 \\ 0 & 0 & \frac{3}{8} & \frac{1}{8} \\ 0 & 0 & -\frac{1}{8} & \frac{3}{8} \end{bmatrix}$$

2. Separable state $w = \sum \lambda_i \rho_i \otimes \sigma_i$ with $\sum \lambda_i = 1$ and $\lambda_i \geq 0$