Lecture 1. Overview

Where do we come from?

Information Theory studies the processing, storage and communication of information.

1948. Claude Shannon

"A Mathematical Theory of Communication"

What is information? How can we transmit it?

Shannon entropy

A lot of citations. importance in engineering

data compression (ZIP files)

Communication (error correction)

cryptography (encryption)

statistics (data analysis)

Joseph Doob. "It is not clear that author's mathematical intentions are honorable."

2001. Shannon Award

⋮

Active research area

Quantum Mechanic

1900 Max Planck: Black-body radiation

thermal electromagnetic wave

energy is radiated in discrete package "quanta"

Latin "How much"

Einstein, Bohr

1920s Heisenberg, Born, Jordon, Schrödinger

Matrix Mechanic:

1930s Hilbert Paul Dirac John von Neumann

Hilbert space Mathematic Foundations of Quantum Mechanics

von Neumann Entropy

⋮

Today: Quantum Physics: a major branch of Morden physics.

The world is quantum.

Wilde Book

Interaction: Information Theory & Quantum Mechanic

Quantum Information Theory

1950s - 1970s: Mathematics works on Entropies on Quantum systems

1970s: Information transmission via coherent lasers

Alexander. Holevo. (2017 Shannon award)

other important theoretical work

1980s Richard Feynman: Computing with quantum mechanical model for stimulating quantum systems.

1990s Increased activities and intereses

Peter Shor: Quantum algorithm for prime factorization

$4801 \times 35317 = 169556917$

Breaks RSA encryption

after 2000s: expontial growth research on Quantum information

Where we are at Today: Quantum information Science and engineering.

A major task: Build a quantum computer

two leading player: IBM 127 quatum bits (qubits)

Google 72 qubits

"achieved some computional task that can not be done by the best current classical computer in the life of universe" Recently channelged.

Future: Where are wegoing (Probably)?

IBM & Google expect to build "useful" quantum computer in this decade.

Where should we start?

Next time: Probability Theory

Entropy. What is "bit"? The first half of Shannon's 1948 paper.

A crash course on Probability: (Not a replacement for a proper text book)

A discrete probality space (Ω, P) is given by

- A finite set or a countable set Ω
  e.g. $\{a,b,c,d\}$ $\{1,2,3,4,\cdots\}$
- A probality mass function $P: \Omega \to [0,1]$ s.t.
  ① $\forall w \in \Omega$, $P(w) \geq 0$
  ② $\sum_{w \in \Omega} P(w) = 1$

For $w \in \Omega$, $P(w)$: the probality the case w happen

An <u>event</u> A is a subset $A \subseteq \Omega$.

$$P(A) = \sum_{w \in A} P(w)$$

$P: \Omega \to [0,1]$ induce a probability distribution $P: 2^{\Omega} \to [0,1]$ (also denoted by P.)

$\{A \mid A \subseteq \Omega\}$ = 2^{Ω}

① if $A \cap B = \emptyset$ $P(A \cup B) = P(A) + P(B)$

② $P(\Omega) = 1$

p.m.f $\longleftrightarrow$ p.d.

$P(w)$ $P(\{w\})$

Example 1: Rolling a fair die 

$\Omega = \{1,2,3,4,5,6\}$ $P(i)$ = Probality of face i happens

$P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$

Event A = {outcome is even}

= $\{2,4,6\}$

$P(A) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}$

Let $A, B \subseteq \Omega$.

The condition probability $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Example: A fair die $\Omega = \{1, 2, 3, 4, 5, 6\}$

$A = \{\omega \text{ is even}\}$ $B = \{\omega \geq 4\}$ $P(A) = \frac{1}{2} = P(B)$

$P(A \cap B) = P(\{4, 6\}) = \frac{1}{3}$

$P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$

Bayes' Rule $P(B|A) = \dfrac{P(A|B) \cdot P(B)}{P(A)}$

pf: $P(B|A)\,P(A) = P(A \cap B) = P(A|B) \cdot P(B)$

Two events A and B are independent if

$P(A \cap B) = P(A) \cdot P(B)$ $(\Leftrightarrow P(A|B) = P(A))$

Example. Flip a fair coin twice.

$\Omega = \{HH, HT, TT, TH\}$

$A = \{\text{first outcome is } H\}$ $P(A) = \frac{1}{2}$ $P(B) = \frac{1}{2}$

$B = \{\text{second } \_\_ \text{ is } T\}$ $P(A \cap B) = P(\{HT\}) = \frac{1}{4}$

A Random variable is a function $X: \Omega \to \mathcal{X}$ from a prob. space (Ω, P) to a target space $\mathcal{X}$

X is discrete if $\mathcal{X}$ is discrete. (We always in this case as $X(\Omega)$ is discrete)

$\mathcal{X} = \{x_1, x_2, \cdots\}$ is called the alphabet of X

X induce a distribution on $\mathcal{X}$

$\forall x \in \mathcal{X}, P_X(x) = P(\{w \mid X(w) = x\})$

In many cases, $(\mathcal{X}, P_X)$ capture all the information we need from RV X.

(or law)

$X \sim P_X$ means X has distribution P_X on $\mathcal{X}$.

Example, X be the rank of a poker card randomly picked from a 52-card deck

$\Omega = \{$ all cards in a 52-card deck $\}$ $\quad P(w) = \frac{1}{52}$ $\forall w \in \Omega$

$X: \Omega \to \mathcal{X} = \{2, 3, \cdots, 10, J, Q, K, A\}$

$P_X(A) = P_X(2) = \cdots = \frac{1}{13}$

$X: \Omega \to \mathcal{X}$, $Y: \Omega \to \mathcal{Y}$ two random variables

Joint distribution on $\mathcal{X} \times \mathcal{Y}$: $P_{XY}(X=x, Y=y) = P(\{X(w)=x, Y(w)=y\})$

$A \subseteq \mathcal{X}$, $B \subseteq \mathcal{Y}$ $\quad P_{XY}(X \in A, Y \in B) = P(\{X(w) \in A, Y(w) \in B\})$

P_{XY} is a distribution on the produce space $(\mathcal{X} \times \mathcal{Y}, P_{XY})$

Example. A fair die $\Omega = \{1, 2, 3, 4, 5, 6\}$

$X(w) = \begin{cases} \text{large if } w \geq 4 \\ \text{Small if } w \leq 3 \end{cases}$

$Y(w) = \begin{cases} \text{Even} & \text{if } w \text{ even} \\ \text{odd} & \text{if } w \text{ odd} \end{cases}$

P_{XY} (Large & Even) $= P(\{w \text{ even}, w \geq 4\}) = P(\{4,6\}) = \frac{1}{3}$

P_{XY} (B & Odd) $= P(\{5\}) = \frac{1}{6}$

P_{XY} (Small & E) $= P(\{2\}) = \frac{1}{6}$

P_{XY} (S & O) $= P(\{1,3\}) = \frac{1}{3}$

Example. Flip a fair coin twice. X : outcome of first flip

Y : - - - second - -

$X = \{H, T\}$ $Y = \{H, T\}$ $X \times Y = \{HH, TH, HT, TT\}$

$P_X(H) = P_X(T) = \frac{1}{2} = P_Y(H) = P_Y(T)$

$P_{XY}(HH) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = P_{XY}(HT)$ - - -

Two R.V. X and Y are independent if $P_{XY}(X=x, Y=y) = P_X(x) P_Y(y)$

$\Leftrightarrow P_{XY}(X \in A, Y \in B) = P_X(A) P_Y(B)$

Product prob. (Ω_1, P_1) (Ω_2, P_2). two prob. spaces.

$P_1 \times P_2 (A \times B) = P_1(A) P_2(B)$ $A \subseteq \Omega_1$, $B \subseteq \Omega_2$

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

Prop. X, Y independent $\Leftrightarrow$ $P_{XY} = P_X \times P_Y$ product prob.

Example. Randomly pick one from 52-card deck

X : the rank, Y the type

$X : \Omega \to X = \{2, 3, \cdots 10, J, Q, K, A\}$

$Y: \Omega \to \mathcal{Y} = \{ \underset{S}{\text{spade}}, \underset{H}{\text{heart}}, \underset{C}{\text{club}}, \underset{D}{\text{diamond}} \}$

$P_{XY}(\diamondsuit 10) = \frac{1}{52} = P_X(X=10)\, P_Y(Y=\diamondsuit) = \frac{1}{13} \times \frac{1}{4}$

Real Random variables.

A Real R.V. is a function $X: \Omega \to \mathbb{R}$.

e.g. The height of a random person., Value of a die

$X :=$ the rank of poker card is real R.V.

if we identify A=1 J=11. Q=12. K=13

$Y :=$ type of poker card is not

In the discrete case: if $X: \Omega \to \mathcal{X}$ is a R.V.

the prob. distribution $P_X: \mathcal{X} \to [0,1]$ is a Real R.V.

For two R.V. $X: \Omega \to \mathcal{X}$

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

one can define $P_{X|Y}: \mathcal{X} \times \mathcal{Y} \to [0,1]$ a Real R.V.

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

What is special of Real random variables?

Given $X, Y: \Omega \to \mathbb{R}$

We can define: $X+Y$, $X \cdot Y$, $f(X)$ as real R.V.

where $f: \mathbb{R} \to \mathbb{R}$ is a real function.

Expection and Variance.

Let $X: \Omega \to \mathcal{X} \subseteq \mathbb{R}$ be a discrete real R.V.

① Expectation (or mean)

$$\mathbb{E}X = \sum_{x \in \mathcal{X}} x \, P_X(x) = \sum_{w \in \Omega} X(w) P(w)$$

② Variance

$$\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}|X - \mathbb{E}X|^2$$

Example: $X :=$ Value of a fair die $\quad X : \Omega \to \{1,2,3,4,5,6\}$

$$\mathbb{E}X = \sum_{j=1}^{6} \frac{1}{6} j = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

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

$$= \sum_{j=1}^{6} \frac{1}{6} j^2 - (3.5)^2 = \frac{91}{6} - (3.5)^2 = \frac{35}{12}$$

Prop. Let $X, Y : \Omega \to \mathbb{R}$.

① $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$, $\quad \mathbb{E}(cX) = c\,\mathbb{E}X \quad c \in \mathbb{R}$

$\mathrm{Var}(cX) = c^2 \mathrm{Var}(X)$

If X, Y are independent

② $\mathbb{E}(X \cdot Y) = (\mathbb{E}X)(\mathbb{E}Y)$

③ $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Pf: $\mathbb{E}(X+Y) = \sum_{w \in \Omega} (X(w)+Y(w))P(w) = \sum_{w \in \Omega} X(w)P(w) + \sum_{w \in \Omega} Y(w)P(w)$

$= \mathbb{E}X + \mathbb{E}Y$

② $\mathbb{E}(X \cdot Y) = \sum_{x \in X, y \in Y} xy\, P_{XY}(X=x, Y=y) = \sum xy\, P_X(X=x) \cdot P_Y(Y=y)$ (independence)

$= \left(\sum_x x P_X(X=x)\right)\left(\sum_y y P_Y(Y=y)\right)$

$= \mathbb{E}X\ \mathbb{E}Y$

③ $\mathrm{Var}(X+Y) = \mathbb{E}(X+Y)^2 - (\mathbb{E}(X+Y))^2$

$= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}X + \mathbb{E}Y)^2$

$= \mathbb{E}X^2 + \mathbb{E}Y^2 + 2\mathbb{E}XY - \left[(\mathbb{E}X)^2 + 2\mathbb{E}X \cdot \mathbb{E}Y + (\mathbb{E}Y)^2\right]$

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

Law of large number

A sequence of R.V. $X_1, X_2, \cdots X_n$ i.i.d $\sim P_X$ if

① $\forall i$ $X_i \sim P_X$

② $X_1 \cdots X_n$ mutually independent

e.g. $(X_1, X_3, X_6): \Omega \to \mathbb{R}^3$

is independent to (X_2, X_4, X_7)

Thm (Weak L.L.N)

Let $X_i, i \in \mathbb{N}$ be an infinite i.i.d sequence subject to P_X.

Denote $\overline{X_n} = \frac{1}{n}(X_1 + \cdots + X_n)$. Suppose $Var(X)$ and $\mathbb{E}X < +\infty$

For $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|\overline{X_n} - \mathbb{E}X| < \varepsilon) = 1$$

Chebyshev's Inequality

$$P(|X - \mathbb{E}X| > \varepsilon) \le \frac{Var(X)}{\varepsilon^2}$$

Pf: $Var(X) = \mathbb{E}|X - u|^2 \ge \sum_{|X-u|>\varepsilon} |X-u|^2 P(x) + \sum_{|X-u|<\varepsilon} |X-u| P(x)$

($u = \mathbb{E}X$)

$$\ge \sum_{|X-u|>\varepsilon} \varepsilon^2 P(x) \ge \varepsilon^2 P(|X-u|>\varepsilon)$$

Pf of Weak L.L.N: $\mathbb{E}\overline{X_n} = \mathbb{E}\frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}\mathbb{E}X_1 + \cdots \mathbb{E}X_n$

$$= \frac{1}{n} \cdot n \mathbb{E}X = \mathbb{E}X$$

$$Var(\overline{X_n}) = Var(\frac{1}{n}(X_1 + \cdots X_n))$$

$$= \frac{1}{n^2} Var(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2}\mathrm{Var}(X_1) + \cdots \mathrm{Var}(X_n)$$

$$= \frac{1}{n^2} \cdot n\,\mathrm{Var}(X) = \frac{1}{n}\mathrm{Var}(X)$$

Then $P(|\overline{X_n} - \mathbb{E}X| \geq \varepsilon) = P(|\overline{X_n} - \mathbb{E}\overline{X_n}| > \varepsilon) \leq \dfrac{\mathrm{Var}(\overline{X_n})}{\varepsilon^2}$

$$= \frac{\mathrm{Var}(X)}{n\varepsilon^2} \to 0$$

So $P(|\overline{X_n} - \mathbb{E}X| < \varepsilon) = 1 - P(|\overline{X_n} - \mathbb{E}X| \geq \varepsilon) \to 1$ □

$X \in \{0,1\}$

Example: A Bernoulli R.V. has distribution

$P_X(X=1) = p \quad P_X(X=0) = 1-p$

$\mathbb{E}X = p \quad \mathrm{Var}(X) = p(1-p) \quad \lim_{n\to\infty} \frac{1}{n}(X_1 + \cdots X_n) = p$ almost surely.

Vector valued R.V.: $\mathbb{R}^n = \{(a_1, a_2 \cdots a_n) \mid a_i \in \mathbb{R}\}$

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

has $\mathbb{E}X$, $\mathrm{Var}(X)$, L.L.N as above

Def: Let P be a prob. distribution on a discrete Ω. The (Shannon) entropy

$$H(P) := \sum_{w\in\Omega} P(w) \log \frac{1}{P(w)}$$

For a discrete R.V. $X: \Omega \to \mathcal{X}$

$$H(X) := H(P_X) = \sum_{x\in\mathcal{X}} P_X(x) \log \frac{1}{P_X(x)} = \mathbb{E}\left(\log \frac{1}{P_X(x)}\right)$$

(real R.V.)

$\log \frac{1}{P_X(x)}$: the suprisal of $X=x$ happens, $H(X)$: the uncertainty/Randomness of P_X.

Rem 1. Basis of log

$\log_2 \longleftrightarrow$ bits $\qquad \log_{256} \longleftrightarrow$ bytes $=$ 8 bits

2. We agree $0 \log \frac{1}{0} = 0$ by $\lim_{x\to 0} x \log \frac{1}{x} = 0$.

Example (Bernoulli): $X \in \{0,1\}$. $P(X=1)=p$ $\quad P(X=0)=1-p$

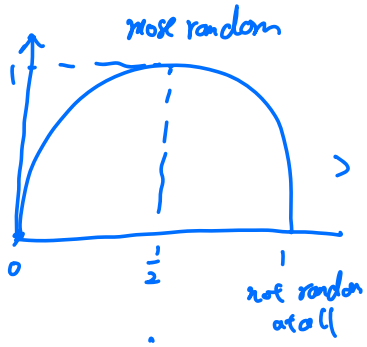
$$H(X) = p\log\frac{1}{p} + (1-p)\log\frac{1}{(1-p)} := h(p)$$

where $h(\cdot)$ is called binary entropy function

$h(\frac{1}{2}) = 1 \qquad h(0) = h(1) = 0$

In $\log_2$ basis, $\quad h(p) \le 1$ and

$h(p) = 1$ iff $p = \frac{1}{2}$



Example (Geometric): $X \in \{0,1,2,\dots\}$ $\qquad P(X=i) = p(1-p)^i$ $\quad i = 0,1,2\cdots$

$$H(X) = \sum_{i=0}^{\infty} p(1-p)^i \log \frac{1}{p(1-p)^i}$$

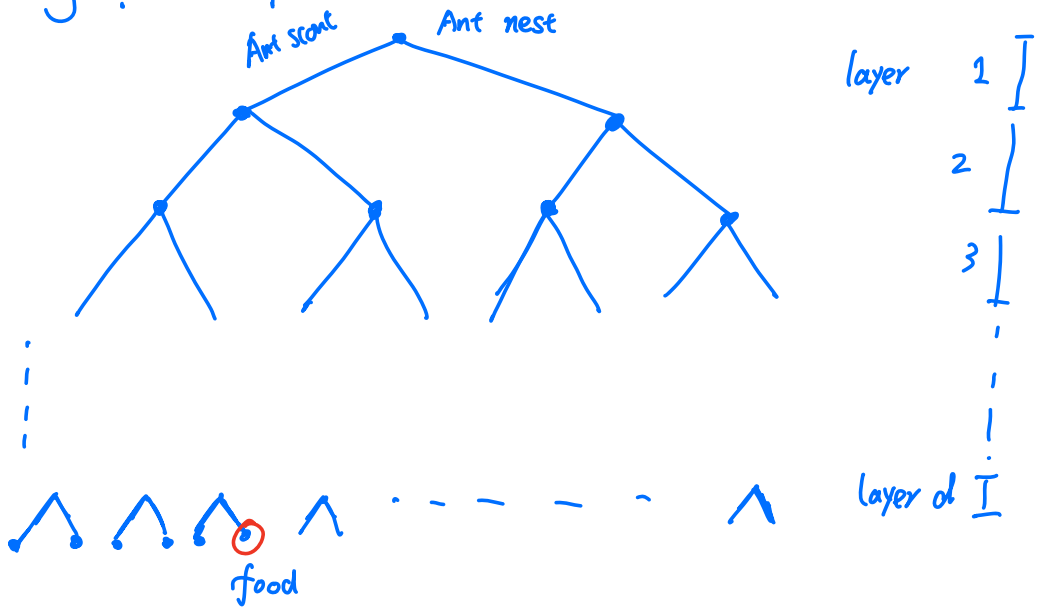
$$= \sum_{i=0}^{\infty} p(1-p)^i \left(\log\frac{1}{p} + i \log\frac{1}{(1-p)}\right)$$

$$= \log\frac{1}{p} \sum_{i=0}^{\infty} p(1-p)^i + p\log\frac{1}{1-p} \sum_{i=0}^{\infty} i\, p(1-p)^i$$

$$= \log\frac{1}{p} + p\log\frac{1}{(1-p)} \cdot \frac{1-p}{p^2} = \frac{h(p)}{p} \to +\infty \quad \text{as } p \to 0$$

Example (∞ entropy): Can $H(X) = +\infty$? Yes, $P(X=k) = \frac{c}{k\ln^2 k}$, $k = 2,3,\cdots$

Why "log": A experiment



Time for ant scout to describe the location of food $\sim \log_2 2^d = d$

left, right left · · · d binary digit

ant communication ~ 1 bit/min

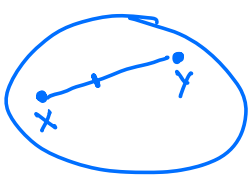
Convexity

V a vector space ($V \cong \mathbb{R}^n$),

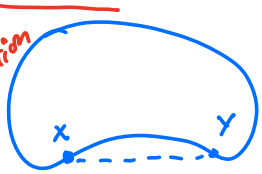
A subset $S \subseteq V$ is convex if

$\forall\ x, y \in S$, $\lambda x + (1-\lambda) y \in S$ for $\lambda \in [0,1]$

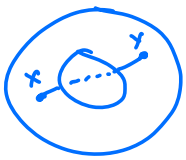
↓ Convex combination



Convex



not convex



Example: ① $\mathbb{R}^n$ is convex

$[0,1] \subseteq \mathbb{R}$, $(a,b) \subseteq \mathbb{R}$

② $\mathcal{P}(\mathcal{X}) = \{$ prob. distribution on $\mathcal{X}\}$

③ $\mathcal{P}_0(\mathbb{R}) = \{P_X \mid \mathbb{E}(X)=0\} \subseteq \mathcal{P}(\mathbb{R})$

$\mathbb{E}(\lambda X+(1-\lambda)Y) = \lambda\mathbb{E}X+(1-\lambda)\mathbb{E}Y = 0$

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

(i) convex if $f(\lambda x+(1-\lambda)y) \le \lambda f(x)+(1-\lambda)f(y)$, $\forall x,y \in S$. $\lambda \in [0,1]$

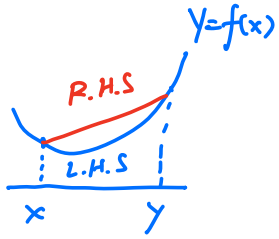
(ii) strictly convex if $f(\lambda x+1-\lambda) < \lambda f(x)+(1-\lambda)f(y)$, $\forall x \neq y \in S$, $\lambda \in (0,1)$

(iii) (strictly) concave if $-f$ is (strictly) convex

Example: ① $x \longmapsto x\log x$ convex strictly

$x \longmapsto \log x$ concave strictly

② $X \longmapsto \mathbb{E}X$ convex but not strictly (proof?)



Jensen inequality: $\forall\ X: \Omega \to S \subseteq \mathbb{R}^n$ vector valued R.V.

f convex $\Rightarrow f(\mathbb{E}X) \le \mathbb{E}f(X)$

If strictly convex, then $f(\mathbb{E}X) = \mathbb{E}f(X)$ iff $X = \mathbb{E}X$ a.s.

constant R.V.

Pf: Convexity $\Rightarrow$ $f(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n)$

$\leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots \lambda_n f(x_n)$

$\lambda_i \geq 0 \quad \sum_{i=1}^{n} \lambda_i = 1$

$f(\mathbb{E}X) = f\left(\sum_w P(w) X(w)\right) \leq \sum_w P(w) f(X(w))$

$P(w) \geq 0 \quad \sum P(w) = 1$ $\qquad = \mathbb{E} f(X)$

Properties of H

① $H(X) \geq 0$. $H(X) = 0$ iff X is constant

② If $\mathcal{X}$ is finite, $H(X) \leq \log |\mathcal{X}|$ with equality iff P_X is uniform on $\mathcal{X}$

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

④ $P \longmapsto H(P)$ is strictly concave

Pf: ① $H(X) = \mathbb{E}\left[\log \frac{1}{P_X}\right] \geq 0$ $\qquad P_X(x) \leq 1, \ \log \frac{1}{P_X(x)} \geq 0$

② $H(X) = \mathbb{E}\left[\log \frac{1}{P_X}\right] \leq \log \mathbb{E}\left(\frac{1}{P_X}\right)$

$= \log \sum_x P(x) \frac{1}{P(x)} = \log |\mathcal{X}|$

equality iff $\log \frac{1}{P_X}$ is constant

$\Leftrightarrow P_X$ constant

$\sum_{x \in \mathcal{X}} P_X(x) = 1 \Rightarrow P_X(x) = \frac{1}{|\mathcal{X}|}$

③ $P_X(x) = P(\{w \mid X(w) = x\}) = P(\{w \mid f \circ X(w) = f(x)\}) = P_{f(X)}(f(x))$

$H(X) = \sum_x P_X(x) \log \frac{1}{P_X(x)} = \sum_x P_{f(X)}(f(x)) \log \frac{1}{P_{f(X)}(f(x))} = H(X)$

④: $H(\lambda P_1 + (1-\lambda)P_2) = \sum_w f(\lambda P_1(w) + (1-\lambda)P_2(w))$ $\quad f(t) = t\log\frac{1}{t}$

$\geq \sum_w \lambda f(P_1(w)) + (1-\lambda) f(P_2(w))$ $\quad = -t\log t$ concave

$= \lambda \sum_w f(P_1(w)) + 1-\lambda \sum_w f(P_2(w))$

$= \lambda H(P_1) + \lambda H(P_2)$

Def: Let P be a prob. distribution on a discrete Ω. The (Shannon) entropy

$$H(P) := \sum_{w\in\Omega} P(w) \log \frac{1}{P(w)}$$

For a discrete R.V. $X: \Omega \to \mathcal{X}$

$$H(X) := H(P_X) = \sum_{x\in\mathcal{X}} P_X(x) \log \frac{1}{P_X(x)} = \mathbb{E}\left(\log \frac{1}{P_X(x)}\right)$$

(real R.V.)

$\log \frac{1}{P_X(x)}$: the suprisal of $X=x$ happens, $H(X)$: the uncertainty/Randomness of P_X.

Rem 1. Basis of log

$\log_2 \longleftrightarrow$ bits $\qquad \log_{256} \longleftrightarrow$ bytes $=$ 8 bits

2. We agree $0 \log \frac{1}{0} = 0$ by $\lim_{x\to 0} x \log \frac{1}{x} = 0$.

Example (Bernoulli): $X \in \{0,1\}$. $P(X=1)=p$ $\quad P(X=0)=1-p$

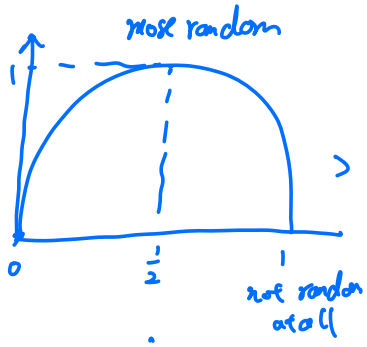
$$H(X) = p\log\frac{1}{p} + (1-p)\log\frac{1}{(1-p)} := h(p)$$

where $h(\cdot)$ is called binary entropy function

$h(\frac{1}{2}) = 1 \qquad h(0) = h(1) = 0$

In $\log_2$ basis, $\quad h(p) \le 1$ and

$h(p) = 1$ iff $p = \frac{1}{2}$



Example (Geometric): $X \in \{0,1,2,\dots\}$ $\qquad P(X=i) = p(1-p)^i$ $\quad i = 0,1,2\cdots$

$$H(X) = \sum_{i=0}^{\infty} p(1-p)^i \log \frac{1}{p(1-p)^i}$$

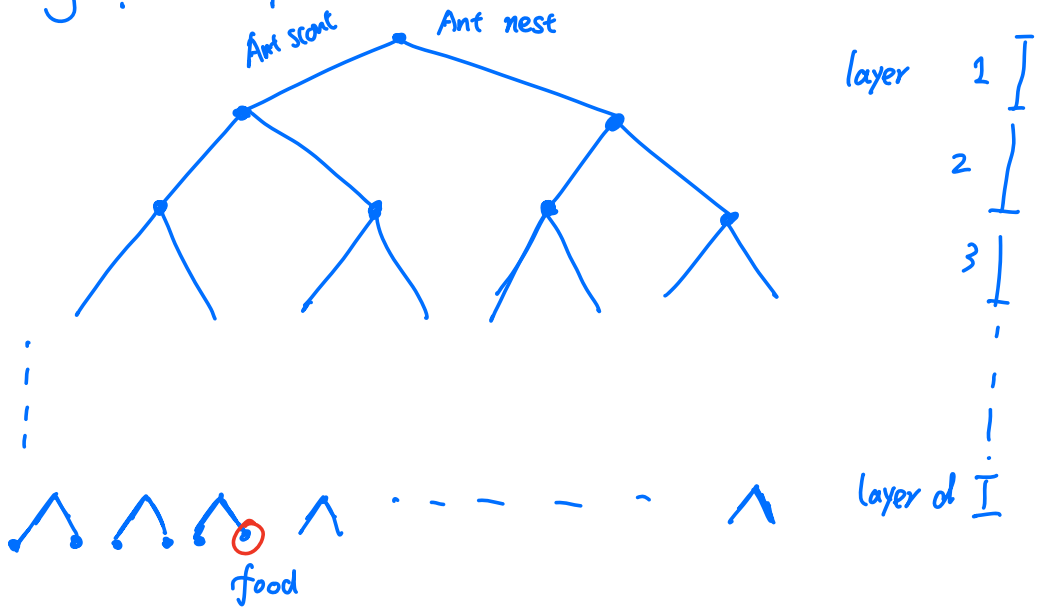
$$= \sum_{i=0}^{\infty} p(1-p)^i \left(\log\frac{1}{p} + i \log\frac{1}{(1-p)}\right)$$

$$= \log\frac{1}{p} \sum_{i=0}^{\infty} p(1-p)^i + p\log\frac{1}{1-p} \sum_{i=0}^{\infty} i\, p(1-p)^i$$

$$= \log\frac{1}{p} + p\log\frac{1}{(1-p)} \cdot \frac{1-p}{p^2} = \frac{h(p)}{p} \to +\infty \quad \text{as } p \to 0$$

Example (∞ entropy): Can $H(X) = +\infty$? Yes, $P(X=k) = \frac{c}{k\ln^2 k}$, $k = 2,3,\cdots$

Why "log": A experiment



Time for ant scout to describe the location of food $\sim \log_2 2^d = d$

left, right left · · · d binary digit

ant communication ~ 1 bit/min

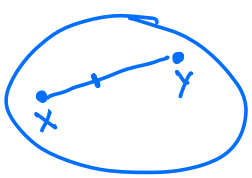
Convexity

V a vector space ($V \cong \mathbb{R}^n$),

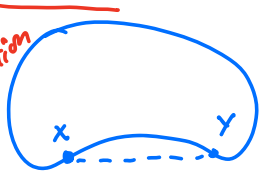
A subset $S \subseteq V$ is convex if

$\forall\ x, y \in S$, $\lambda x + (1-\lambda) y \in S$ for $\lambda \in [0,1]$

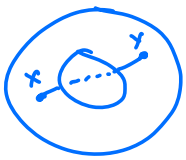
↓ Convex combination



Convex



not convex



Example: ① $\mathbb{R}^n$ is convex

$[0,1] \subseteq \mathbb{R}$, $(a,b) \subseteq \mathbb{R}$

② $\mathcal{P}(\mathcal{X}) = \{$ prob. distribution on $\mathcal{X}\}$

③ $\mathcal{P}_0(\mathbb{R}) = \{P_X \mid \mathbb{E}(X)=0\} \subseteq \mathcal{P}(\mathbb{R})$

$\mathbb{E}(\lambda X+(1-\lambda)Y) = \lambda\mathbb{E}X+(1-\lambda)\mathbb{E}Y = 0$

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

(i) convex if $f(\lambda x+(1-\lambda)y) \le \lambda f(x)+(1-\lambda)f(y)$, $\forall x,y \in S$. $\lambda \in [0,1]$

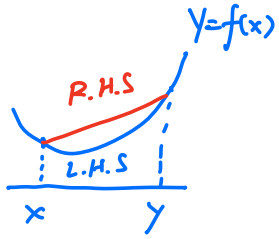
(ii) strictly convex if $f(\lambda x+1-\lambda) < \lambda f(x)+(1-\lambda)f(y)$, $\forall x \neq y \in S$, $\lambda \in (0,1)$

(iii) (strictly) concave if $-f$ is (strictly) convex

Example: ① $x \longmapsto x\log x$ convex strictly

$x \longmapsto \log x$ concave strictly

② $X \longmapsto \mathbb{E}X$ convex but not strictly (proof?)



Jensen inequality: $\forall\ X: \Omega \to S \subseteq \mathbb{R}^n$ vector valued R.V.

f convex $\Rightarrow f(\mathbb{E}X) \le \mathbb{E}f(X)$

If strictly convex, then $f(\mathbb{E}X) = \mathbb{E}f(X)$ iff $X = \mathbb{E}X$ a.s.

constant R.V.

Pf: Convexity $\Rightarrow$ $f(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n)$

$\leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots \lambda_n f(x_n)$

$\lambda_i \geq 0 \quad \sum_{i=1}^{n} \lambda_i = 1$

$f(\mathbb{E}X) = f\left(\sum_w P(w) X(w)\right) \leq \sum_w P(w) f(X(w))$

$P(w) \geq 0 \quad \sum P(w) = 1$ $\qquad = \mathbb{E} f(X)$

Properties of H

① $H(X) \geq 0$. $H(X) = 0$ iff X is constant

② If $\mathcal{X}$ is finite, $H(X) \leq \log |\mathcal{X}|$ with equality iff P_X is uniform on $\mathcal{X}$

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

④ $P \longmapsto H(P)$ is strictly concave

Pf: ① $H(X) = \mathbb{E}\left[\log \frac{1}{P_X}\right] \geq 0$ $\qquad P_X(x) \leq 1, \ \log \frac{1}{P_X(x)} \geq 0$

② $H(X) = \mathbb{E}\left[\log \frac{1}{P_X}\right] \leq \log \mathbb{E}\left(\frac{1}{P_X}\right)$

$= \log \sum_x P(x) \frac{1}{P(x)} = \log |\mathcal{X}|$

equality iff $\log \frac{1}{P_X}$ is constant

$\Leftrightarrow P_X$ constant

$\sum_{x \in \mathcal{X}} P_X(x) = 1 \Rightarrow P_X(x) = \frac{1}{|\mathcal{X}|}$

③ $P_X(x) = P(\{w \mid X(w) = x\}) = P(\{w \mid f \circ X(w) = f(x)\}) = P_{f(X)}(f(x))$

$H(X) = \sum_x P_X(x) \log \frac{1}{P_X(x)} = \sum_x P_{f(X)}(f(x)) \log \frac{1}{P_{f(X)}(f(x))} = H(X)$

④: $H(\lambda P_1 + (1-\lambda)P_2) = \sum_w f(\lambda P_1(w) + (1-\lambda)P_2(w))$ $\quad f(t) = t\log\frac{1}{t}$

$\geq \sum_w \lambda f(P_1(w)) + (1-\lambda) f(P_2(w))$ $\quad = -t\log t$ concave

$= \lambda \sum_w f(P_1(w)) + 1-\lambda \sum_w f(P_2(w))$

$= \lambda H(P_1) + \lambda H(P_2)$

Random vector

Let $X_1, \cdots X_n : \Omega \to \mathcal{X}$ be R.V.s

Define $X^n = (X_1, \cdots X_n) : \Omega \to \mathcal{X}^n$ n-dim random vector.

Entropy: $H(X^n) = H(X_1, X_2, \cdots X_n)$

$$= \mathbb{E}\left[\frac{1}{\log P_{X_1 X_2 \cdots X_n}}\right]$$

In particular, for two R.V. X and Y

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

Definition (Condition Entropy).

$$H(X|Y) = \mathop{\mathbb{E}}_{y \sim P_Y}\left[H(P_{X|Y=y})\right] = \mathbb{E}\left[\log \frac{1}{P_{X|Y}}\right]$$

Expected uncertainty $H(P_{X|Y=y})$ over $y \sim P_Y$.

Notation ① $P_{X|Y=y}$ is a distribution on $\mathcal{X}$ $\quad P_{X|Y=y}(x) = P(X=x|Y=y)$

$$= \frac{P(X=x, Y=y)}{P(Y=y)}$$

② $P_{X|Y}$ is a R.V. on $\mathcal{X} \times \mathcal{Y}$: $\quad P_{X|Y}(x, y) = P(X=x|Y=y)$

Example: A fair die $\Omega = \{1, 2, \cdots, 6\}$

$X = \begin{cases} \text{large} \\ \text{small} \end{cases}$ $Y = \begin{cases} \text{even} \\ \text{odd} \end{cases}$

$H(X) = 1$ $H(Y) = 1$

$P(S \& E) = \frac{1}{6}$ $P(L \& E) = \frac{1}{3}$ $P(S \& O) = \frac{1}{3}$ $P(L \& O) = \frac{1}{6}$

$H(XY) = \frac{1}{6}\log 6 + \frac{1}{3}\log 3 + \frac{1}{3}\log 3 + \frac{1}{6}\log 6$

$= \log 3 + \frac{1}{3}\log 2$

$P_{X|Y}(S|E) = \frac{1}{3}$ $P_{X|Y}(L|E) = \frac{2}{3}$

$P_{X|Y}(S|O) = \frac{2}{3}$ $P_{X|Y}(L|O) = \frac{1}{3}$

$H(X|Y) = P_{XY}(SE) \log \frac{1}{P_{X|Y}(S|E)} + \cdots$

$= \frac{1}{6}\log 3 + \frac{1}{3}\log \frac{3}{2} + \frac{1}{3}\log \frac{3}{2} + \frac{1}{6}\log 3$

$= \log 3 + \frac{2}{3}\log \frac{1}{2}$

Properties of $H(X|Y)$

① $H(X|Y) \le H(X)$ with "=" iff X and Y independent

② $H(XY) = H(Y) + H(X|Y) \le H(Y) + H(X)$

with "=" iff X and Y independent

③ $H(XY) \ge \max\{H(X), H(Y)\}$

Pf: ① Using concavity of $P \mapsto H(P)$

$$H(X|Y) = \mathop{\mathbb{E}}_{y\sim P_Y} H(P_{X|Y=y}) \le H\left(\mathop{\mathbb{E}}_{y\sim P_Y} P_{X|Y=y}\right)$$

$$= H(P_X) = H(X)$$

② $H(XY) = \mathbb{E}\left[\log \frac{1}{P_{XY}}\right]$ $\quad P_{XY}(x,y) = P_{X|Y}(x|y)\,P_Y(y)$

$$= \mathbb{E}\left[\log \frac{1}{P_{X|Y}\cdot P_Y}\right]$$

$$= \mathbb{E}\left[\log \frac{1}{P_{X|Y}} + \log \frac{1}{P_Y}\right]$$

$$= \mathbb{E}\left[\log \frac{1}{P_{X|Y}}\right] + \mathbb{E}\left[\log \frac{1}{P_Y}\right]$$

$$= H(X|Y) + H(Y)$$

③ $H(XY) = H(X) + H(Y|X) \ge H(X)$

Corollary. For any function f,

① $H(X) = H(X, f(X))$, ② $H(f(X)|X) = 0$

① $H(X) \geq H(f(X))$

with equality iff f injective

Pf: ① $P_{X,f(X)}(x,y) = \begin{cases} P_X(x), & \text{if } y = f(x) \\ 0, & \text{otherwise} \end{cases}$

② $H(X) = H(X, f(X)) = H(X) + H(f(X)|X)$

More explicitly. $P_{f(X)|X}(y|x) = \begin{cases} 1 & \text{if } y = f(x) \\ 0 & \text{otherwise} \end{cases}$

$$H(f(X)|X) = \mathop{\mathbb{E}}_{x \sim P_X} \left(H(f(X)|X=x) \right) = 0$$

③ $H(X) = H(X|f(X)) + H(f(X))$

$H(X|f(X)) \geq 0$

Equality $\Rightarrow$ $H(X|f(X)) = 0$

$\Rightarrow \mathop{\mathbb{E}}_{y \sim P_{f(X)}} H(X|f(X)=y) = 0$

$\Rightarrow H(X|f(X)=y) = 0 \ \forall y$

$\Rightarrow X = g f(x) \ \forall y \in f(\mathcal{X}) \Rightarrow f$ injective.

History: Thermo dynamics

No Perpetual Motion Machine by conservation of energy.
1st law

2nd law: No machine can produce work by only drawing heat from a warm body but without expend heat to environment.

heat ⇄ work (No free conversion from heat to work)

3rd law: Entropy cannot reduce.

Boltzmann & Gibbs

Entropy of ideal gas $S = kn \sum_{j=1}^{l} p_j \log \frac{1}{p_j}$

k Boltzmann constant

n # of particle

Lossless Data Compression

"Today is Wednesday" a message as a sequence of letter

a letter $\longleftrightarrow$ bytes $\{1 \cdots 256\} \longleftrightarrow$ 8 bits $\{0,1\}^8$

"a" 97 2^6 110 0001

See table below

Binary	Oct	Dec	Hex	Glyph	Binary	Oct	Dec	Hex	Glyph	Binary	Oct	Dec	Hex	Glyph
010 0000	040	32	20	(space)	100 0000	100	64	40	@	110 0000	140	96	60	`
010 0001	041	33	21	!	100 0001	101	65	41	A	110 0001	141	97	61	a
010 0010	042	34	22	"	100 0010	102	66	42	B	110 0010	142	98	62	b
010 0011	043	35	23	#	100 0011	103	67	43	C	110 0011	143	99	63	c
010 0100	044	36	24	$	100 0100	104	68	44	D	110 0100	144	100	64	d
010 0101	045	37	25	%	100 0101	105	69	45	E	110 0101	145	101	65	e
010 0110	046	38	26	&	100 0110	106	70	46	F	110 0110	146	102	66	f
010 0111	047	39	27	'	100 0111	107	71	47	G	110 0111	147	103	67	g
010 1000	050	40	28	(	100 1000	110	72	48	H	110 1000	150	104	68	h

Is this optimal in # of bits? No, if only english words we only need

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

one letter — 6 bits

Lossless compression

Source X (alphabet) $\xrightarrow[\text{compressor or encoder}]{f}$ $\{0,1\}^*$ $\xrightarrow[\text{Decompressor or decodor}]{g}$ X

① $\{0,1\}^* = \{\phi, 0, 1, 00, 01, 10, 11, 000 \cdots\}$ $\{$ bit string

② for $x \in X$, $f(x) \in \{0,1\}^*$ code word. $\{f(x) \mid x \in X\}$ code book (countable)

③ Lossless compression: $g \circ f = I_X \Rightarrow f$ injective

need $|f(X)| = |X|$ code word

④ length function: $l : \{0,1\}^* \to \mathbb{N}$, e.g. $l(01101) = 5$

an alphabet X need code word with maximal length

$\sup l(f(x)) = \log_2 |X|$

Can we compress more? In terms of maximal codeword length, no
expected codeword length, Yes

Example: $X = \{a, b, c, d\} \xrightarrow{w} \{0,1\}^2$

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

each 2 bits $\qquad \forall x,\ L(w(x)) = 2$ (length)

Now given the fact

$P(a) = 1/2 \quad P(b) = 1/8 \quad P(c) = 1/4, \quad P(d) = 1/8$

Consider $a \to 0, \quad b \to 110, \quad c \to 10, \quad d \to 111$ ← variable length code

$0011010111 \longleftrightarrow a\ a\ b\ c\ d\ a$

Expected codeword length

$$\frac{1}{2}(1) + \frac{1}{8}(3) + \frac{1}{4}(2) + \frac{1}{8}(3) = \frac{7}{4} < 2 \quad !$$

In the long turn, do better than 2 bits/letter.

What does "$P(a) = \frac{1}{2} \quad P(b) = \frac{1}{8}$ - - - - " mean?

frequency of letters in two different English novel are approximately the same.

English text $\xrightarrow{\text{empirical}}$ frequency of letters $\xrightarrow{\text{modeled by}}$ probabilistic feature

distribution on Alphabet
R.V.

—Shannon

Objective: minimize $\nearrow$ $\sup \ell[f(X)]$
$\searrow$ $\mathbb{E}\ \ell[f(X)]$

There is an optimal compressor f^* minimize both!

Main idea — Assigen short code words for more probable symbols

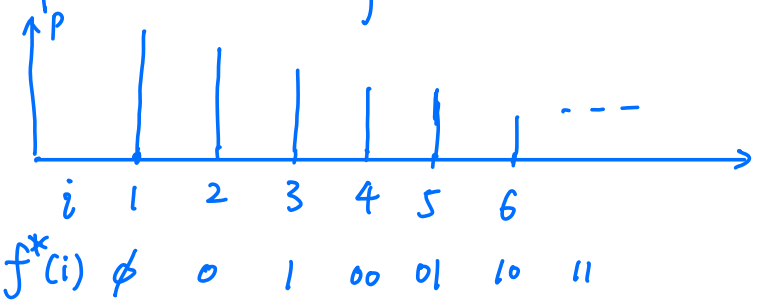
Longer - - - - - - less probable symbols.

WLOG, $\mathcal{X} = \{1, 2, \cdots, |\mathcal{X}|\} \subseteq \mathbb{N}$ and reorder p.m.f $\searrow$

$P_X(i+1) \le P_X(i)$

Theorem (Optimal Compressor)

Define the encoder f^*



Then

1. length of code word

$$\ell(f^*(i)) = \lfloor \log_2 i \rfloor \qquad (\lfloor a \rfloor := \text{largese integer} \le a)$$

2. Stochastically optimal: $\forall$ encoder f,

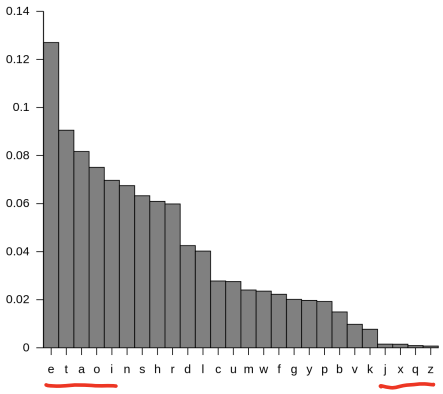
$$\forall k,\ \mathbb{P}[\ell(f(X)) \le k] \le \mathbb{P}[\ell(f^*(X)) \le k] \qquad (\ell f^*(X) \overset{st.}{\le} \ell f(X))$$

As a consequence,

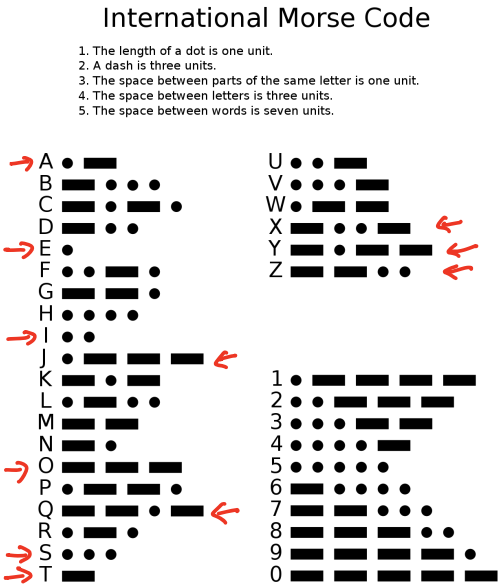
$$\sup \ell f(X) \ge \sup \ell f^*(X).\quad \mathbb{E}\,\ell f(X) \ge \mathbb{E}\,\ell f^*(X)$$

Example (Telegraph & Morse code)

Letter frequency in English text



Morse code



For the other side, we recall the lemma:

Lemma. For $Z \in \mathbb{N}$ and $\mathbb{E}Z < \infty$, $H(Z) \le \mathbb{E}[Z]\, h\left(\frac{1}{\mathbb{E}Z}\right)$

Entropy of Geometric distribution $Q_p(i) = p(1-p)^i$ $\quad H(Q_p) = \frac{h(p)}{p}$

$\mathbb{E}(Q_p) = \frac{1}{p}$

$\frac{1}{p} = \mathbb{E}Z \Rightarrow H(Z) \le H(Q_p)$

$Z \in \mathbb{N}$

Geometric distribution has largese entropy for integer valued RV of given mean $p = \mathbb{E}X$

How to prove? ① Lemma $\forall$ P,Q prob. $D(P\|Q) := \sum p(w) \log \frac{p(w)}{Q(w)} \ge 0$ (Relative entropy)

② Calculate $D(P\|Q_p) = \sum p(w) \log p(w) - p(w) \log Q(w)$

where $p = \frac{1}{\mathbb{E}X}$.

Pf: $|A_k| := |\{x \mid l \circ f(x) \le k\}| \le \sum_{k=0}^{k} 2^i = 2^k - 1$

$= |\{x \mid l^* \circ f(x) \le k\}| = |A_k^*|$

Because $A_k^* = \{1, 2 \cdots, 2^{k-1}\}$

$$P[l \circ f(x) \le k] = \sum_{x \in A_k} P_X(x) \le \sum_{i=1}^{2^{k-1}} P_X(i) = \sum_{x \in A_k^*} P_X(x) = P[l^* \circ f(x) \le k]$$

□

$X, Y : \Omega \to \mathbb{R}$ real R.Vs.

Def. We say X is stochastically dominated by Y, denoted $Y \overset{st.}{\le} X$ if

$\forall k$, $\quad P_X(X \le k) \le P_Y(Y \le k)$

Prop. If $Y \overset{st.}{\le} X$, then

i) $\sup Y \le \sup X$ and ii) $\mathbb{E}Y \le \mathbb{E}X$

Pf: i) $\sup X = \sup\{k \mid P_X(X \le k) < 1\} \ge \sup\{k \mid P_Y(X \le k)\} = \sup Y$

ii) special case: $X, Y : \Omega \to \mathbb{N}$.

$$\boxed{\mathbb{E}X = \sum_{n=1} P_X(X \ge n)}$$

Indeed, $\mathbb{E}X = \sum_{n=1} n\, P_X(n) = P_X(X \ge 1) + \sum_{n=2} P_X(X \ge n)$

$$= \sum_{n=1} P_X(X \ge n)$$

$$\mathbb{E}X = \sum_{n=1} P_X(X \ge n) \ge \sum_{n=1} P_Y(Y \ge n) = \mathbb{E}Y$$

Theorem (Optimal Average code length)
Given $X \in \mathbb{N}$ and $P_X(1) \geq P_X(2) \geq \cdots$. Then

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

② $H(X) - \log_2(eH(X)+e) \leq \mathbb{E}[l(f^*(X))] \leq H(X)$

Pf: ① $\mathbb{E}[l(f^*(X))] = \mathbb{E}(\lfloor \log_2 X \rfloor) = \sum_{k\geq 1} P(\lfloor \log_2 X \rfloor \geq k)$

$= \sum_{k\geq 1} P(\log_2 X \geq k)$

② Denote $L(X) = l(f^*(X))$

$P_X(m) \leq \frac{1}{m}$ b/c $P_X(i)$ decreasing

$\Rightarrow l(f^*(m)) = \lfloor \log_2 m \rfloor \leq \log_2 \frac{1}{P_X(m)}$

$\Rightarrow \mathbb{E}(L(X)) \leq \mathbb{E}(\log \frac{1}{P_X(m)}) = H(X)$

For the other side, we recall the lemma:

Lemma. For $Z \in \mathbb{N}$ and $\mathbb{E}Z < \infty$, $H(Z) \leq \mathbb{E}[Z]\, h(\frac{1}{\mathbb{E}Z})$

$\{0,1,2,\cdots,n,\cdots\}$

$H(X) = H(X,L) = H(X|L) + H(L)$

$= \sum P_L(k) H(X|L=k) + h(\frac{1}{1+\mathbb{E}L})(1+\mathbb{E}L)$

$L \in \{0,1,2,\cdots\}$

$L+1 \in \mathbb{N}$

$\leq \sum P_L(k) \log 2^k + \cdots$

$\geq k$

$= \mathbb{E}L + \log(1+\mathbb{E}L) + (\mathbb{E}L)\log(1+\frac{1}{\mathbb{E}L})$

$$\leq \mathbb{E}L + \log_2(e(1+H(X)))$$

$$\left(x\log\left(1+\frac{1}{x}\right) \leq \log e \quad \forall x>0\right)$$

□

Cor. If $X=S^n$ i.i.d. sequence, $H(S^n) = nH(S)$

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

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

Expected length/message

($\mathbb{E}\ell^*(S^n) = nH(S) - \frac{1}{2}\log n + O(1)$ Szpankowski & Verdú . 2011)

Remark Actually. $\frac{\ell^*(S^n)}{n} \longrightarrow H(S)$ in probability by WLLN.

Almose loseless / error less. Coding

Coding scheme: encoder f, decoder g

$$X \xrightarrow{f} \{0,1\}^* \xrightarrow{g} X$$

Idea: If we want perfect (error less) code, then $k \geq \log_2 |X|$

But any physyic process is subjected to noise,

the transmission is erroneous.

In reality, we do things with certain accuracy / error tolerance.

Def: A pair (f, g) (encoder, decoder) is called a (k, ε)-code for X if (fixed length)

$$f: X \to \{0,1\}^k \qquad g: \{0,1\}^k \to X$$

such that $P(g(f(x)) \neq X) \leq \varepsilon$

k bits string.

undectable error

Remark: Alternative setting with detectable error

$$f: X \to \{0,1\}^k \qquad g: \{0,1\}^k \to X \cup \{e\}$$

such that $\exists S \subseteq X$ s.t. $g(f(x)) = \begin{cases} x & x \in S \\ e & x \notin S \end{cases}$

$$P(g(f(X)) \neq X) = P(g(f(X)) = e) = P_X(S)$$

S loss less part. e detectable error

Def: (Optimal error probability)

$$\varepsilon^*(X,k) := \inf\{\varepsilon : \exists (k,\varepsilon)\text{-code for } X\}$$

Smallese error that a lengeh k code can aeheive.

Thm: $\varepsilon^*(X,k) = P[L(f^*(X)) \geq k] = 1 - \sum_{i=1}^{2^{k-1}} P_X(i)$

(Recall that we assume $P_X(i) \geq P_X(i+1)$)
decreasing

Pf: We assign a code word to each of $\underbrace{2^k-1 \text{ more likely realization of } X}_{S}$. and all the rese to one word "error"

$$\varepsilon^*(X,k) = P[X \notin S] = P[\text{of } f^*(X) \geq k]$$

Actual code book: Variable lengeh $\{\emptyset, 0, 1, 00, 01, \cdots \underbrace{11\cdots1}_{k-1}\}$

fixed length $\{(00\cdots00) \quad (00\cdots01) \quad \cdots \quad (11\cdots10)\}$ 2^{k-1}

$(11\cdots11)$ for "error".

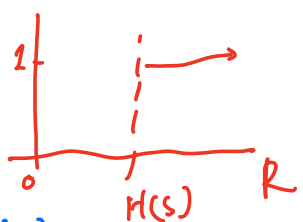
□

Shannon's Source coding Thm 1948

Let S^n be i.i.d. Then

Rate ↓

$$\lim_{n\to\infty} \varepsilon^*(S^n, nR) = \begin{cases} 0 & \text{if } R > H(S) \\ 1 & \text{if } R < H(S) \end{cases}$$



Lemma (Achievability)

$$\varepsilon^*(X,k) \leq P\left[\log_2 \frac{1}{P_X(X)} \geq k\right]$$

Indeed, $\varepsilon^*(X,k) = \sum_{m \geq 2^k} P_X(m) = \sum 1_{\{m \geq 2^k\}} P_X(m) \leq \sum 1_{\{\frac{1}{P_X(m)} \geq 2^k\}} P_X(m)$

$$= \mathbb{E}\, 1_{\{\log_2 \frac{1}{P_X(X)} \geq k\}}$$

Lemma (converse)

$$\varepsilon^*(X,k) \geq P\left[\log \frac{1}{P_X(X)} > k+\tau\right] - 2^{-\tau} \quad , \ \forall \tau > 0$$

Denote, $L = l \circ f^*$. $\quad L(m) = \lfloor \log m \rfloor \leq \log m \leq \log \frac{1}{P_X(m)}$

$$1 - \varepsilon^*(X,k) = P[L \leq k]$$

$$= P\left[L \leq k, \log_2 \frac{1}{P_X} \leq k+\tau\right] + P\left[L \leq k, \log_2 \frac{1}{P_X} > k+\tau\right]$$

$$\leq P\left[\log_2 \frac{1}{P_X} \leq k+\tau\right] + \sum_m P_X(m) \, 1_{\{L(m) \leq k\}} 1_{\{P_X(m) \leq 2^{-k-\tau}\}}$$

$$\leq P\left[\log_2 \frac{1}{P_X} \leq k+\tau\right] + (2^{k+1}-1) \cdot 2^{-k-\tau}$$

(Pf of Thm)

$$P\left[\log_2 \frac{1}{P_X} > k+\tau\right] - 2^{-\tau} \leq \varepsilon^*(X,k) \leq P\left[\log \frac{1}{P_X} \geq k\right]$$

Now take $X = S^n$, $\quad P_{S^n}(s_1 \cdots s_n) = P_S(s_1) \cdot P_S(s_2) \cdots P_S(s_n)$

By WLLN, $\frac{1}{n} \log \frac{1}{P_{S^n}} = \frac{1}{n}\left(\log \frac{1}{P_{S_1}} + \log \frac{1}{P_{S_2}} + \cdots \log \frac{1}{P_{S_n}}\right)$

$$\xrightarrow{P} \quad \mathbb{E}\left(\log \frac{1}{P_S}\right) = H(S)$$

Then

$$\varepsilon^*(S^n, nR) \leq P\left[\log \frac{1}{S^n} > nR\right] \quad \text{(constant)}$$

$$= P\left[\frac{1}{n} \log \frac{1}{S^n} > R\right] \longrightarrow P[H(S) > R] = 0$$

if $R > H(S)$

$$\varepsilon^*(S^n, nR) \geq P\left[\log \frac{1}{P_{S^n}} > nR - \sqrt{n}\right] - 2^{-\sqrt{n}}$$

$$\geq P\left[\frac{1}{n} \log \frac{1}{P_{S^n}} > R - \frac{1}{\sqrt{n}}\right] - 2^{-\sqrt{n}}$$

$$\geq P\left[\frac{1}{n} \log \frac{1}{P_{S^n}} > R\right] - 2^{-\sqrt{n}}$$

$$\lim_{n\to\infty} \varepsilon^*(S^n, nR) \geq \lim_{n\to\infty} P\left[\frac{1}{n}\log\frac{1}{P_{S^n}} > R\right] - \lim_{n\to\infty} 2^{-\sqrt{n}}$$

$$\geq P[H(S) > R] = 1 \quad \text{if } H(S) > R.$$

A Direct argument. Typical Sequence

Let $\mathcal{X}$ be a finite alphabet.

For a sequence $x^n = x_1 \cdots x_n \in \mathcal{X}^n$, we can define

$$N_{x^n}(a) = \#\{i \mid x_i = a\}$$

Then $\frac{N_{x^n}(a)}{n}$ is the frequency "a" appears in x^n

$P_{x^n}(a) := \frac{N_{x^n}(a)}{n}$ emperical distribution of x^n.

Now

$X \sim P$ R.V. on $\mathcal{X}$. $X^n = X_1 \cdots X_n$ i.i.d

X^n is a random sequence.

What would be mosely likely sequence?

$P_{X^n} \cong P_X$

A sequence x^n is strong (n,ε)-typical for P if

$$|P_{X^n}(a) - P_X(a)| < \frac{\varepsilon}{|\mathcal{X}|} \quad \forall a \in \mathcal{X}$$

is weak (n,ε)-typical if

$$H(P)+\varepsilon < \frac{1}{n}\log\frac{1}{P_{X^n}(x^n)} < H(P)-\varepsilon$$

We denote $T_\varepsilon^n(P) \subseteq \mathcal{X}^n$ be the set of strong typical sequence of P

$A_\varepsilon^n(P) \subseteq \mathcal{X}^n$ --- weak ---- of P

Theorem (Typicality) ① $T_\varepsilon^n(P) \subseteq A_\varepsilon^n(P)$

② $(1-\varepsilon)2^{n(H(X)-\varepsilon)} \le |A_\varepsilon^n(P)| \le 2^{n(H(P)+\varepsilon)}$

③ $2^{n(H(X)-\delta(\varepsilon))}|T_\varepsilon^n(P)| \le 2^{n(H(P)+\delta(\varepsilon))}$

where $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$

④ $P_{X^n}(X^n \in A_\varepsilon^n) \to 1$

$P_{X^n}(X^n \in T_\varepsilon^n) \to 1$ as $n \to \infty$

$x^n \in A_\varepsilon^n(P) \Leftrightarrow 2^{-n(H(P)+\varepsilon)} < P_{X^n}(x^n) < 2^{-n(H(P)-\varepsilon)}$

$P_{X^n}(X^n \in A_\varepsilon^n) = P_{X^n}\left(\left|\frac{1}{n}\sum_{i=1}^n \log\frac{1}{P_{X_i}} - H(X)\right| \le \varepsilon\right) \to 1$

by W.L.L.N

Alternative proof for Shannon's Coding Theorem.

Choose error free set $S(n) = A_\varepsilon^n$. Assign a codeword to each $x^n \in S(n)$.

need $|S(n)| \leq 2^{n(H(X)+\varepsilon)}$ codeword

word length $\leq \lfloor n(H(X)+\varepsilon) \rfloor$

error probability $\varepsilon(n) = 1 - P(S(n)) \longrightarrow 0$

□

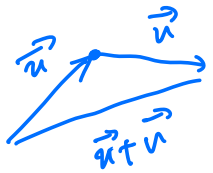
Preliminary on Linear Algebra

1. Complex vector space

n-dim (complex) v.s. $\mathbb{C}^n = \left\{ \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \mid u_i \in \mathbb{C} \right\}$

Two operations:

① Addition: $\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} u_1+v_1 \\ \vdots \\ u_n+v_n \end{pmatrix}$



② Scalar multiplication: $\alpha \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} \alpha u_1 \\ \alpha u_2 \\ \vdots \\ \alpha u_n \end{pmatrix}$

$\alpha u + \beta v$ is a linear combination of u and v

Standard basis:

$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ ---- $e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$

$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = u_1 e_1 + u_2 e_2 + \cdots u_n e_n$

$u \xleftrightarrow{\{e_i\}} (u_1 \cdots u_n)$

In general, let I be a finite set. We define

$$\mathbb{C}^I = \{u : I \to \mathbb{C}\} \text{ all complex valued functions on } I$$

Addition: $\forall\, u, v \in \mathbb{C}^I$, $u+v \in \mathbb{C}^I$

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

Scalar multiplication: $\forall\, u \in \mathbb{C}^I$, $\alpha \in \mathbb{C}$, $\alpha u = \alpha \cdot u \in \mathbb{C}^I$

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

$\mathbb{C}^n := \mathbb{C}^{\{1,\cdots n\}}$ n-dimension complex v.s.

$$u \in \mathbb{C}^{\{1,\cdots n\}} \longleftrightarrow \begin{pmatrix} u(1) \\ u(2) \\ \vdots \\ u(n) \end{pmatrix}$$

Thm: $\mathbb{C}^I \cong \mathbb{C}^n$ iff $n = |I|$

$I = \{x_1 \cdots x_n\} \longleftrightarrow \{1 \cdots n\}$

$e_x(y) = \delta_{xy} = \begin{cases} 1 & \text{if } x=y \\ 0 & \text{if } x \neq y \end{cases}$ $\qquad e_i = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ - ith

Example $X = \{0,1\}^2 = \{00, 01, 10, 11\}$

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

Inner product : $\langle \cdot, \cdot \rangle : \mathbb{C}^I \times \mathbb{C}^I \to \mathbb{C}$

$$\langle u, v \rangle = \sum_{i \in I} \overline{u(i)}\, v(i)$$

$\overline{a+bi} = a-bi$ $\quad a, b \in \mathbb{R}$

Definition - Properties

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

$$\langle u, v \rangle = \overline{\langle v, u \rangle} \qquad \forall\, u, v$$

③ Positivity:

$$\langle u, u \rangle \geq 0$$

with equality iff $u = 0$.

Note: ① + ② implies anti-linear in first input

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

$\bar{\alpha} \alpha = |\alpha|^2$

Norm: $\|u\| := \sqrt{\langle u, u \rangle} = \sqrt{\sum_{i \in I} |u(i)|^2}$ $\quad u$ is a unit vector if $\|u\| = 1$

distance $d(u, v) = \|u - v\| = \sqrt{\sum_i (u(i) - v(i))^2}$ Euclidean distance

Defining - Properties

1. Positivity: $\|u\| \geq 0$ with equality iff $u = 0$
2. $\|\alpha u\| = |\alpha| \, \|u\| \quad \forall\, \alpha \in \mathbb{C}$
3. $\|u + v\| \leq \|u\| + \|v\| \quad \forall\, u, v \in \mathbb{C}^n$ triangle inequality.

⇔ Cauchy - Schwarz inequality ("The inequality" in Hilbert space)

$\forall\, u, v \in \mathbb{C}^n \qquad |\langle u, v \rangle| \leq \|u\| \, \|v\|$

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

CS inequality ⇔ triangle inequality.

Other example of norms: $\|u\|_p = \left(\sum_{i\in I} |u(i)|^p\right)^{\frac{1}{p}}$ (How to prove triangle inequality?)

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

Hilbert space norm $\|u\| := \|u\|_2$ $p=2$

unit: $\|u\|=1$

A set $\{u_1, u_2, \cdots u_k\}$ is orthogonal if $\langle u_i, u_j\rangle = 0$ if $i \neq j$.

is orthonormal if orthogonal and $\|u_i\|=1$

is a orthonormal basis if orthonormal and basis

E.g. $e_1=(1,0\cdots 0)$, $e_2=(0,1,0\cdots 0)$ $\cdots$ $e_n=(0,0\cdots 1)$

O.N.B. $\mathbb{C}^n$

$X=\{x_1 \cdots x_n\}$ $e_{x_1} \cdots$ O.N.B $\mathbb{C}^X$

$(\mathbb{C}^n, \langle\cdot,\cdot\rangle)$ is called n-dim complex Hilbert space (or Euclidean)

Thm Every n-dim $\mathbb{C}$ Hilbert space is isomorphic to $(\mathbb{C}^n, \langle\cdot,\cdot\rangle)$

In general, Hilbert space := vector space + inner product. "Completeness" infinite dim

E.g. $L_2(\mathbb{R}) = \{ f: \mathbb{R} \to \mathbb{C} \mid \int |f(x)|^2 dx < \infty \}$ (measurable). $\langle f, g\rangle = \int \overline{f(x)} g(x)\, dx$

Linear operator:

Let V, W be complex v.s. A map $L: V \to W$ is linear if

$$L(\alpha u + \beta v) = \alpha L(u) + \beta L(v) \quad \forall \alpha, \beta \in \mathbb{C},\ u, v \in V$$

$\mathcal{L}(V, W) :=$ the space of all linear operator

$\mathcal{L}(V, W)$ is a complex vector space with

Addition: $(A+B)u := Au + Bu$

Scalar multiplication: $(\alpha A)u = \alpha Au$

$\dim V = n \quad \dim W = m$

Given an basis $\{v_1 \cdots v_n\}$ of V

basis $\{w_1 \cdots w_m\}$ of W,

$a_{ij} \in \mathbb{C}$

$$A v_1 = a_{11} w_1 + a_{12} w_2 + \cdots a_{1m} w_m = \begin{pmatrix} a_{11} \\ \vdots \\ a_{1m} \end{pmatrix}$$

$$A v_2 = a_{21} w_1 + a_{22} w_2 + \cdots a_{2m} w_m$$

$$\vdots$$

$$A v_n = a_{n1} w_1 + \cdots a_{nm} w_m = \begin{pmatrix} a_{n1} \\ \vdots \\ a_{nm} \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \xrightarrow{M} \begin{pmatrix} \sum a_{1j} b_j \\ \sum a_{2j} b_j \\ \vdots \\ \sum a_{nj} b_j \end{pmatrix} = \begin{pmatrix} a_{11} a_{12} \cdots \\ a_{21} a_{22} \cdots \\ \vdots \qquad a_{nm} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Definition of matrix multiply vector.

$\{v_i\}$ $\qquad$ $\{w_j\}$

$$V = b_1 v_1 + \cdots b_n v_n \xrightarrow{A} AV = A(b_1 v_1 + \cdots b_n v_n)$$

$$= b_1 A v_1 + \cdots b_n A v_n$$

$$= \sum_{i=1} \sum_{j=1} (a_{ij} b_j) w_i$$

$M = (a_{ij})_{\substack{1\le i\le m \\ 1\le j\le n}}$ is $m \times n$ complex matrix

$M_{ij} = \langle w_i, A\, v_j \rangle$

$M_{n\times m}$ space of all $n\times m$ complex matrices $\cong L(\mathbb{C}^n, \mathbb{C}^m)$ space of Linear operator

matrix M $\xleftrightarrow{\{v_j\} \text{ basis of } \mathbb{C}^n}$ A linear operator

$M\begin{pmatrix} u(1) \\ \vdots \\ u(n) \end{pmatrix}$ $\{w_j\}$ basis of $\mathbb{C}^m$ Au

For $A \in L(V, W)$ $B \in L(W, Z)$. $AB \in L(V, Z)$

$A\circ B(u) := A(B(u))$ "$\circ$" often omitted

$A\circ B(\alpha u + \beta v) := A\,(\,B(\alpha u + \beta v)\,)$

$= A(\alpha Bu + \beta Bv)$

$= \alpha ABu + \beta ABv$. Linear

$$\begin{array}{ccccc} V & \xrightarrow{A} & W & \xrightarrow{B} & Z \\ \{v_j\}\uparrow & & \uparrow\{w_j\} & & \uparrow\{z_i\} \\ \mathbb{C}^n & \xrightarrow{M} & \mathbb{C}^m & \xrightarrow{N} & \mathbb{C}^d \end{array}$$

$A\circ B \longleftrightarrow MN$

$$(MN)_{ik} = \sum_{j=1}^{m} M_{ij}\, N_{jk}$$ Matrix multiplication

$1 \le i \le d$, $1 \le k \le n$

Basis for $M_{n\times m}$

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

$M = \sum M(i,j)\, E_{i,j}$ $\quad \{E_{i,j}\}$ basis for $M_{n\times m}$

Basis for $L(V,W)$

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

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

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

$\{E_{w_j,v_i} \mid \begin{matrix} i=1\cdots n \\ j=1\cdots m \end{matrix}\}$ forms a basis for $L(V,W)$.

So $\dim(M_{n\times m}) = nm$

$\dim(L(V,W)) = \dim V \dim W$

Direct sum of V.S. and Operators

$V_1 = \mathbb{C}^{X_1}$, $V_2 = \mathbb{C}^{X_2}$ $\cdots$ $V_n = \mathbb{C}^{X_n}$ X

$$V_1 \oplus V_2 \oplus \cdots V_n = \left\{ \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \mid v_i \in V_i \right\} := v_1 \oplus v_2 \oplus \cdots \oplus v_n$$

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

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

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

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

$\forall i$, $\{e_k^i\}_{k=1}^{|V_i|}$ O.N.B of V_i $\Rightarrow$ $\{e_k^i\}_{k,i}$ O.N.B of $V_1 \oplus \cdots V_n$

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

Let $A_1 \in L(V_1, W_1) \cdots A_n \in L(V_n, V_m)$

Define $A_1 \oplus A_2 \oplus \cdots A_n \in L(V_n, V_m)$

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

If A_1 has matrix M_1 ... A_n has matrix M_n, $A_1 \otimes \cdots A_n \longleftrightarrow$ $\begin{bmatrix} M_1 & & \\ & M_2 & \\ & & \ddots \\ & & & M_n \end{bmatrix}$ $\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$, with $M_{n_i \times m_i}$ and $\mathbb{C}^{m_i}$

Tensor product

$V = \mathbb{C}^X \quad W = \mathbb{C}^Y$

$$V \otimes W := \mathbb{C}^{X \times Y} = \{u : \overset{(a,b)}{X \times Y} \to \mathbb{C}\}$$

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

$V \otimes W = \text{Span}\{v \otimes w \mid u \in V, v \in W\}$

$= \{\sum \alpha_i v_i \otimes w_i \mid \forall \alpha_i \in \mathbb{C}, v_i \in V, w_i \otimes W\}$

all linear combinations of elementary tensor.

$v_1 \otimes w + v_2 \otimes w = (v_1 + v_2) \otimes w \qquad v \otimes w_1 + v \otimes w_2 = v \otimes (w_1 + w_2)$

$\alpha(v \otimes w) = \alpha v \otimes w = v \otimes \alpha w.$

$\langle v_1 \otimes w_1, v_2 \otimes w_2 \rangle = \langle v_1, v_2 \rangle \langle w_1, w_2 \rangle$

If $\{e_i\} \subseteq V$ basis

$\{f_j\} \subseteq W$ basis.

$$V \otimes W = (a_1 e_1 + \cdots a_n e_n) \otimes (b_1 f_1 + \cdots b_m f_m)$$

$$= \sum_{i,j} a_i b_j \, e_i \otimes f_j$$

$\{e_i \otimes f_j\}_{i,j}$ is a basis for $V \otimes W$

$\dim(V \otimes W) = \dim(V) \dim(W)$

One can similarly define $V_1 \otimes V_2 \otimes \cdots \otimes V_n$

General rule

$$V \otimes W = W \otimes V \qquad V \otimes W \otimes Z \cong (V \otimes W) \otimes Z$$
$$V \oplus W = W \oplus V \qquad V \oplus W \oplus Z = (V \oplus W) \oplus Z$$
$$(V \oplus W) \otimes Z = V \otimes Z + W \otimes Z$$

Tensor product operator

$$A \in L(V_1, W_1) \qquad B \in L(V_2, W_2)$$

Define $A \otimes B \in L(V_1 \otimes V_2, W_1 \otimes W_2)$

$$A \otimes B(v \otimes w) = Av \otimes Bw$$

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

Last time:

Operator $L(\mathbb{C}^n, \mathbb{C}^m)$ $\xleftrightarrow{\text{basis}}$ $M_{m\times n}$ matrix

A $\quad (a_{ij})_{i,j} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$

$A + B$ $\quad (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})_{i,j}$

$A \cdot B$ $\quad (a_{ij}) \cdot (b_{kl}) = (\sum a_{ij} b_{jl})_{i,l}$

$L(\mathbb{C}^n, \mathbb{C}^n) = B(\mathbb{C}^n)$ $\xleftrightarrow{\text{basis}}$ $M_{n\times n} = M_n$

Operator on $\mathbb{C}^n$ $\quad$ square matrix

.

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

① Addition: $A+B$

$A+B = B+A$ commutative

$0u = \vec{0}$ zero operator $\qquad A + 0 = A$,

② Multiplication: $A \cdot B$

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

$\begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$

Identity operator: $Iu = u$, $\qquad I = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$

$A \cdot I = A = I \cdot A$

③ A djoint: define A^* as $\qquad A = [a_{ij}] \qquad A^* = [\overline{a_{ji}}]$

$\langle A^* v, u \rangle = \langle v, Au \rangle$

$(A^*)^* = A$

So $A^* = \overline{(A^T)}$ transpose and conjugate

$(AB)^* = B^* A^* \qquad (A+B)^* = B^* + A^*$ e.g. $\begin{bmatrix} 1+i & 2-i \\ 4 & 3 \end{bmatrix}^* = \begin{bmatrix} 1+i & 4 \\ 2+i & 3 \end{bmatrix}$

①+② is called an algebra

①+②+③ is called an $*$-algebra

$B(\mathbb{C}^n)$ Algebra of (linear) operators on $\mathbb{C}^n$

M_n Algebra of $n\times n$ complex matrix

Other example of algebra? $\{u:\Omega\to\mathbb{C}\}$

Function algebras

with $fg=gf$!

Operators / Matrix with special property

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

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

$\Leftrightarrow \langle v, Av\rangle \in \mathbb{R}\ \forall\ v\in\mathbb{C}^n$ $\left(\overline{\langle v, Av\rangle}=\langle Av, v\rangle = \langle A^*v, v\rangle = \langle v, Av\rangle\right)$

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

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

$\Leftrightarrow \langle v, Av\rangle \geq 0\ \forall\ v\in\mathbb{C}^n$ $(\langle v, Av\rangle = \langle v, B^*Bv\rangle = \langle Bv, Bv\rangle \geq 0)$

e.g. $A=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}$ $\left\langle \begin{pmatrix}a\\ b\end{pmatrix}, \begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}\begin{pmatrix}a\\ b\end{pmatrix}\right\rangle = \left\langle \begin{pmatrix}a\\ b\end{pmatrix}, \begin{pmatrix}a\\ 2b\end{pmatrix}\right\rangle = \bar{a}a+2\bar{b}b$

$= |a|^2+2|b|^2\geq 0$

$A=\begin{bmatrix}5 & 1\\ 1 & 1\end{bmatrix}$ $\left\langle \begin{pmatrix}a\\ b\end{pmatrix}, \begin{bmatrix}5 & 1\\ 1 & 1\end{bmatrix}\begin{pmatrix}a\\ b\end{pmatrix}\right\rangle = 5|a|^2+\bar{a}b+a\bar{b}+|b|^2$

$= 4|a|^2+|a|^2+\bar{a}b+a\bar{b}+|b|^2$

$= 4|a|^2+|a+b|^2\geq 0.$

$A\geq 0 \Rightarrow A=A^*$ b/c $A=B^*B \Rightarrow A^*=B^*(B^*)^*=B^*B=A$

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

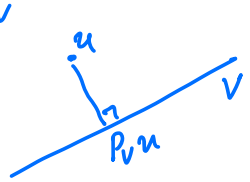
e.g. $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ or $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ or $P(u) = \langle v, u \rangle v$ for some $\|v\| = 1$ rank one projection

P projection $\Leftrightarrow \forall v$, $\langle Pv, (I-P)v \rangle = 0$

$v = Pv + (I-P)v$, $Pv \perp (I-P)v$

$V \subseteq \mathbb{C}^n$ a subspace ($\forall v, u \in V$, $\alpha v + \beta u \in V$)

$\Longleftrightarrow$ P_V s.t. $\forall u \in \mathbb{C}^n$, $\min_{v \in V} \|u - v\| = \|u - P_V u\|$



$V \perp W$ if $\forall v \in V, w \in W$, $\langle v, w \rangle = 0$

$\Leftrightarrow$ $P_V \cdot P_W = 0$ orthogonal

$V \subseteq W$ $\Leftrightarrow$ $P_V \cdot P_W = P_V$

4. U is a unitary if $U^* U = U U^* = I$

e.g. $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $Y = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$ $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ Pauli Matrix

U Unitary $\Leftrightarrow$ $\langle Uv, Uw \rangle = \langle v, w \rangle$ $\forall v \in \mathbb{C}^n$

$\langle v, U^* U w \rangle$, $U^* U = I$

Change of basis: Given an O.N.B $\{v_i\} \subseteq \mathbb{C}^n$ $\langle v_i, v_j \rangle = \delta_{i,j}$

$\{Uv_i\}$ is also O.N.B b/c $\langle Uv_i, Uv_j \rangle$

Given any two basis $\{v_i\}, \{w_i\} \subseteq \mathbb{C}^n$

$\exists!$ Unitary U s.t. $Uv_i = w_i$ $\forall i$

If U unitary. $\{e_i\} \to \{v_i\}$, then U^* unitary $U^* : \{v_i\} \to \{e_i\}$. $U^* = U^{-1}$

A operator is diagonalizable if $\exists$ some O.N.B $\{v_i\} \subseteq \mathbb{C}^n$ s.t.

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

A matrix $A=(a_{ij})$ is diagonalizable if $\exists$ U unitary

$$UAU^* = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad \text{diagonal matrix}$$

U is the unitary $\{e_i\} \longrightarrow \{v_i\}$

Spetrum Theorem : A is diganizable if and only if $AA^* = A^*A$

In particular, $A = \sum \lambda_i E_i$ where $\lambda_i \in \mathbb{C}$

E_i mutual orthogonal projections

s.t. $\sum_{i=1}^{k} E_i = I$

Remark: A satisfy $AA^* = A^*A$ is called a normal operator

What are λ_i and E_i,

Recall that $\lambda \in \mathbb{C}$ is an eigen value of A if $\exists$ $u \neq 0$

$$Au = \lambda u$$

Eigen space $V_\lambda = \{u \in \mathbb{C}^n \mid Au = \lambda u\}$ E_λ projection onto V_λ

$\text{Spec}(A) = \{\lambda \in \mathbb{C} \mid \lambda \text{ is an eigen value of } A\}$ finite set

If $A^*A = AA^*$, $A = \sum_{\lambda \in \text{Spec}(A)} \lambda E_\lambda$ for some E_λ with $E_\lambda \perp E_u$ $\forall \lambda \neq u \in \text{Spec}(u)$

In some basis, $UAU^* = \begin{bmatrix} \lambda_1 E_1 & & & \\ & \lambda_2 E_2 & & \\ & & \ddots & \\ & & & \lambda_n E_n \end{bmatrix} = \begin{bmatrix} \overbrace{\lambda_1 \cdots \lambda_1}^{\dim(E_1)} & & & \\ & \overbrace{\lambda_2 \cdots \lambda_2}^{\dim(E_2)} & & \\ & & \ddots & \\ & & & \lambda_n \cdots \lambda_n \end{bmatrix}$

Example: $\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} = 1 \cdot E_1 + (-2) E_2$ eigen vector $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ $E_2 = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$

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

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

(Why $\mathbb{C}$? Real matrix can have $\mathbb{C}$ eigenvalues)

Fuctional calculus

Now for a normal A, $A = U^* \begin{bmatrix} u_1 & & & \\ & u_2 & & \\ & & \ddots & \\ & & & u_n \end{bmatrix} U = \sum \lambda_i E_i$ (D_u)

$u_i \in \text{spec}(A)$

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

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

Justification. $A^2 = A \cdot A = U^* D_u \underbrace{U U^*}_{I} D_u U = U^* D_u^2 U = U^* \begin{bmatrix} u_1^2 & & \\ & \ddots & \\ & & u_n^2 \end{bmatrix} U$

$A^k = A \cdot A \cdots A = U^* D_u U \cdots U^* D_u U = U^* D_u U = U^* \begin{bmatrix} u_1^k & & \\ & \ddots & \\ & & u_n^k \end{bmatrix} U$

Thus for any polynomial $f(x) = a_k x^k + a_{k-1}x^{k-1} + \cdots a_1 x + a_0$

$$f(A) = a_k A^k + \cdots a_1 A + a_0 I$$

$$= a_k U^* \begin{bmatrix} u_1^k & & \\ & \ddots & \\ & & u_n^k \end{bmatrix} U + \cdots a_1 U^* \begin{bmatrix} u_1 & & \\ & \ddots & \\ & & u_n \end{bmatrix} U + a_0 I$$

$$= U^* \begin{bmatrix} a_k u_1^k & & \\ & \ddots & \\ & & a_k u_n^k \end{bmatrix} U + \cdots U^* \begin{bmatrix} a_1 u_1 & & \\ & \ddots & \\ & & a_1 u_n \end{bmatrix} + U^* \begin{bmatrix} a_0 & & \\ & \ddots & \\ & & a_0 \end{bmatrix} U$$

$$= U^* \begin{bmatrix} f(u_1) & & \\ & \ddots & \\ & & f(u_n) \end{bmatrix} U \quad \text{for any polynomial}$$

Now for general $f: \mathrm{spec}(A) \to \mathbb{C}$.

$\exists$ polynomial $P_k \to f$ uniformly

$$P_k(A) \longrightarrow f(A)$$

$$\| \qquad\qquad \|$$

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

Example: $f(x) = e^x$ $\quad e^A = \sum_{k=0} \frac{1}{k!} A^k$ $\quad X = \begin{bmatrix} 0 & t \\ -t & 0 \end{bmatrix}$ $\quad e^X = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}$

$f(x) = \sqrt{x}$, $x \geq 0$. For $A \geq 0$, $\sqrt{A}$:= unique positive operator $\sqrt{A}$. $\quad A^p$

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

General A, $\quad |A| = \sqrt{A^* A}$ $\quad$ (Note that $A^* A \neq A A^*$)

$\forall\ f, g : \mathrm{spec}(A) \to \mathbb{C}, \quad f(A)\,g(A) = fg(A) = g(A)\,f(A)$

$f(t) = \log t \quad t>0$ For $A>0$, $\log A = U^* \begin{bmatrix} \log u_1 & & \\ & \ddots & \\ & & \log u_n \end{bmatrix} U$

$f(t) = t\log t \qquad A\log A = U^* \begin{bmatrix} u_1 \log u_1 & & \\ & \ddots & \\ & & u_n \log u_n \end{bmatrix} U = A \cdot \log A.$

In general, $AB \neq BA$. When $AB = BA$ can happen?

$A = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad B = \begin{bmatrix} u_1 & & \\ & \ddots & \\ & & u_n \end{bmatrix}$ diagonal matrix

$AB = BA$ if and only if $\exists$ unitary U

$A = U D_1 U^* \qquad B = U D_2 U^*$

D_1, D_2 diagonal matrix

Dirac's Bra-ket Notation

$u \in \mathbb{C}^n \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \longrightarrow |u\rangle$ "ket"

$v \in (\mathbb{C}^n)^* \quad v = (\bar{v}_1, \bar{v}_2 \cdots \bar{v}_n) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}^* \longrightarrow \langle v|$ "bra"

$(\mathbb{C}^n)^* = L(\mathbb{C}^n, \mathbb{C})$ a bracket $\langle v|u\rangle = \langle v, u\rangle$ inner product

$$= (\bar{v}_1 \cdots \bar{v}_n) \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \sum_{i=1}^{n} \bar{v}_i u_i$$

$$(\bar{v}_1 \cdots \bar{v}_n) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}^* \quad \text{so} \quad (|u\rangle)^* = \langle u| \qquad |u\rangle \in L(\mathbb{C}, \mathbb{C}^n), \ \langle u| \in L(\mathbb{C}^n, \mathbb{C})$$

$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \cdots$ $\quad \mathbb{C}^n = \{ a_0|0\rangle + a_1|1\rangle + \cdots a_{n-1}|n-1\rangle \mid a_i \in \mathbb{C} \}$

How it works:

① Inner product : $\langle v|u\rangle$

② Matrix action : $\langle v|A|u\rangle = \langle v, Au\rangle$

③ Rank one operator : $E_{vu} = |v\rangle\langle u|$ outer bracket

$E_{v,u}(|w\rangle) = (|v\rangle\langle u|)|w\rangle$

$= |v\rangle\langle u|w\rangle$

check linearity $= (\langle u|w\rangle)|v\rangle$ $\quad \langle u|w\rangle \in \mathbb{C}$

④ General operator : $A = \sum a_j |v_j\rangle\langle u_j|$ for $|v_j\rangle\langle u_j|$ e.g. $A = [a_{ij}]$

$= \sum_{ij} a_{ij} |i\rangle\langle j|$ ↓ standard basis

⑤ Matrix multiplication : $A = \sum a_j |v_j\rangle\langle u_j|$

$B = \sum b_k |w_k\rangle\langle x_k|$

$$A \cdot B = \sum_{j,k} a_j b_k |v_j\rangle\langle u_j|w_k\rangle\langle x_k|$$

$$= \sum_{j,k} a_j b_k \langle u_j|w_k\rangle |v_j\rangle\langle x_k| \qquad \langle u_j|w_k\rangle \in \mathbb{C}$$

⑤ diagonal operator : $A = \sum \lambda_j |V_j\rangle\langle V_j|$ $\{|V_i\rangle\}$ O.N.B of eigenvalues of A

orthogonal decomposition λ_j eigenvalue of $|V_j\rangle$

Spetrum theorem

A Positive $\Rightarrow \lambda_j \geq 0$ A projection $\Rightarrow \lambda_j \in \{0,1\}$

A Hermitian $\Rightarrow \lambda_j \in \mathbb{R}$ A unitary $\Rightarrow |\lambda_j| = 1 \quad \forall j$

Postulates: Working Assumption / Framework / Model

Watch: Feynman — Difference between Mathematics and Physics.

Real World $\xrightarrow{\text{Poseulates}}$ Quantum Mechanic using Mathematics

↑ Experiment / theory

Postulate 1. (State space) Any isolated (quantum) system is associated a $\mathbb{C}$-Hilbertspace as its "state space". The state of system is given by a unit vector $|\varphi\rangle$, $\| |\varphi\rangle \| = 1$

called the "state vector".

Example (Qubit)

$\mathbb{C}^2 = \{ a|0\rangle + b|1\rangle \mid a, b \in \mathbb{C} \}$ $|a|^2$ prob observing $|\varphi\rangle$ in $|0\rangle$.

logic basis $|0\rangle, |1\rangle$ Other basis. $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$, $|-\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$

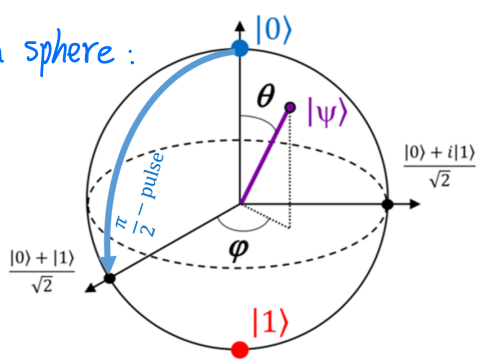
(superposition : $|\varphi\rangle = a|0\rangle + b|1\rangle$ unit / state $\Leftrightarrow |a|^2+|b|^2 = 1$

$|\varphi\rangle = e^{i\delta}\cos(\theta/2)|0\rangle + e^{i(\delta+\phi)}\sin(\theta/2)|1\rangle$ Bloch sphere:

$0 \leq \delta \leq 2\pi \quad 0 \leq \theta \leq \pi \quad 0 \leq \phi \leq 2\pi$

$e^{i\delta}$ global phase not physical relevalent

ϕ relative phase physical



$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$$

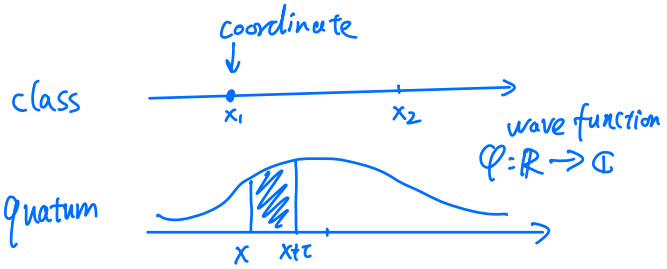
Example: (1-dim Wave function)

Position of a particle 0, $x_0 \in \mathbb{R}$

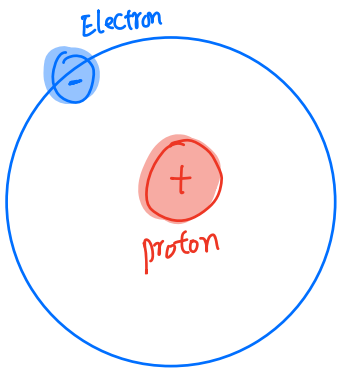
prob. particle appear in $(a.b)$

$$P(0 \in (a.b)) = \int_a^b |\varphi(x)|^2 dx$$

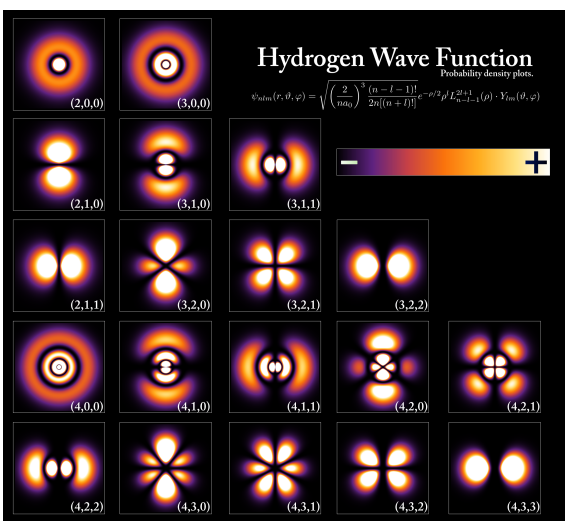
$$\|\varphi\| = \int_{-\infty}^{\infty} |\varphi(x)|^2 dx = 1 \longrightarrow P((-\infty, \infty)) = 1$$



Chemistry



Quantum Mechanic Model



Young Double split, Single photon Inference

Bell inequality

Postulate 2. (Measurement)

Quantum Information Form (fid.)

Every quantum measurement is given by a collection $\{E_m\}$ of Measurement Operator satisfying the completeness equation

$$\sum_m E_M = I \qquad E_M \geq 0$$

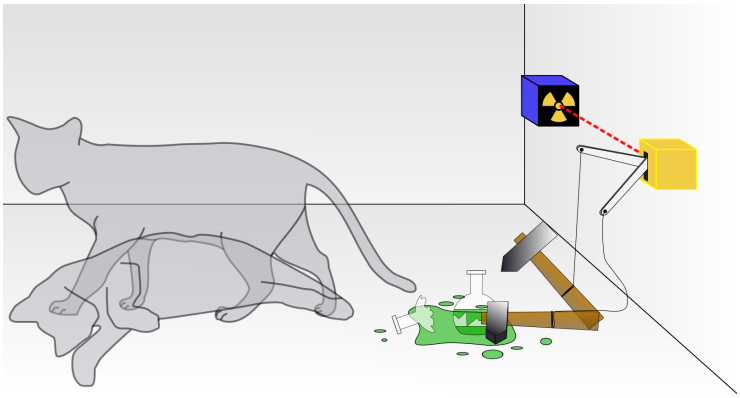
$\{E_m\}$ is called a POVM (Positive Operator-Valued Measurement)

Prob seeing outcome m: $\langle \varphi | E_M | \varphi \rangle$ prob. distr

$$\sum_m \langle \varphi | E_M | \varphi \rangle = 1 \quad \forall |\varphi\rangle \Rightarrow \sum E_M = I$$

Pose measurement state $\dfrac{E_M|\varphi\rangle}{\|E_M|\varphi\rangle\|} = \dfrac{E_M|\varphi\rangle}{\sqrt{\langle\varphi|E_M|\varphi\rangle}}$

Example. Schrödinger's cat



Picture from Wikipedia by Dhatfield.

$|\varphi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$

$E_0 = |0\rangle\langle 0| \quad E_1 = |1\rangle\langle 1|$

After measurement

$\frac{1}{2}$ prob $E_0|\varphi\rangle = \frac{1}{\sqrt{2}}|0\rangle$

$E_1|\varphi\rangle = \frac{1}{\sqrt{2}}|1\rangle$

Projection Value Measurement (PVM): $\{P_m\}$ $\sum P_m = 1$

von Neumann

P_m . mutually orthogonal projection

Example: Let $\{|v_i\rangle\}$ be a O.N.B $\{E_i = |v_i\rangle\langle v_i|\}$ PVM

$|\varphi\rangle = a|0\rangle + b|1\rangle$ $|a|^2$-prob seeing $|\varphi\rangle$ in $|0\rangle$

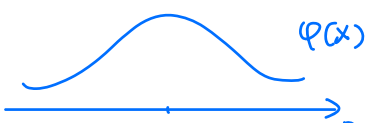
Quantum Mechanics version

Every Observable corresponds to a Hermitian operator. (∞-dim)

The only value that will obsered are eigen values.

Example: Position operator : $X\ \varphi(x) = x\varphi(x)$

$e^{-\frac{x^2}{2}} \to xe^{-\frac{x^2}{2}}$



Expected value of position $\langle\varphi|X|\varphi\rangle = \int_{-\infty}^{\infty} X\,|\varphi(x)|^2 dx$

Momentum ... : $P = -i\hbar\frac{\partial}{\partial x}$

$\hbar$ Planck constant

Kinetic Energy : $H = -\frac{\hbar^2}{2m}\frac{\partial}{\partial x^2}$ free Hamiltonian

mass

In fd. $A = A^*$ Hermitian $\Rightarrow$ $A = \sum \lambda_i E_i$ $\{E_i\}$ PVM

$\lambda_i \in \mathbb{R}$ all physical quantity are real.

$|\varphi_i\rangle$ eigen vector of $\lambda_i \Rightarrow \langle\varphi_i|A|\varphi_i\rangle = \lambda_i$

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

$$|\varphi\rangle \longrightarrow U|\varphi\rangle$$

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

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

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

Why unitary. $|\varphi\rangle \longrightarrow |\psi\rangle=U|\varphi\rangle$ So $\||U|\varphi\rangle\|=\||\varphi\rangle\|$ $\forall|\varphi\rangle$

$\||\varphi\|=1$ $\||\psi\rangle\|=1$ $\Downarrow$ unitary.

Postulate 3' (Continous time). The time evolution of a closed syseem is given Schrödinger equation:

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

(H: Hamiltonian)

$\langle\psi|H|\psi\rangle$ expected energy. ① $H(t)$ time dependent

② $H(t)\equiv H$ time independent

② $|\psi(t)\rangle=e^{iHt}|\psi(0)\rangle$ H hermitian $\Rightarrow$ e^{iHt} unitary.

Example: $H=\sum_E E|\varphi_E\rangle\langle\varphi_E|$ $|\varphi_E\rangle$ energy basis.

$e^{iHt}|E\rangle=e^{iEt}|E\rangle$ $|\varphi\rangle=\sum a_E|E\rangle$ $e^{iHt}=\sum a_E e^{iEt}|\varphi_E\rangle$

Postulate 1. Quantum System $\longrightarrow$ $\mathbb{C}^n$

State $\longrightarrow$ $|\varphi\rangle$ unit vector

Postulate 2. Measurement $\longrightarrow$ POVM $\sum_m E_m = I \quad E_m \geq 0$

Postulate 3. Evolution $\longrightarrow$ unitary $U: |\varphi\rangle \longrightarrow U|\varphi\rangle$

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

Tensor product space

$V = \mathbb{C}^X \quad W = \mathbb{C}^Y$

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

$\varphi \otimes \psi (x,y) = \varphi(x)\psi(y)$

$|\varphi\rangle \otimes |\psi\rangle$

$V \otimes W = \text{span}\{|\varphi\rangle \otimes |\psi\rangle \mid |\varphi\rangle \in V, |\psi\rangle \in |\psi\rangle\}$

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

$(a|\varphi_1\rangle \otimes |\psi\rangle + \beta|\varphi_2\rangle \otimes |\psi\rangle) = (a|\varphi_1\rangle + \beta|\varphi_2\rangle) \otimes |\psi\rangle$

$(\langle\varphi_1| \otimes \langle\varphi_2|)(|\psi\rangle_1 \otimes |\psi_2\rangle) = \langle\varphi_1|\psi_1\rangle \langle\varphi_2|\psi_2\rangle$

Given $\{|v_i\rangle\} \leq V \quad \{|w_j\rangle\} \leq W$ O.N.B

$\{|v_i\rangle \otimes |w_j\rangle \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ O.N.B of $V \otimes W$

So $\dim(V \otimes W) = \dim(V) \dim W$

Now back to state space

Example (product state) $V_1 = V_2 = \mathbb{C}^2$.

$|0\rangle \otimes |1\rangle = |01\rangle \qquad |1\rangle \otimes |0\rangle = |10\rangle$

$|0\rangle \otimes |0\rangle = |00\rangle \qquad |1\rangle \otimes |1\rangle = |11\rangle$

$|\varphi\rangle = a|0\rangle + b|1\rangle \quad |\psi\rangle = \alpha|0\rangle + \beta|1\rangle$

$|\varphi\rangle \otimes |\psi\rangle = a\alpha|00\rangle + a\beta|01\rangle + b\alpha|10\rangle + b\beta|11\rangle$

($|\varphi\rangle$: state of 1st system; $|\psi\rangle$: state of 2nd system)

Example (Entangled state / non product state)

$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle |1\rangle)$

$= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$

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

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

$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ O.N.B of $\mathbb{C}^2 \times \mathbb{C}^2 \cong \mathbb{C}^4$

$|\Phi^+\rangle \neq |\varphi\rangle \otimes |\psi\rangle$ for any $|\varphi\rangle \otimes |\psi\rangle$, called entangled state.

super position of produse seate.

Operations on Tensor produce system

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

Define $(A \otimes B)(|\varphi\rangle \otimes |\psi\rangle) = A|\varphi\rangle \otimes B|\psi\rangle$. $A \otimes B \in B(\mathbb{C}^n \otimes \mathbb{C}^m)$

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

Product measurement

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

$(A\otimes B)^* = A^*\otimes B^* = A\otimes B$ is Hermitian in $B(\mathbb{C}^n\otimes\mathbb{C}^m)$

$\langle\varphi|\otimes\langle\psi|\,(A\otimes B)\,|\varphi\rangle\otimes|\psi\rangle = \langle\varphi|A|\varphi\rangle\,\langle\psi|B|\psi\rangle$

Expectation of observable A given φ ... B given ψ

PoVM $\sum_m E_m = I_1$ $\sum_n F_n = I_2$ $\sum_{m,n} E_m\otimes F_n = I_1\otimes I_2$

$E_m \geq 0$, $F_n \geq 0 \Rightarrow E_m\otimes F_n \geq 0$

prob of $|\varphi\rangle\otimes|\psi\rangle$ measured in outcome (m,n)

$\langle\varphi|\otimes\langle\psi|\,E_m\otimes F_n\,|\varphi\rangle\otimes|\psi\rangle = \langle\varphi|E_m|\varphi\rangle\,\langle\psi|F_n|\psi\rangle$

produce probability

Partial Measurement : $A\in B(\mathbb{C}^n)$ $\langle\varphi|\otimes\langle\psi|\,A\otimes I\,|\varphi\rangle\otimes|\psi\rangle$

$= \langle\varphi|A|\varphi\rangle\langle\psi|\psi\rangle = \langle\varphi|A|\varphi\rangle$

or $\sum_m E_m = I$ $\langle\varphi|\langle\psi|(E_m\otimes I)|\psi\rangle|\varphi\rangle = \langle\varphi|E_m|\varphi\rangle$

Product state ~ independent measurement outcome

Entangled state ~ corrolated measurement outcome

Example: $E_0 = |0\rangle\langle 0|$ $E_+ = |+\rangle\langle +|$ $|\Phi^+\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$

$E_1 = |1\rangle\langle 1|$ $E_- = |-\rangle\langle -|$

$\langle\Phi^+|\,E_0\otimes I\,|\Phi^+\rangle = \frac{1}{2} = \langle\Phi^+|\,I\otimes E_0\,|\Phi^+\rangle$

$\langle\Phi^+|\,E_1\otimes I\,|\Phi^+\rangle = \frac{1}{2} = \langle\Phi^+|\,I\otimes E_1\,|\Phi^+\rangle$

$\langle\Phi^+|\,E_i\otimes E_j\,|\Phi^+\rangle = \begin{cases}\frac{1}{2} & \text{if } i=j\\ 0 & \text{if } i\neq j\end{cases}$ not independent on i,j

$\langle\Phi^+|\,E_i\otimes E_\pm\,|\Phi^+\rangle = 0$ for $i=1,2$, $+$ or $-$

$$\langle \Phi^+ | A \otimes I | \Phi^+ \rangle = \frac{1}{2}\langle 0|A|0\rangle + \frac{1}{2}\langle 1|A|1\rangle$$

$$\neq \langle \varphi|A|\varphi\rangle \quad \forall \ |\varphi\rangle \in \mathbb{C}^2$$

The observation of $|\Phi^+\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ on the first system doesnot match any vector state. Does this violates Postulate 1?

No. B/C A closed system has a vector state $|\varphi\rangle$

The state of a general system is given by a density operator.

A vector state $|\varphi\rangle \in \mathbb{C}^n$ is also called pure state

In general, a quantum system can have mixed state. given by

$\{p_i, |\psi_i\rangle\}$ ensemble of pure states. p_i is the prob. system in $|\psi_i\rangle$

Then $\rho \equiv \sum p_i |\psi_i\rangle\langle\psi_i|$ is the density operator

Example: $\rho = |\varphi\rangle\langle\varphi|$. $\lambda|0\rangle\langle 0| + (1-\lambda)|1\rangle\langle 1|$ $u|+\rangle\langle +| + (1-u)|-\rangle\langle -|$. $\sum p_i |i\rangle\langle i|$

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

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

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

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

$$p(m) = \mathrm{tr}(\rho E_m)$$

to be outcome E_m.

(Observable) The expected value of an observable $A = A^\dagger$ given state ρ is $\mathrm{tr}(\rho A)$

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

$$\rho \to U \rho U^*$$

Density Matrix / Operator

Last time: composite system $\mathbb{C}^m \otimes \mathbb{C}^n$ $\quad |\varphi\rangle = \mathbb{C}^n \times \mathbb{C}^m$ a vector state

Do a partial measurement $A = A^* \in B(\mathbb{C}^m)$. $\quad \langle\varphi| A \otimes I |\varphi\rangle$.

E.g. $|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$ $\quad \langle\Phi^+| A \otimes I |\Phi^+\rangle = \frac{\langle 0|A|0\rangle}{2} + \frac{\langle 1|A|1\rangle}{2}$ (*)

$\mathbb{C}^2 \otimes \mathbb{C}^2$

$\neq \langle\varphi|A|\varphi\rangle$ for some $|\varphi\rangle \in \mathbb{C}^2$?

No. Any $\alpha|1\rangle + \beta|0\rangle = |\varphi\rangle$ $\quad \langle\varphi|A|\varphi\rangle = |\alpha|^2 \underbrace{\langle 0|A|0\rangle}_{a_{00}} + |\beta|^2 \underbrace{\langle 1|A|1\rangle}_{a_{01}}$

$+ \alpha\bar{\beta} \underbrace{\langle 1|A|0\rangle}_{a_{10}} + \beta\bar{\alpha} \underbrace{\langle 0|A|1\rangle}_{a_{11}}$

$$A = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix}$$

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

$\{p_i, |\psi_i\rangle\}$ ensemble of pure states. p_i is the prob. system in $|\psi_i\rangle$

E.g. $\langle\Phi^+|A \otimes I|\Phi^+\rangle = \frac{1}{2}\langle 0|A|0\rangle + \frac{1}{2}\langle 1|A|1\rangle \rightsquigarrow \{(\frac{1}{2}, |0\rangle), (\frac{1}{2}, |1\rangle)\}$

Vector space $|\varphi\rangle \rightsquigarrow \{1, |\varphi\rangle\}$ single ensemble

Given an ensemble $\{p_i, |\varphi_i\rangle\}$, measurements

Expected value of $A = A^*$: $\sum_i p_i \langle\varphi_i|A|\varphi_i\rangle$

POVM $\{E_m\}$: $\sum_i p_i \langle\varphi_i|E_m|\varphi_i\rangle$ prob of outcome m

Unitary transformation $\{p_i, |\varphi_i\rangle\} \xrightarrow{U} \{p_i, U|\varphi_i\rangle\}$

However, in terms of measurement the ensemble representation is not unique

For any $A = A^*$ observable, $\frac{1}{2}\langle +|A|+\rangle + \frac{1}{2}\langle -|A|-\rangle = \frac{1}{2}\langle 0|A|0\rangle + \frac{1}{2}\langle 1|A|1\rangle$

So from physic measurement, we can not distinguish $\{(\frac{1}{2}, |0\rangle), (\frac{1}{2}, |1\rangle)\}$

and $\{(\frac{1}{2}, |+\rangle), (\frac{1}{2}, |-\rangle)\}$

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

$\varphi: B(\mathbb{C}^n) \to \mathbb{C}$, $\varphi(A) = \langle \vec{x}^\dagger | A \otimes I | \vec{x}^\dagger \rangle$

φ is linear: So $\varphi \in L(B(\mathbb{C})^n, \mathbb{C}) = B(\mathbb{C})^*$

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

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

$$V^* \cong \mathbb{C}^n = \{ (v_1 \cdots v_n) \mid v_i \in \mathbb{C} \}$$

$(v_1 \cdots v_n) : \mathbb{C}^n \to \mathbb{C}$

$(v_1 \cdots v_n) \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \sum_{i=1}^n v_i u_i$ linear functional

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

There exists $\{e_i^*\} \subseteq V^*$ dual basis s.t.

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

Indeed $e_i^* = (1, 0 \cdots 0)$.

Thm: $V \cong V^*$ as $\mathbb{C}$ vector space if $\dim(V) < \infty$.

Now consider $V = B(\mathbb{C})^n \cong M_n$. What is V^*?

Recall the trace functional:

$$tr((a_{ij})) = \sum a_{ii}$$

equivalently $tr(A) = \sum_i \langle e_i | A | e_i \rangle$

① Independence of basis

② Tracial Property.

(Tracial property) ① $\forall A, B \in M_n(\mathbb{C})$ $tr(AB) = tr(BA)$

② $\forall U$ unitary, $tr(U^* A U) = tr(A)$

③ $tr(A) = \sum_i \langle \varphi_i | A | \varphi_i \rangle$ for any O.N.B. $\{\varphi_i\}$.

Proof: ① $A = (a_{ij})$ $B = (b_{ij})$

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

$BA = \left(\sum_l b_{il} a_{lj}\right)_{ij}$

$tr(AB) = \sum_{i,k=1}^{n} a_{ik} b_{ki}$

$tr(BA) = \sum_{i,l=1}^{n} b_{il} a_{li}$

Lemma: $M_n \cong M_n^*$ by the following bijection

$f \in M_n^* \longleftrightarrow$ a operator X_f s.t. $f(A) = tr(AX)$.

How to see it in an elementary way?

Consider $\{E_{ij} = |i\rangle\langle j|\} \subseteq M_n$ basis

$\{f_{ij}(e_{kl}) = \delta_{(ij)(kl)}\}$ is M_n^* dual basis

$f_{ij} \longleftrightarrow E_{ji} \in M_n$ basis $f_{ij}(A) = tr(A\,|j\rangle\langle i|)$

$= tr(\sum a_{kl}\,|k\rangle\langle l|\,|j\rangle\langle i|)$

$= a_{ij}$

Span $\{E_{ji}\} = M_n \cong M_n^*$

Now for $|\varphi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$,

$\varphi(A) = \langle\varphi|A\otimes I|\varphi\rangle$ corresponds to a operator ρ. s.t. $\varphi(A) = tr(A\rho)$

What property φ have?

① if $A \geq 0$, $A \otimes I \geq 0$, then $\varphi(A) \geq 0$

② $\varphi(I) = \langle\varphi|I\otimes I|\varphi\rangle = \langle\varphi|\varphi\rangle = 1$

A linear function satisfy ① ② is called a seate.

What property should the operator ρ have?

① $tr(\rho A) = \langle\varphi|A\otimes I|\varphi\rangle \geq 0 \Rightarrow \rho \geq 0$ (choose $A = |h\rangle\langle h|$, $|h\rangle \in \mathbb{C}^n$)

② $tr(\rho \cdot I) = tr(\rho) = 1$

$\rho \geq 0 \Rightarrow \rho = \sum p_i |\psi_i\rangle\langle\psi_i|$, $p_i \geq 0$ (by orthogonal decomposition)

$tr(\rho) = 1 \Rightarrow \sum p_i = 1$ (by basis independence of trace)

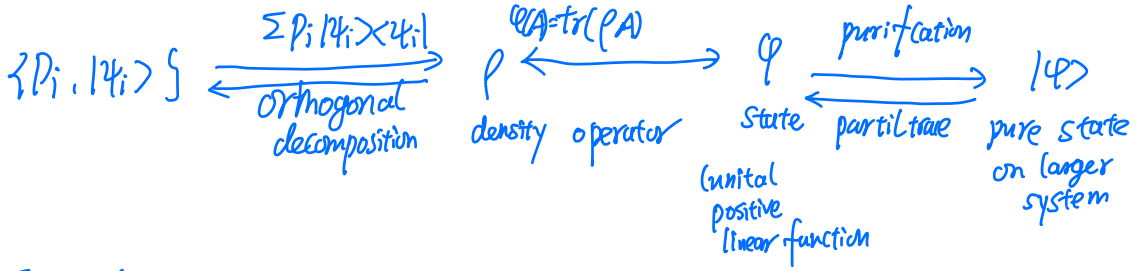
A state of a quantum system $\mathbb{C}^n$ can be equivalently described by one of the following

① An ensemble of pure state $\{p_i, |\psi_i\rangle\}$ $\sum p_i = 1$, $p_i \geq 0$ $|\psi_i\rangle \in \mathbb{C}^n$

② A density operator $\rho \in B(\mathbb{C}^n)$ s.t. $\rho \geq 0$, $\mathrm{tr}(\rho) = 1$

③ A linear functional $\varphi: B(\mathbb{C})^n \to \mathbb{C}$ s.t. $\varphi(I) = 1$ and $\varphi(A) \geq 0$ if $A \geq 0$

④ A state vector $|\psi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$ for some m



Examples

① Pure state = vector state: $|\psi\rangle \leftrightarrow \varphi(A) = \langle\psi|A|\psi\rangle \leftrightarrow \rho = |\psi\rangle\langle\psi|$ density operator

② A mixed state $\rho = \sum p_i |\psi_i\rangle\langle\psi_i|$ $\{|\psi_i\rangle\}$ orthonormal set

$\sum p_i = 1$ $p_i \geq 0$

Mixed state are convex combination of pure state

If $p_i = 1$, $p_j = 0$ $\forall j \neq i$ $\Rightarrow$ $\rho = |\psi_i\rangle\langle\psi_i|$ pure state

For $\mathbb{C}^n$, $p_i = \frac{1}{n}$, $\rho = \sum \frac{1}{n} |\psi_i\rangle\langle\psi_i| = \frac{1}{n} I$, completely mixed state

like uniform distribution $(\frac{1}{n}, \frac{1}{n}, \cdots \frac{1}{n})$

③ Flat state: P projection. $\rho = \frac{P}{\mathrm{tr}(P)} = \frac{1}{k} \sum_{i=1}^{k} |\psi_i\rangle\langle\psi_i|$ $\{|\psi_i\rangle\}$ O.N.B of $\mathrm{Ran}(P)$.

$k = \mathrm{tr}(P) \in \mathbb{N}$

④ Ensemable of pure states $\{p_i, |\psi_i\rangle\} \to \rho = \sum p_i |\psi_i\rangle\langle\psi_i|$

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

Product state

Let $\rho \in D(\mathbb{C}^n)$ and $\sigma \in D(\mathbb{C}^m)$. Then $\rho \otimes \sigma \in D(\mathbb{C}^{nxm})$

Fact: $tr(A \otimes B) = tr(A)\,tr(B)$.

Then $tr(\rho \otimes \sigma) = tr(\rho)\,tr(\sigma) = 1$

$\rho \geq 0$, $\sigma \geq 0 \Rightarrow \rho \otimes \sigma \geq 0$

Joint States / density operator

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

Examples: 1. Product state $\rho \otimes \sigma$ $\quad \rho \in D(H_A)$ $\quad \sigma \in D(H_B)$

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

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

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

2. Separable state $\omega = \sum_i \lambda_i \rho_i \otimes \sigma_i$ $\quad \sum \lambda_i = 1$ $\lambda_i \geq 0$

$\rho_i \in D(H_A)$, $\sigma_i \in D(H_B)$

$\omega = \sum \lambda_i |i\rangle\langle i| \otimes \sigma_i$ $\quad \sigma_i \in D(H_B)$, $\sum \lambda_i = 1$, $\lambda_i \geq 0$ $\quad$ classical-quantum state

3. Pure joint state: $|\varphi\rangle \in H_A \otimes H_B$ unit vector

$\varphi = |\varphi\rangle\langle\varphi|$ is the joint density

e.g. $|\varphi\rangle = |0\rangle \otimes |0\rangle$ $\quad \varphi = |00\rangle\langle 00| = |0\rangle\langle 0| \otimes |0\rangle\langle 0|$ product state

$|\varphi\rangle = \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}}$

$$\varphi = |\varphi\rangle\langle\varphi| = \frac{1}{2}\left(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\right)$$

Using $|00\rangle \to |1\rangle$ $|01\rangle \to |2\rangle$ $|10\rangle \to |3\rangle$ $|11\rangle \to |4\rangle$

$$\begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

Is this a seperable state?

Definition: A joint state ρ_{AB} is called entangled if it is not seperable

Important concept! A lot more to explore later.

Marginal seate / Reduced density

Given a joint state $\rho_{AB} \in B(H_{AB})$. it induce a seate on A.

$\varphi_A : B(H_A) \to \mathbb{C}$. $\varphi_A(X) = \operatorname{tr}_{AB}((X_A \otimes I_B)\rho_{AB})$

$= \operatorname{tr}_A(X_A \rho_A)$ for some ρ_A

$\rho_A = Id \otimes \operatorname{tr}_B(\rho_{AB})$

e.g. $\rho_{AB} = \sum a_{ij,kl}\, e_{ij} \otimes e_{kl}$

$id_A \otimes \operatorname{tr}_B(\rho_{AB}) = \sum a_{ij,kl}\, Id(e_{ij}) \otimes \operatorname{tr}_B(e_{kl})$

$= \sum_{k,ij} a_{ij,kk}\, e_{ij}$

For example: ① $W_{AB} = \rho_A \otimes \sigma_B \Rightarrow W_A = id_A \otimes \operatorname{tr}_B = \rho_A \operatorname{tr}_B(\sigma_B) = \rho_A$

$W_B = \sigma_B$

② $W_{AB} = \sum \lambda_i |i\rangle\langle i| \otimes \sigma_i$ $W_A = \sum \lambda_i |i\rangle\langle i|$ $W_B = \sum \lambda_i \sigma_i$

③ $W_{AB} = |\varphi\rangle\langle\varphi|$. $|\varphi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$ $W_A = W_B = \frac{1}{2}$

Purification.

Given a mixed state $\rho \in D(\mathcal{H}_A)$

① Does there exises a joint state ω_{AB} such that $\omega_A = \rho$?

② . . - - - - - a pure joint state $\varphi_{AB} = |\varphi\rangle\langle\varphi|$ such that $\varphi_A = \rho$?
Such $|\varphi\rangle$ is called a purification of ρ.

① Yes. $\omega_{AB} = \rho_A \otimes \sigma_B$

② Given $\rho = \sum p_i |\varphi_i\rangle\langle\varphi_i|$. Define $|\varphi\rangle_{AA'} = \sum_i \sqrt{p_i}\, |\varphi_i\rangle_A \otimes |i\rangle_{A'}$ $\quad A \cong A'$

Then $\varphi = |\varphi\rangle\langle\varphi|$ has reduced density ρ on A.

$$id_A \otimes tr_{A'}(|\varphi\rangle\langle\varphi|) = id_A \otimes tr_{A'}\left(\sum_{ij} \sqrt{p_i}\sqrt{p_j}\, |\varphi_i\rangle\langle\varphi_j| \otimes |i\rangle\langle j|\right)$$
$$= \sum_i p_i |\varphi_i\rangle\langle\varphi_i| = \rho.$$

Theorem (Ulmann): Let $|\psi\rangle_{AC}$ be any purification of ρ. Then there exises a isometry $V \in L(\mathcal{H}_{A'}, \mathcal{H}_C)$ such that $V|\varphi\rangle = |\psi\rangle$.

Last time: $|\varphi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$ induce $\varphi \in B(\mathbb{C}^n)^* \cong \mathcal{L}(B(\mathbb{C}^n), \mathbb{C})$ $= \mathcal{M}_n^*$

$\varphi(A) = \langle\varphi| A \otimes I |\varphi\rangle$

Fact: $\forall f \in \mathcal{M}_n^*$, $\exists$ operator $\rho \in \mathcal{M}_n$ s.t. $f(A) = tr(A\rho)$

$\varphi \in \mathcal{M}_n^*$ is called a linear functional of $\mathcal{M}_n$

What property φ have $\quad \rho \in B(\mathbb{C}^n)$, $\varphi(A) = tr(A\rho)$

① $\varphi(I) = \langle\varphi| I \otimes I |\varphi\rangle = \langle\varphi|\varphi\rangle = 1 \iff tr(\rho I) = tr(\rho) = 1$

called, **unital** $\qquad$ trace 1

② If $A \geq 0$, $\varphi(A) = \langle\varphi| A \otimes I |\varphi\rangle \geq 0 \iff \forall A \geq 0 \ tr(\rho A) \geq 0$

called **positive** $\qquad$ $\rho \geq 0$

Lemma $tr(\rho A) \geq 0$ for $\forall A \geq 0 \iff \rho \geq 0$

Pf: (See Homework)

$\varphi \in \mathcal{M}_n^*$ is called a state $\qquad$ $\rho \in \mathcal{M}_n$ is a density operator

if φ is positive and unital $\qquad$ if $\rho \geq 0$ and $tr(\rho) = 1$

$S(\mathcal{M}_n) = \{\varphi \in \mathcal{M}_n |$ φ positive unital$\}$ $\underset{1\text{ to }1}{\cong}$ $D(\mathcal{M}_n) = \{\rho \in \mathcal{M}_n \,|\, \rho \geq 0, tr\rho\}$

$S(\mathcal{M}_n) \ni \varphi \longrightarrow \exists!\ d_\varphi \in \mathcal{M}_n$ s.t. $\varphi(A) = tr(A d_\varphi)$

$D(\mathcal{M}_n) \ni \rho \longrightarrow \varphi_\rho(A) := tr(A\rho)$

Fact. $D(\mathcal{M}_n) \cong S(\mathcal{M}_n)$ is a convex set.

If $\rho_i \in D(\mathcal{M}_n)$, $\sum_{\rho_i \geq 0} p_i = 1$, then $\sum \lambda_i \rho_i \in D(\mathcal{M}_n)$

A state of a quantum system $\mathbb{C}^n$ can be equivalently described by one of the following

① A state $\varphi: B(\mathbb{C}^n) \to \mathbb{C}$, i.e. $\varphi(I)=1$ and $\varphi(A)\geq 0$ if $A\geq 0$

② A density operator $\rho \in B(\mathbb{C}^n)$ s.t. $tr(\rho)=1$, $\rho \geq 0$

③ An ensemble of pure state $\{p_i, |\psi_i\rangle\}$ $\sum p_i = 1$. $p_i \geq 0$ $|\psi_i\rangle \in \mathbb{C}^n$

④ A state vector $|\varphi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$ for some m

$\{p_i, |\psi_i\rangle\}$ $\underset{\text{orthogonal decomposition}}{\overset{\rho=\sum p_i |\psi_i\rangle\langle\psi_i|}{\rightleftarrows}}$ ρ (density operator) $\overset{\varphi(A)=tr(\rho A)}{\longleftrightarrow}$ φ (state; unital positive linear function) $\underset{\varphi(A)=\langle\varphi|A\otimes I|\varphi\rangle}{\overset{\text{purification}}{\rightleftarrows}}$ $|\varphi\rangle$ (pure state on larger system)

Examples

① Pure state = vector state: $|\varphi\rangle \leftrightarrow \varphi(A)=\langle\varphi|A|\varphi\rangle \leftrightarrow \rho=|\varphi\rangle\langle\varphi|$ density operator

② A mixed state $\rho=\sum p_i |\varphi_i\rangle\langle\varphi_i|$ $\{|\varphi_i\rangle\}$ orthonormal set

$\sum p_i = 1$ $p_i \geq 0$

Mixed state are convex combination of pure state

If $p_i = 1$. $p_j = 0$ $\forall j \neq i$ $\Rightarrow$ $\rho = |\varphi_i\rangle\langle\varphi_i|$ pure state

For $\mathbb{C}^n$, $p_i = \frac{1}{n}$, $\rho = \sum \frac{1}{n} |\varphi_i\rangle\langle\varphi_i| = \frac{1}{n} I$, completely mixed state

like uniform distribution $(\frac{1}{n}, \frac{1}{n} \cdots \frac{1}{n})$

③ Flat state: P projection. $\rho = \frac{P}{tr(P)} = \frac{1}{k}\sum_{i=1}^{k} |\varphi_i\rangle\langle\varphi_i|$ $\{|\varphi_i\rangle\}$ O.N.B of $Ran(P)$.

$k = tr(P) \in \mathbb{N}$

④ Ensemable of pure states $\{p_i, |\varphi_i\rangle\} \to \rho = \sum p_i |\varphi_i\rangle\langle\varphi_i|$

e.g. $\lambda|0\rangle\langle 0| + (1-\lambda)|1\rangle\langle 1| = \begin{pmatrix} \lambda & \\ & 1-\lambda \end{pmatrix}$ $\lambda |+\rangle\langle +| + (1-\lambda)|-\rangle\langle -| = \begin{bmatrix} \frac{1}{2} & 2\lambda-1 \\ 2\lambda-1 & \frac{1}{2} \end{bmatrix}$

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

Postulate 1 : Quantum system $\mathbb{C}^n$

State $\rho = \sum p_i |\varphi_i\rangle\langle\varphi_i|$ density operator

$\{p_i, |\varphi_i\rangle\}$ ensemable

$\varphi(\cdot) = tr(\rho \cdot)$ "state"

Postulate 2: Observables $A = A^*$ expected value

$\varphi(A) = tr(\rho A) = \sum_i p_i \langle\varphi_i| A |\varphi_i\rangle$

Measurement $\{E_m\}$: prob. of out come m

POVM $\varphi(E_m) = tr(\rho E_m) = \sum_i p_i \langle\varphi_i| E_m |\varphi_i\rangle$

$E_m \geq 0 \Rightarrow \varphi(E_m) \geq 0$. , $\sum_m E_m = I \Rightarrow \sum \varphi(E_m) = \varphi(\sum E_m)$

$= \varphi(I) = 1$

So $\{\varphi(E_m)\}$ prob. density function

Postulate3: Transformation of closed system : Unitary U

$|\varphi\rangle \to U|\varphi\rangle$

$\{p_i, |\varphi_i\rangle\} \to \{p_i, U|\varphi_i\rangle\}$ $(U|\varphi_i\rangle)^* = \langle\varphi_i| U$

$\rho = \sum p_i |\varphi_i\rangle\langle\varphi_i| \to \sum p_i U|\varphi_i\rangle\langle\varphi_i|U^*$

$= U(\sum p_i |\varphi_i\rangle\langle\varphi_i|)U^*$

$= U \rho U^*$ unitary conjugate

Postulate 4: Composite System $\longleftrightarrow$ Tensor product space $\mathbb{C}^n \otimes \mathbb{C}^m$

What type of density operators we can have on $\mathbb{C}^n \otimes \mathbb{C}^m$?

Joint States

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

① Product state

Let $\rho \in D(\mathbb{C}^n)$ and $\sigma \in D(\mathbb{C}^m)$. Then $\rho \otimes \sigma \in D(\mathbb{C}^{nxm})$

Fact: a) $tr(A \otimes B) = tr(A)\,tr(B)$. b) $\rho \geq 0$, $\sigma \geq 0 \Rightarrow \rho \otimes \sigma \geq 0$

Example:

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

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

Completely mixed state $\frac{1_n}{n} \otimes \frac{1_m}{m} = \begin{bmatrix} \frac{1}{n} & & \\ & \ddots & \\ & & \frac{1}{n} \end{bmatrix} \otimes \begin{bmatrix} \frac{1}{m} & & \\ & \ddots & \\ & & \frac{1}{m} \end{bmatrix} = \begin{bmatrix} \frac{1}{nm} & & \\ & \ddots & \\ & & \frac{1}{nm} \end{bmatrix} = \frac{1_{nm}}{nm}$

Classical analog: $P_{XY} = P_X \times P_Y$ independent distribution

2. Separable state $w = \sum_i \lambda_i \rho_i \otimes \sigma_i$ $\quad \sum \lambda_i = 1$, $\lambda_i \geq 0$

convex combination of product state $\quad \rho_i \in D(H_A)$, $\sigma_i \in D(H_B)$

Example. a) $\rho \otimes \sigma$ product state

b) $w = \sum \lambda_i |i\rangle\langle i| \otimes \sigma_i$ $\quad \sigma_i \in D(H_B)$, $\sum \lambda_i = 1$, $\lambda_i \geq 0$ $\quad$ classical-quantum state

If $[\sigma_i, \sigma_j] = 0$ $\forall i, j$ Then $\exists$ a O.N.B $|\varphi_i\rangle$ s.t. $\forall i$, $\sigma_i = \sum p_{i,k} |\varphi_k\rangle\langle\varphi_k|$

Then $w = \sum \lambda_i |i\rangle\langle i| \otimes \sigma_i = \sum_i \lambda_i |i\rangle\langle i| \otimes \sum_k p_{ik} |\varphi_k\rangle\langle\varphi_k|$

$$= \sum_{i,k} \underbrace{\lambda_i p_{ik}}_{q_{i,k}} |i\rangle\langle i| \otimes |\varphi_k\rangle\langle\varphi_k|$$

Definition: A joint state ρ_{AB} is called entangled if it is not seperable

Example: $|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \in \mathbb{C}^2 \otimes \mathbb{C}^2$

density $|\Phi^+\rangle\langle\Phi^+| = \left(\frac{|00\rangle + |11\rangle}{\sqrt{2}}\right)\left(\frac{\langle 00| + \langle 11|}{\sqrt{2}}\right)$

$= \frac{1}{2}\left(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\right)$

O.N.B $\{|00\rangle \quad |01\rangle \quad |10\rangle \quad |11\rangle\}$

$$|\Phi^+\rangle\langle\Phi^+| = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

How to see this is not seperable state

In general, $|\varphi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$. if $|\varphi\rangle \neq |h_1\rangle \otimes |h_2\rangle$ produce vector

then $|\varphi\rangle\langle\varphi|$ is an entangled state

Marginal seate / Reduced density

Given a joint state $\rho_{AB} \in B(H_{AB})$. it induce a seate on A.

$\varphi_A : B(H_A) \to \mathbb{C}$. $\varphi_A(X) = tr_{AB}(X_A \otimes I_B)\,\rho_{AB})$

$= tr_A(X_A \rho_A)$ for some ρ_A

$\rho_A = Id \otimes tr_B(\rho_{AB})$

More explicitly, $\rho_{AB} = \sum a_{ij,kl}\, e_{ij} \otimes e_{kl}$

$\rho_A = id_A \otimes tr_B(\rho_{AB}) = \sum a_{ij,kl}\, Id(e_{ij}) \otimes tr_B(e_{kl})$

$= \sum_{k,ij} a_{ij,kk}\, e_{ij}$

For example: ① $W_{AB} = \rho_A \otimes \sigma_B \Rightarrow W_A = id_A \otimes tr_B = \rho_A\, tr_B(\sigma_B) = \rho_A$

$$w_B = \sigma_B$$

② $w_{AB} = \sum \lambda_i |i\rangle\langle i| \otimes \sigma_i$ $\quad w_A = \sum \lambda_i |i\rangle\langle i|$ $\quad w_B = \sum \lambda_i \sigma_i$

③ $\varphi_{AB} = |\varphi\rangle\langle\varphi|$. $\quad |\varphi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$ $\quad \varphi_A = \varphi_B = \frac{1}{2}$

$|\varphi\rangle = a|00\rangle + b|11\rangle$ $\quad \varphi_A = |a|^2 |0\rangle\langle 0| + |b|^2 |1\rangle\langle 1|$

$\varphi_B = |a|^2 |0\rangle\langle 0| + |b|^2 |1\rangle\langle 1|$

Now: back to our equivalence:

Given $|\varphi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$, $\longrightarrow$ $\varphi : B(\mathbb{C})^n \to \mathbb{C}$ $\longrightarrow$ $\rho_\varphi \in B(\mathbb{C})^n$

vector state $\quad$ state $\quad \varphi(X) = \langle\varphi| X \otimes I |\varphi\rangle$ $\quad \varphi(X) = tr(\rho_\varphi X)$

How to compute ρ_φ from $|\varphi\rangle$

$$\langle\varphi| X \otimes I |\varphi\rangle = tr_{AB}(X \otimes I\, |\varphi\rangle\langle\varphi|)$$
$$= tr_A \otimes tr_B (X \otimes I\, |\varphi\rangle\langle\varphi|)$$
$$= tr_A (X\ tr_B(|\varphi\rangle\langle\varphi|))$$
$$= tr_A (X\ tr_B(\varphi_{AB})) = tr_B(X\, \varphi_A)$$
$$\rho = \varphi_A = tr_B(|\varphi\rangle\langle\varphi|)$$

Can we go back? $\quad \rho \in B(\mathbb{C})^n$ (density) $\longrightarrow$ find $|\rho\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$ s.t. $tr_B(|\rho\rangle\langle\rho|) = \rho$

Purification.

Given a mixed state $\rho \in D(H_A)$

① Does there exises a joint state w_{AB} such that $w_A = \rho$?

② . . - - - - a pure joint state $\varphi_{AB} = |\varphi\rangle\langle\varphi|$ such that $\varphi_A = \rho$?

Such $|\varphi\rangle$ is called a purification of ρ.

① Yes. $W_{AB} = \rho_A \otimes \sigma_B$

② Given $\rho = \sum p_i |\varphi_i\rangle\langle\varphi_i|$. Define $|\varphi\rangle_{AA'} = \sum_i \sqrt{p_i}\, |\varphi_i\rangle_A \otimes |i\rangle_{A'}$ ($A \cong A'$)

Then $\varphi = |\varphi\rangle\langle\varphi|$ has reduced density ρ on A.

$$id_A \otimes tr_{A'}(|\varphi\rangle\langle\varphi|) = id_A \otimes tr_{A'}\left(\sum_{ij} \sqrt{p_i}\sqrt{p_j}\, |\varphi_i\rangle\langle\varphi_j| \otimes |i\rangle\langle j|\right)$$

$$= \sum_i p_i |\varphi_i\rangle\langle\varphi_i| = \rho.$$

Theorem (Ulmann): Let $|\psi\rangle_{AC}$ be any purification of ρ. Then there exises an isometry $V \in L(H_{A'}, H_C)$ (i.e. $V^\dagger V = I_{H_A}$) such that $V|\varphi\rangle = |\psi\rangle$.

Quantum Channels

Transformation of closed quantum system

$$\Phi: B(\mathbb{C}^n) \to B(\mathbb{C})^n$$
$$\Phi(\rho) = U\rho U^*$$

How about open quantum systems?

Two approaches < Mathematical: CPTP map ; Physical: interaction with environment > Quantum channels

Mathematical approach

$H = \mathbb{C}^n \quad K = \mathbb{C}^m, \quad \Phi: B(H_A) \to B(H_B)$ linear

$\rho \in D(H_A) \Rightarrow \Phi(\rho) \in D(H_B)$

Def: A linear map $\Phi: B(H) \to B(K)$ is positive if

$\forall A \geq 0, \quad \Phi(A) \geq 0$

Example: ① $a \in L(K, H)$ $\quad \Phi(X) = a^* X a$

$\forall |k\rangle \in K, \quad \langle k|aXa^*|k\rangle = \langle k|a^* X a|k\rangle$

$= \langle ak|X|ak\rangle \geq 0$

In particular: Unitary U, $\quad \Phi(\rho) = U\rho U^*$

isometry. $V: H \to K$. $\quad V^*V = I_H$

$\Phi(\rho) = V\rho V^*$

② Trace map: $tr: B(H) \to \underset{\text{s.t.}\ B(\mathbb{C})}{\mathbb{C}}$, $X \to tr(X)$

If $X \geq 0$, $tr(X) = \sum_{i=1}^{n} \langle e_i | X | e_i \rangle \geq 0$

③ Transpose: Fixed a basis $\{|i\rangle\} \subseteq H$

$T: B(H) \to B(H)$, $(\rho_{ij})_{i,j} \to (\rho_{ji})_{i,j}$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \to \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$

$A \geq 0 \iff A = B^* B \Rightarrow A^t = B^t (B^*)^t = B^t (B^t)^*$, ≥ 0

④ $P(B(H), B(K)) = \{$positive map from $B(H)$ to $B(K)\}$ is a convex cone

If Φ_1, Φ_2 positive, $\forall\ \lambda \Phi_1 + u \Phi_2$ positive if $\lambda, u \geq 0$

So $\Phi(X) = \sum a_i X a_i^*$ is positive

Def: A quantum channel from system H to K is

a linear map $\mathcal{I}: B(H) \to B(K)$ satisfying

① $\mathcal{I}$ is completely positive: $\forall$ Hilbert space H' s.t. $id_{B(H')} \otimes \mathcal{I}$ is positive

② $\mathcal{I}$ is trace preserving: $tr(\rho) = tr(\mathcal{I}(\rho))$

Idea: If $\rho \geq 0$, $tr(\rho) = 1$, then $\mathcal{I}(\rho) \geq 0$, $tr(\mathcal{I}(\rho)) = 1$

$tr(\rho) = tr(\mathcal{I}(\rho)) = 1$ by TP

$\rho \geq 0 \Rightarrow \mathcal{I}(\rho) \geq 0$ by positivity

So $\mathcal{I}(D(H)) \subseteq D(K)$.

Why complete positivity instead of positivity?

It's all about entanglement

$\exists\ \mathcal{I}$ positive, $Id_{B(H')} \otimes \mathcal{I}$ is not positive for some H'.

So $\mathcal{I}$ positive $\not\Rightarrow$ $\mathcal{I}$ CP

$\exists\ W \in B(H \otimes H')$, $W \geq 0$, but $\mathcal{I} \otimes id_{H'}(W)$ not positive

Note that if $W = \sum p_i\ \rho_i \otimes \sigma_i$ separable state

$$\mathcal{I} \otimes id(W) = \sum p_i \underbrace{\mathcal{I}(\rho_i)}_{D(K)} \otimes \sigma_i \geq 0$$ seperable states

This implies if $W \geq 0$ & $\mathcal{I} \otimes id_{H'}(W)$ is not positive for some $\mathcal{I}$ positive

W is entangled!

Example: $\Phi = t$ transpose, on the $\{|i\rangle\}$ basis

$$\Phi \otimes id \left(\sum a_{ij,kl} |i\rangle\langle j| \otimes |k\rangle\langle l| \right)$$
$$= \sum a_{ij,kl} |i\rangle\langle j|^t \otimes |k\rangle\langle l|$$
$$= \sum a_{ij,kl} |j\rangle\langle i| \otimes |k\rangle\langle l|$$

Now consider $|\varphi\rangle = \frac{1}{\sqrt{n}} \sum |i\rangle \otimes |i\rangle$ maximally entangled state on $\mathbb{C}^n \otimes \mathbb{C}^n$

density
$$\varphi = |\varphi\rangle\langle\varphi| = \frac{1}{n} \sum_{i,j} |i\rangle|i\rangle \langle j|\langle j|$$
$$= \frac{1}{n} \sum \underbrace{|i\rangle\langle j|}_{e_{ij}} \otimes \underbrace{|i\rangle\langle j|}_{e_{ij}}$$

Apply partial transpose,
$$\Phi \otimes id(\varphi) = \frac{1}{n} \sum (|i\rangle\langle j|)^t \otimes |i\rangle\langle j|$$
$$= \frac{1}{n} \sum_{i,j} \underbrace{|j\rangle\langle i| \otimes |i\rangle\langle j|}_{F}$$

One can verify $F = F^*$, $F^2 = I$ self-adjoint unitary

$F \neq I \Rightarrow$ F has spectrum $\{-1, 1\}$

Not positive!

So transpose t map is positive, but not completely positive

it does not perseve positivity for all entangled state.

What are CP?

① $\Phi(\rho) = a\rho a^*$ b/c $\Phi\otimes id(\rho) = (a\otimes 1)\rho(a^*\otimes 1)$

$\Phi(\rho) = \sum a_i \rho a_i^*$ cp

② Trace: $tr: B(H) \to \mathbb{C}$, $tr(\rho) = 1 \quad \forall \rho \in D(H)$

Physically, this is just ignore the system

called forgetful channel

Partial trace $tr\otimes id_K : B(H\otimes K) \to B(H)$

forget part of the system.

Pf: $\forall A \geq 0$, $A \in B(K)$

$tr_K(tr_H \otimes id_K(\rho) A)$

$= tr_K \otimes tr_H (\rho A \otimes I)$

Quantum Channel II

Stinspring Dilation

Theorem. The following are equivalent

① A linear map $\Phi: B(H) \to B(H)$ is a quantum channel (cptp map)

② There exists a partial isometry $V: H \to K \otimes H_E$ s.t.

$$\Phi(\rho) = tr_E(V \rho V^*) \qquad \text{(Stinespring Dilation)}$$

③ There exists a family of operators $\{a_i\} \subseteq B(H,K)$ s.t.

$$\sum a_i^* a_i = I \quad \text{and} \quad \Phi(\rho) = \sum_{i=1}^{k} a_i \rho a_i^* \qquad \text{(Kraus operators)}$$

Pf (③ ⇒ ② ⇒ ①)

Given $\Phi(\rho) = \sum_{i=1}^{k} a_i \rho a_i^*$

Define $H_E = \mathbb{C}^k$ $\qquad H \otimes H_E = H \otimes \mathbb{C}^k \cong H \oplus H \oplus \cdots \oplus H$

$V: H \to H \otimes H_E$

$$V(|\varphi\rangle) = \sum_i a_i |\varphi\rangle \otimes |i\rangle$$

V is a isometry b/c $\langle\varphi|V^*V|\varphi\rangle = \sum_{i=1}^{k} \langle\varphi| a_i^* a_i |\varphi\rangle = \langle\varphi| \sum_{i=1}^{k} a_i^* a_i |\varphi\rangle$

$= \langle\varphi|\varphi\rangle$

Then for any $|\varphi\rangle\langle\varphi|$, pure state

$$tr_E(V|\varphi\rangle\langle\varphi|V^*) = tr_E\left(\left(\sum_i a_i|\varphi\rangle \otimes |i\rangle\right)\left(\sum_j \langle\varphi| a_j^* \otimes \langle j|\right)\right)$$

$$= tr_E\left(\sum_i a_i|\varphi\rangle\langle\varphi| a_j^* \otimes |i\rangle\langle j|\right)$$

$$= \sum_i a_i|\varphi\rangle\langle\varphi| a_i^* = \Phi(|\varphi\rangle\langle\varphi|)$$

② ⇒ ① is obvious b/c

$V \cdot V^*$: $\rho \to V\rho V^*$ are CPTP

tr_E : $\mathcal{L}(\mathcal{H}_{AE}) \to \mathcal{L}(\mathcal{H}_A)$

So does their composition $\mathcal{I}(\rho) = tr_E(V\rho V^*)$

Now for the hard part ① ⇒ ③

CPTP ⇒ Kraus operators.

We introduce several interesting tools.

Let $\mathcal{H}_A \cong \mathcal{H}_{A'} \cong \mathbb{C}^d$. Fix a O.N.B $\{|i\rangle\} \subset \mathcal{H}_A$

Maximally entangled state $\Phi_{AA'} = |\Phi\rangle\langle\Phi|$

$$|\varphi\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle|i\rangle$$

Lemma: $\forall a \in \mathcal{B}(\mathcal{H})$, $a \otimes 1 |\varphi\rangle = 1 \otimes a^t |\varphi\rangle$

where a^t is the transpose of a w.r.t to $\{|i\rangle\}$

Pf: Sufficient to consider $a = |k\rangle\langle l|$

$$|k\rangle\langle l| \otimes 1 |\varphi\rangle = (|k\rangle\langle l| \otimes 1) \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle \otimes |i\rangle$$

$$= \frac{1}{\sqrt{d}} |k\rangle \otimes |l\rangle$$

$$= (1 \otimes |l\rangle\langle k|)|\varphi\rangle = (1 \otimes |k\rangle\langle l|^t)|\varphi\rangle$$

The same extends to general $a = \sum a_{kl} |k\rangle\langle l|$

① ⇒ ③ $\mathcal{I} : B(H_A) \to B(H_B)$ CP

$\Rightarrow id_{A'} \otimes \mathcal{I} : B(H_{A'} \otimes H_A) \to B(H_{A'} \otimes H_B)$ CP $\quad H_{A'} \cong H_A$

Take (non normalized) maximal entangled state (MES)

$|\varphi\rangle = \sum |i\rangle|i\rangle \quad \varphi_{A'A} = |\varphi\rangle\langle\varphi| \quad \varphi_A = \varphi_{A'} = 1$

Then $w = id \otimes \mathcal{I}(\varphi) \geq 0$. Note that $\forall\, \rho \in D(H_A)$

Define $|\rho\rangle = 1 \otimes \sqrt{\rho}\,|\varphi\rangle$ $\quad |\rho\rangle$ is a purification of ρ

$= \sqrt{\rho}^t \otimes 1 |\varphi\rangle \quad tr_{A'}(|\rho\rangle\langle\rho|) = tr_{A'}(1 \otimes \sqrt{\rho}|\varphi\rangle\langle\varphi| 1 \otimes \sqrt{\rho})$

$= \sqrt{\rho}\ 1\ \sqrt{\rho} = \rho$

Then $\mathcal{I}(\rho) = \mathcal{I}\left(tr_{A'}(|\rho\rangle\langle\rho|)\right)$

$= tr_{A'}\left(id_{A'} \otimes \mathcal{I}(|\sqrt{\rho}\rangle\langle\sqrt{\rho}|)\right)$

$= tr_{A'}\left(id_{A'} \otimes \mathcal{I}(\sqrt{\rho}^t_{A'} \otimes 1\, |\varphi\rangle\langle\varphi|\, \sqrt{\rho}^t_{A'} \otimes 1)\right)$

$= tr_{A'}\left(\sqrt{\rho_{A'}}^t w_{AA'} \sqrt{\rho_{A'}}^t\right) = tr_{A'}\left(\rho_{A'}^t w_{AA'}\right)$

$tr(AB) = tr(AB)^t = tr(B^t A^t) = tr(A^t B^t)$

Recall $H_A \otimes H_{A'} \cong B(H_A) \quad Vec(a) \longleftarrow a$

$e_i \otimes e_j \leftrightarrow e_{ij} \quad |h\rangle \to op(|h\rangle)$

Then $tr\left(\rho_{A'}^t |h\rangle\langle k|\right) = op(|h\rangle)\, \rho\, op(|k\rangle)^*$

Indeed, $tr_A\left(\rho\, (|i\rangle|j\rangle)(\langle k|\langle l|)\right) = |i\rangle\langle k|\, \langle j|\rho|l\rangle$

$= |i\rangle\langle j|\rho|l\rangle\langle k|$

$= e_{ij}\, \rho\, e_{kl}^*$

Assume $W_{AA'} = \sum |\psi_i\rangle\langle\psi_i|$

$$tr_{A'}(\rho_{A'}^t W_{AA'}) = tr\left(\rho_{A'}^t |\psi_i\rangle\langle\psi_i|\right) = \sum a_i \rho a_i^*$$

As $a_i = op(|\psi_i\rangle)$

Since $\mathcal{E}(\rho) = \sum a_i \rho a_i^*$ and $tr(\mathcal{E}(\rho)) = tr(\rho)$

$$tr(\rho \textstyle\sum a_i^* a_i) = tr(\rho)$$

$$\Rightarrow \sum_i a_i^* a_i = I$$

□

Note that in the above proof,

we used $\mathcal{E}$ CP $\Rightarrow W = id_{A'} \otimes \mathcal{E}(\varphi) \geq 0 \Rightarrow ③ \Rightarrow ② \Rightarrow \mathcal{E}$ CP

So $\mathcal{E}$ CP $\Leftrightarrow W = id_{A'} \otimes \mathcal{E}(\varphi) \geq 0$

Theorem (Choi) TFAE

① $\mathcal{E} = B(H_A) \to B(H_B)$ is CP

② $id_{A'} \otimes \mathcal{E}_A : B(H_{A'} \otimes H_A) \to B(H_{A'} \otimes H_B)$ is positive

③ $W_{A'B} = id_{A'} \otimes \mathcal{E}_{A \to B}(\varphi_{AA'})$ is positive

where $\varphi_{AA'} = \frac{1}{d}\sum |i\rangle|i\rangle$ is the m.e.s.

In particular $W_{A'B}$ is called the choi density of $\mathcal{E}$

If $\dim(H_A) = d$, $\hat{W}_{A'B} = id_{A'} \otimes \mathcal{E}_{A \to B}(d\varphi_{AA'}) = d W_{A'B}$ Choi matrix of $\mathcal{E}$

$\hat{W}_{A'B} = \sum |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|)$ determines $\mathcal{E}$.

Pf: ① ⇒ ② ⇒ ③ ⇒ ① by argument in previous proof

(① ⇒ ② obvious)

What is really a quantum channel?

Mathematically. ① CP map

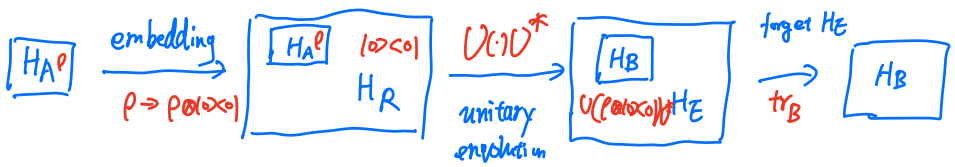
② Embedding and Partial trace

$\rho \to V\rho V^*$ $\quad G_{AE} \to G_A$

V isometry

Physically: ① A linear map sending all states (including entangled state with enviroment to states

② Unitary evolution with environment



Def: The quantum entropy of a state $\rho \in D(H)$ is defined as

$$H(\rho) = -\mathrm{tr}(\rho \ln \rho)$$

- $\rho \ln \rho = f(\rho)$ where $f(t) = t \ln t$ and $f(0) = \lim_{t \to 0} t \ln t = 0$
- Given $\rho = \sum p_i |\varphi_i\rangle\langle\varphi_i|$ orthogonal decomposition

then $f(\rho) = \sum p_i \ln p_i \, |\varphi_i\rangle\langle\varphi_i|$

$$H(\rho) = -\mathrm{tr}(\rho \ln \rho) = -\sum_i p_i \ln p_i \, \mathrm{tr}(|\varphi_i\rangle\langle\varphi_i|$$

$$= -\sum_i p_i \ln p_i = H(P)$$

where $P = \{p_i\}_{i=1}^d$ prob. distribution of spectrum of ρ

Why $\{p_i\}_{i=1}^d$ is a prob distribution?

- Also called von Neumann entropy (or simply entropy) as introduced by John von Neumann in 1930s.
- The unit is called "Qubit" Quantum bit.

Example: ① Pure state $\rho = |\rho\rangle\langle\rho|$ $H(|\rho\rangle\langle\rho|) = 0$ $\forall$ $|\rho\rangle \in H$ unit

② Maximally Mixed state: $\rho = \frac{1}{d} I = \sum \frac{1}{d} |i\rangle\langle i|$ where $d = \dim(H)$

$$H(\rho) = H\left(\left\{\frac{1}{d}, \frac{1}{d} \cdots \frac{1}{d}\right\}\right) = \log d$$

In particular, $d = 2$. $\rho = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $H(\frac{1}{2} I_2) = \log_2 2 = 1$ qubit

③ Embedding classical state into quantum system.

Given a classical information source P_X on $\mathcal{X}$, fix an O.N.B $\{|x\rangle\}$ in $\mathbb{C}^{\mathcal{X}}$

$$P_X \longrightarrow \rho_X = \sum P_X(x) |x\rangle\langle x|$$

Preparation process: $x \longmapsto |x\rangle\langle x|$.

Note that if $\rho = \sum \lambda_i |\varphi_i\rangle\langle\varphi_i|$ with $|\varphi_i\rangle$ not mutually orthogonal

then $H(\rho) \neq H(\{\lambda_1 \cdots \lambda_n\})$

Indeed $\rho = \frac{1}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1| + \frac{1}{4}|+\rangle\langle +| + \frac{1}{4}|-\rangle\langle -|$

$= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$ $\quad H(\rho) = \log 2 \neq \log 4$

(Properties of H) ① $H(\rho) \geq 0$ with equality iff ρ pure

② $H(\rho) \leq \log d_H$, - —— - iff $\rho = \frac{1}{d_H}$

③ $\sum p_i H(\rho_i) \leq H(\sum p_i \rho_i)$

Namely, entropy is a concave function

④ $H(\rho) = H(V\rho V^*)$ for any $V^*V = I$ isometry

In particular $H(\rho) = H(U\rho U^*)$ for $\forall$ U unitary

⑤ $H(\rho) = \min_{\{E_y\}} H(P_Y)$

where the minimum is over all povm $\{E_y\}$

and $P_Y(y) = \mathrm{tr}(E_y \rho)$ is the measurement outcome.

Pf: ① & ② Note that $H(\rho) = H(P)$ where P is the distribution given by the eigenvalues of ρ. Then $0 \leq H(P) \leq \log d$ as $|\mathrm{spec}(\rho)| \leq d$

$H(P) = 0 \iff P$ is point mass $\{1, 0 \cdots 0\}$

$\iff \rho = |\varphi\rangle\langle\varphi|$

$H(P) = \log d \iff P = \{\frac{1}{d} \cdots \frac{1}{d}\}$

$\iff \rho = \sum \frac{1}{d} |\varphi_i\rangle\langle\varphi_i|$ o.n.b $|\varphi_i\rangle$.

$= \frac{1}{d}$

③ (Nontrivial) Postpone

④ If $\rho = \sum p_i |\varphi_i\rangle\langle\varphi_i|$ $\quad V\rho V^* = \sum p_i V|\varphi_i\rangle\langle\varphi_i|V^*$

$= \sum p_i |V\varphi_i\rangle\langle V\varphi_i|$

$\langle\varphi_j|V^*V|\varphi_i\rangle = \langle\varphi_j|\varphi_i\rangle = \delta_{i,j} \Rightarrow |V\varphi_i\rangle$ o.n.b

where we used $V^*V = I$

⑤ The minimum if attained by choosing $E_i = |\varphi_i\rangle\langle\varphi_i|$ then $p_i = tr(\rho|\varphi_i\rangle\langle\varphi_i|)$

$H(\rho) = H(P)$

For the other direction, we need $H(\rho) \le H(P)$ for any $P = \Lambda(\rho)$ (measurement)

Given $\{E_i\}$ POVM, Λ is the measurement map: $\Lambda : B(H) \to \mathbb{C}^{|\Lambda|}$

$\rho \longrightarrow p_i = tr(\rho E_i)$

Joint Entropy.

Def The joint entropy of a joint state $\rho_{AB} \in D(H_A \otimes H_B)$ is defined as

$H(AB)_\rho = H(\rho_{AB})$

Example: ① Product state $\rho_{AB} = \rho_A \otimes \rho_B$

$H(AB)_\rho = H(\rho_A \otimes \rho_B) = H(\rho_A) + H(\rho_B) = H(A)_\rho + H(B)_\rho$

where $H(A)_\rho = H(\rho_A)$, $\rho_A = tr_B(\rho_{AB})$ similarly for $H(B) = H(\rho_B)$

② Classical-quantum state

$\rho_{XB} = \sum p(x)|x\rangle\langle x| \otimes \rho_x$ $\quad H(XB)_\rho = H(X) + \sum_x p_x(x) H(\rho_x)$

Indeed, suppose $\rho_x = \sum p_{x,y} |\varphi_{x,y}\rangle\langle\varphi_{x,y}|$

$$\rho_{XB} = \begin{bmatrix} P_1\rho_1 & & \\ & P_2\rho_2 & \\ & & \ddots \\ & & & P_n\rho_n \end{bmatrix} = \begin{bmatrix} \sum P_{1,y}|\varphi_{1,y}\rangle\langle\varphi_{1,y}| & & \\ & \ddots & \\ & & \ddots \end{bmatrix}$$

$$H(\rho_{XB}) = -\sum P_{x,y} \log P_{x,y}$$

$$= H(XY) = H(X) + H(Y|X) = H(X) + \sum P_X(x)\, H(\rho_{Y|X=x})$$

$$= H(X) + \sum P_X(x)\, H(\rho_x)$$

③ Pure state $\rho_{AB} = |\varphi\rangle_{AB}\langle\varphi_{AB}|$

$$H(AB)_\rho = H(\rho_{AB}) = 0$$

Conditional Entropy

Def: Given $\rho_{AB} \in D(H_A \otimes H_B)$, the conditional entropy of ρ conditional on B is

$$H(A|B)_\rho = H(AB)_\rho - H(B)_\rho$$

Example: ① classical case. P_{XY} joint distribution

$\rightsquigarrow$ $\rho_{AB} = \sum P_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y|$ for any O.N.B $\{|x\rangle\} \subseteq H_A$, $\{|y\rangle\} \subseteq H_B$

$$H(A|B)_\rho = H(AB)_\rho - H(B)_\rho$$

$$= H(XY)_\rho - H(Y)_\rho = H(X|Y)_\rho$$

② Semi classical case: $\rho_{XB} = \sum P_x |x\rangle\langle x| \otimes \rho_x^B$

$$H(B|X)_\rho = H(BX)_\rho - H(X) = \sum P_x H(\rho_x)$$

③ Entangled pure state: $\rho_{AB} = |\varphi_{AB}\rangle\langle\varphi_{AB}|$ and $|\varphi_{AB}\rangle \neq |\varphi_A\rangle \otimes |\varphi_B\rangle$

$$H(A|B)_\rho = H(AB)_\rho - H(B)_\rho = 0 - H(B)_\rho \leq 0 \ !$$

Negative entropy ~ Negative uncertainty

(How to interpret this? An final project)

Property: ① $-H(B) \le H(A|B) \le H(A)$

② $H(B|X) \ge 0$, $H(X|B) \ge 0$ for cq state $\rho_{XB} = \sum p_x |x\rangle\langle x| \otimes \rho_x^B$

③ $H(A|BC) \le H(A|B)$ conditioning does not increase uncertainty

Mutual Information:

$H(A) - H(A|B) = H(A) - (H(AB) + H(B)) \ge 0 \Rightarrow H(A) + H(B) - H(AB) \ge 0$

Def The mutual Information of ρ_{AB} between A and B is

$I(A:B)_\rho = H(A)_\rho + H(B)_\rho - H(AB)_\rho$

$H(A|BC) \le H(A|B)$

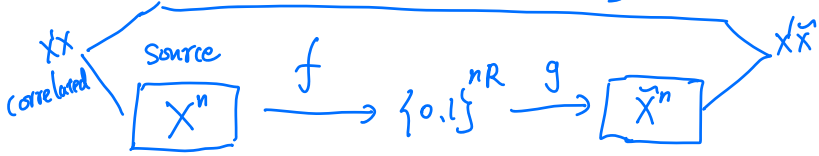
$H(ABC) - H(BC) \le H(AB) - H(B) \Rightarrow H(AB) + H(BC) - H(B) - H(ABC) \ge 0$

Def: The conditional mutual information of ρ_{ABC} between A and C conditional on B is

$I(A:C|B)_\rho = H(AB) + H(BC) - H(B) - H(ABC) \ge 0$

Highly nontrivial final project on this.

Recall classical source coding



Shannon's Theorem

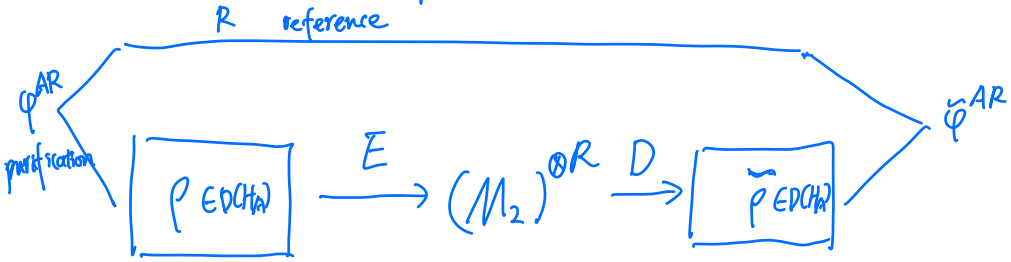
$$\lim_n \varepsilon_n(R) = \begin{cases} 1 & \text{if } R > H(X) \\ 0 & \text{if } R < H(X) \end{cases}$$

$\varepsilon_n(R) = \inf P(X^n \neq \tilde{X}^n)$, R rate of compression

$= \| P_{X'X} - P_{X'\tilde{X}} \|_{TV}$

$$P_{X'X}(x,y) = \begin{cases} P_X(x) & x = y \\ 0 & x \neq y \end{cases}$$

Quantum data compression

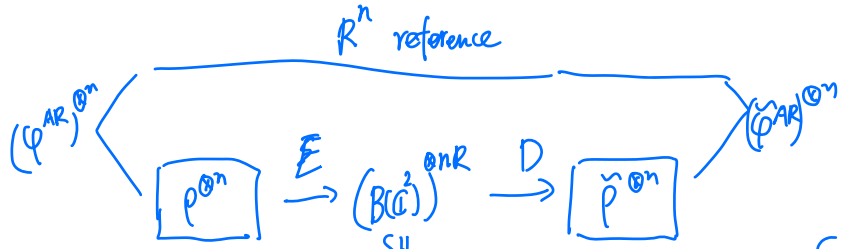


$$\varepsilon(R) = \inf_{(E,D)} \text{"}Pr(\rho \neq \tilde{\rho})\text{"} = \inf_{(E,D)} \| \varphi^{AR} - E \circ D(\varphi^{AR}) \|_1$$

Classical case: $\|P-Q\|_{TV} = \frac{1}{2}\|P-Q\|_1 = \frac{1}{2}\sum_{x \in \mathcal{X}} |P(x)-Q(x)| = \inf_{X \sim P,\, \tilde{X} \sim Q} P(X \neq \tilde{X})$

Quantum case: $\|\rho - \sigma\|_1 = \operatorname{tr}(|\rho-\sigma|) = \sup_{0 \leq P \leq 1} \operatorname{tr}((P-\sigma)P) :=$ maximal guess probability.

I.I.d Setting



$$\varepsilon_n(R) = \inf_{E,D} \| \varphi^{AR} - DE(\tilde{\varphi}^{AR\otimes n}) \|_1$$

Schumacher Compression (1994)

$$\lim_{n\to\infty} \varepsilon_n(R) = \begin{cases} 0 & \text{if } R > H(\rho) \\ 1 & \text{if } R < H(\rho) \end{cases}$$

Direct coding ($R > H(\rho)$)

$\rho = \sum p_x |x\rangle\langle x|$, $H(\rho) = H(P)$. $P_X = \{p_x\}$ over $x \in \mathbb{X}$

Then if $R > H(\rho) = H(P_X)$, by shannon's coding theorem, $\exists \{f_n, g_n\}$ - electetable error

$$P_X^n \xrightarrow{f_n} \{0,1\}^{nR} \xrightarrow{g_n} P_X^n$$

$\exists S_n \subseteq \mathbb{X}^n$ s.t. $g_n \circ f_n(S_n) = I$ $\lim_{n\to\infty} P(X^n \neq \hat{X}^n) = \lim_n P(S_n) = 0$

Define partial isometry $V: (\mathbb{C}^{|\mathbb{X}|})^{\otimes n} \to (\mathbb{C}^2)^{\otimes nR}$

$\mathbb{C}^{\mathbb{X}^n}$ $V(|x_1 \cdots x_n\rangle) = \begin{cases} |f(x_1 \cdots x_n)\rangle & \text{if } x_1 \cdots x_n \in S_n \\ \vec{0} & \text{otherwise} \end{cases}$

$V^*V = \Pi_V$ a projection

Define $E: B(\mathbb{C}^{\mathbb{X}^n}) \to B(\mathbb{C}^{2^{nR}})$

$E(\rho) = V\rho V^* + \mathrm{tr}((I - V^*V)\rho)\,|e\rangle\langle e|$ for some fixed $|e\rangle\langle e|$ for error

$D: B(\mathbb{C}^{2^{nR}}) \to B(\mathbb{C})^{\mathbb{X}^n}$

$D(\rho) = V^*\rho V + \mathrm{tr}((I - VV^*)\rho)\,|e\rangle\langle e|$ $\mathrm{tr}(\Pi_V \rho^{\otimes}) = \mathrm{tr}(V\rho V^*) = P(S_n)$

$$D \circ E(\varphi^{AR^{\otimes n}}) = V^*V (\varphi^{AR})^{\otimes n} V^*V + \mathrm{tr}((I - V^*V)\rho)\, D(|e\rangle\langle e|)$$

$$= \Pi_V (\varphi^{AR})^{\otimes n} \Pi_V + \mathrm{tr}(\Pi_V^c \rho)\, D(|e\rangle\langle e|)$$

$$\| (\varphi^{AR^{\otimes n}}) - D \circ E(\varphi^{AR^{\otimes n}}) \|_1 \leq \| \varphi^{AR} - \Pi_V (\varphi^{AR})^{\otimes n} \Pi_V \|_1 + \mathrm{tr}(\Pi_V^c \rho) \| D(|e\rangle\langle e|) \|$$

$P(S_n^c)$

$$\leq 2\varepsilon_n + \varepsilon_n = 3\varepsilon_n \to 0$$

Weak converse:

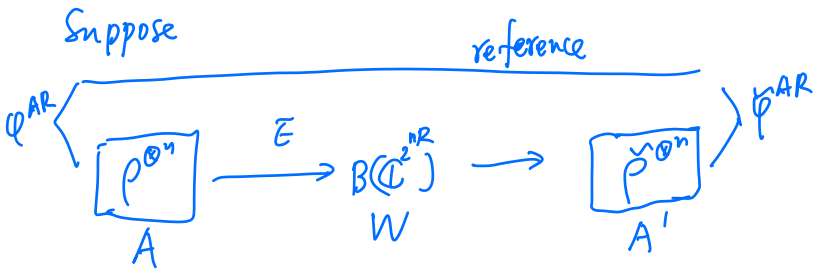
Theorem (Fannes-Audenaert) For $\rho, \sigma \in \mathcal{B}(H)$ and $\varepsilon = \frac{1}{2}\|\rho - \sigma\|_1$,

$$|H(\rho) - H(\sigma)| \leq \varepsilon \log(\dim H - 1) + h(\varepsilon).$$

Corollary: For $\rho^{AB}, \sigma^{AB} \in \mathcal{D}(H_{AB})$ and $\varepsilon = \frac{1}{2}\|\rho_{AB} - \sigma_{AB}\|_1$

$$|I(A.B)_\rho - I(A.B)_\sigma| \leq 3\varepsilon \log(\dim H - 1) + 3h(\varepsilon)$$

Pf: $I(A.B) = H(A) + H(B) - H(AB)$.



$$2nR = 2\log 2^{nR} = 2\log|W| \geq I(W.R^n) \geq I(A'^n.R^n) \overset{\varepsilon}{\geq} I(A^n.R^n)_{\rho^{\otimes n}} - 3\varepsilon\log|AR|^n - 3h(\varepsilon)$$

$$\geq n\, I(A.R) - 3\varepsilon \log|AR|^n - 3h(\varepsilon)$$

$$\geq 2nH(A) - 3\varepsilon n \log|AR| - 3h(\varepsilon)$$

$$\Rightarrow \lim_{n\to\infty} \varepsilon_n(R) \geq \frac{2}{3}\frac{H(A) - R}{\log|AR|} > 0.$$

If $R < H(A)$, $\exists\, \varepsilon$ s.t. $R + 3\varepsilon < H(A)$, then contradiction.

Classical case

Given $f: \mathcal{X} \to \mathbb{C}$, $\operatorname{supp}(f) := \{x \in \mathcal{X} : f(x) \neq 0\}$

Def (Relative entropy). Let P, Q be prob distributions on $\mathcal{X}$.

$$D(P\|Q) = \begin{cases} \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} & \text{if } \operatorname{supp}(P) \subseteq \operatorname{supp}(Q) \\ +\infty & \text{else} \end{cases}$$

- Also called Kullback–Liebler Divergence, 1954
- $D(P\|Q) = \mathop{\mathbb{E}}_{X \sim P}\left(\log \frac{P(x)}{Q(x)}\right) = \sum Q(x) \frac{P(x)}{Q(x)} \log \frac{P(x)}{Q(x)}$

  $= \mathop{\mathbb{E}}_{X \sim Q}\left(\frac{P}{Q} \log \frac{P}{Q}\right)$

  why this is well defined $\operatorname{supp}(P) \subseteq \operatorname{supp}(Q) \Rightarrow \frac{P(x)}{Q(x)}$ could be $\frac{c_1}{c_2}$ or $\frac{0}{c_2} = 0$, $0 \log 0 = 0$

  Not $\frac{c_2}{0}$ (undefined)
- Describe How well we can distinguish P from Q (Hypothesis Testing)

Example: $P, Q \in \{0,1\}$ $P(0) = \lambda$, $P(1) = 1-\lambda$, $Q(0) = \frac{1}{2}$, $Q(1) = \frac{1}{2}$

$$D(P\|Q) = \lambda \log \frac{\lambda}{0.5} + (1-\lambda) \log \frac{1-\lambda}{0.5} = \lambda \log \lambda + (1-\lambda) \log(1-\lambda) - 1$$

$$= 1 - h(\lambda) \to 0 \text{ if } \lambda \to \tfrac{1}{2}$$

Properties of $D(\cdot\|\cdot)$

① $D(P\|Q) \geq 0$, with equality iff $P=Q$

② $D(\sum\lambda_i P_i \| \sum\lambda_i Q_i) \leq \sum\lambda_i D(P_i\|Q_i)$ Joint Convex

Pf: ① $D(P\|Q) = \mathbb{E}_Q\left(\frac{P}{Q}\log\frac{P}{Q}\right) \geq \mathbb{E}_Q\left(\frac{P}{Q}\right)\log\mathbb{E}_Q\left(\frac{P}{Q}\right)$ $\left(t\mapsto t\log t \text{ convex}\right)$

$= 1\cdot\log 1 = 0$ $\quad \mathbb{E}_Q\frac{P}{Q} = \sum_{x\in\mathcal{X}} Q(x)\frac{P(x)}{Q(x)} = \sum P(x) = 1$

② Define an extension of D

$$\bar{D}(P\|Q) = \sum P(x)\log\frac{P(x)}{Q(x)} + (Q(x)-P(x))$$

$\bar{D}(P\|Q) = D(P\|Q)$ for $P, Q \in \mathcal{P}(\mathcal{X})$

We show $\bar{D}: \mathbb{R}_+^{\mathcal{X}} \times \mathbb{R}_+^{\mathcal{X}} \to \mathbb{R}_+$ joint convex

$$\bar{D}(P\|Q) = \sum_{x\in\mathcal{X}} g(P(x), Q(x)) \qquad g(a,b) = a\log\frac{a}{b} + (b-a)$$

Hessian $\nabla^2 g(a,b) = \begin{bmatrix} 1/a & -1/b \\ -1/b & a/b^2 \end{bmatrix} \geq 0$ $\Rightarrow$ g is convex over $\mathbb{R}^2$

$\Rightarrow \bar{D}$ is joint convex

From Relative entropy to Entropies

① For P over $\mathcal{X}$, $\quad H(P) = \log|\mathcal{X}| - D(P\|\frac{1}{|\mathcal{X}|})$

$$D(P\|\tfrac{1}{|\mathcal{X}|}) = \sum P(x)\log\frac{P(x)}{1/|\mathcal{X}|} = \sum P(x)(\log P(x) + \log|\mathcal{X}|)$$
$$= \log$$

② For P_{XY} over $\mathcal{X}\times\mathcal{Y}$,

$$H(X|Y) = \log|\mathcal{X}| - D(P_{XY}\|P_X\times\tfrac{1}{|\mathcal{Y}|})$$
$$I(X:Y) = D(P_{XY}\|P_X\times P_Y)$$

Indeed, $D(P_{XY}\|P_X\times P_Y) = \mathop{\mathbb{E}}_{P_{XY}}\left(\log\frac{P_{XY}}{P_X\times P_Y}\right) = \mathop{\mathbb{E}}_{P_{XY}}\log P_{XY} + \mathop{\mathbb{E}}_{P_{XY}}\log\frac{1}{P_X} + \mathop{\mathbb{E}}_{P_{XY}}\log\frac{1}{P_Y}$

$$= H(X)+H(Y) - H(XY)$$

③ For P_{XYZ} over $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$

$$I(X:Y|Z) := H(XZ)+H(YZ) - H(XYZ) - H(Z)$$
$$= D(P_{XYZ}\|\tfrac{1}{|\mathcal{X}|}\times P_{YZ}) - D(P_{XZ}\|\tfrac{1}{|\mathcal{X}|}\otimes P_Z)$$
$$= (\log|\mathcal{X}| - H(X|YZ) - (\log|\mathcal{X}| - H(X|Z))$$
$$= H(X|Z) - H(X|YZ)$$
$$= H(XZ) - H(Z) - H(XYZ) + H(YZ)$$

Can we prove $I(X:Y|Z)$ using relative entropy? Yes.

Data processing inequality

Denote $C(X) = \{ f : X \to \mathbb{C} \}$ $\quad\quad$ $\mathbb{C}^X$ $\quad$ $\mathbb{C}^Y$

A classical channel from X to Y is a linear map $N : C(X) \to C(Y)$

① Positive: $f \geq 0 \Rightarrow Nf \geq 0$

② Measure preserving: $\mathbb{E}_X f = \mathbb{E}_Y Nf$

So $N : \mathcal{P}(X) \to \mathcal{P}(Y)$ send p.d.f to p.d.f

$N : \mathbb{C}^X \to \mathbb{C}^Y$ is given by a matrix $N(y|x)$

$p \in \mathbb{C}^X \to q(y) = \sum_{x \in X} N(y|x) p(x)$

① $\Rightarrow N(y|x) \geq 0 \quad \forall y, x$

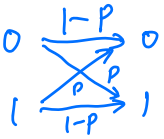
② $\Rightarrow \sum_y N(y|x) = 1$

Such $N(Y|X)$ is called a stochastic matrix

actually the coditional prob. $P_{Y|X}$ for $P_{XY} = N(Y|X) P_X$

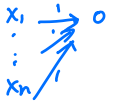
X sender $\xrightarrow{N(Y|X)}$ Y Reciever $\quad$ $N(y|x)$ prob. receiving y given input x.

Example: (Bit flip Channel) Fix $p \in [0,1]$:



Model for noisy communication.

Example: (Forgetting Channel): $\Phi : \mathbb{C}^X \to \mathbb{C}$.



Data processing Inequality

For any channel N,

$$D(Np\|Nq) \le D(p\|q)$$

Pf: $D(Np\|Nq) = \sum_y NP(y) \log \frac{NP(y)}{Nq(y)}$

$$= \sum_{x,y} N(y|x)P(x) \log \frac{NP(y)}{Nq(y)}$$

$$= \sum_x P(x) \left(\sum_y N(y|x) \log \frac{NP(y)}{Nq(y)} \right)$$

$$= \sum_x P(x) \log \exp \left(\sum_y N(y|x) \log \frac{NP(y)}{Nq(y)} \right)$$

$$D(p\|q) = \sum_x P(x) \log \frac{P(x)}{q(x)}$$

$$D(p\|q) - D(Np\|Nq) = \sum P(x) \log \frac{P(x)}{q(x) \exp(\cdots)}$$

$$= \sum_x P(x) \left[\log P(x) - \log \left(q(x) \exp \sum_y N(y|x) \log \frac{NP(y)}{Nq(y)} \right) \right]$$

$$\ge \sum_x P(x) \log P(x) - \log \left(q(x) \sum_y N(y|x) \frac{NP(y)}{Nq(y)} \right)$$

$$= \sum_x P(x) \log \frac{P(x)}{\left(q(x) \sum_y N(y|x) \frac{NP(y)}{Nq(y)} \right)}$$

Claim $r(x) = q(x) \sum_y N(y|x) \frac{Np(y)}{Nq(y)}$ is a prob distribution

Indeed $r = R \; Np$

$R = \mathbb{C}^Y \to \mathbb{C}^X$ is the channel

$Ru(x) = \sum_y \frac{N(y|x)q(x)}{Nq(y)} u(y)$ $\qquad R(x|y) = \frac{N(y|x)q(x)}{Nq(y)}$

$\sum_x R(x|y) = \sum_x \frac{N(y|x)q(x)}{Nq(y)} = \frac{Nq(y)}{Nq(y)} = 1$ Bayes Rule!

Thus $D(p\|q) - D(Np\|Nq) \geq D(p\|RNp) \geq 0$

If $D(p\|q) = D(Np\|Nq) \Rightarrow p = RNp$

Moreover $RNq = q$

$D(p\|q) = D(Np\|Nq)$ iff $\exists R: \mathbb{C}^X \to \mathbb{C}^Y$,

$R \circ N(p) = p \qquad R \circ N(q) = q.$

$\rho \in B(H)_+$ $\quad s(\rho) := $ minimal projection s.t. $s(\rho)\rho s(\rho) = \rho$

Def (Umegaki). Given $\rho, \sigma \in D(H)$, the relative entropy from ρ to σ is

$$D(\rho\|\sigma) = \begin{cases} \mathrm{tr}(\rho\log\rho - \rho\log\sigma) & \text{if } s(\rho) \le s(\sigma) \\ +\infty & \text{otherwise} \end{cases}$$

- Measuring how different ρ is w.r.t. to sigma

E.g. $\rho = \frac{1}{3}|0\rangle\langle 0| + \frac{2}{3}|1\rangle\langle 1|$ $\quad \sigma = \frac{3}{4}|+\rangle\langle +| + \frac{1}{4}|-\rangle\langle -|$

$= \begin{bmatrix} \frac{1}{3} & \\ & \frac{2}{3} \end{bmatrix}$ $\quad = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{bmatrix}$

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

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

$= \begin{bmatrix} \frac{1}{2}\log\frac{3}{16} & \frac{1}{2}\log 3 \\ \frac{1}{2}\log 3 & \frac{1}{2}\log\frac{3}{16} \end{bmatrix}$

$$D(\rho\|\sigma) = \mathrm{tr}(\rho\log\rho - \rho\log\sigma) = \mathrm{tr}\left(\begin{bmatrix} \frac{1}{3}\log\frac{1}{3} & 0 \\ 0 & \frac{2}{3}\log\frac{2}{3} \end{bmatrix} - \begin{bmatrix} \frac{1}{3} & \\ & \frac{2}{3} \end{bmatrix}\begin{bmatrix} \frac{1}{2}\log\frac{3}{16} & \frac{1}{2}\log 3 \\ \frac{1}{2}\log 3 & \frac{1}{2}\log\frac{3}{16} \end{bmatrix}\right)$$

$$= \mathrm{tr}\left(\begin{bmatrix} \frac{1}{3}\log\frac{1}{3} - \frac{1}{6}\log\frac{3}{16} & * \\ * & \frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{3}{16} \end{bmatrix}\right)$$

$$= \frac{1}{6}\log\frac{16}{27} + \frac{1}{3}\log\frac{64}{27} \approx 0.2$$

- Non symmetric $D(\rho\|\sigma) \neq D(\sigma\|\rho)$ (e.g. if $S(\rho) \neq S(\sigma)$)
- If $\rho = \sum p_x |\varphi_x\rangle\langle\varphi_x|$ $\sigma = \sum q_x |\varphi_x\rangle\langle\varphi_x|$ for the same O.N.B $|\varphi_x\rangle$

$$D(\rho\|\sigma) = D(p\|q) \quad \text{classical RE}$$

Relation to other entropies.

① $\rho \in D(H)$, $H(\rho) = \log d_H - D(\rho \| \frac{1}{d_H})$

② $\rho_{AB} \in D(H_A \otimes H_B)$, $I(A:B)_\rho = D(\rho_{AB} \| \rho_A \otimes \rho_B) \geq 0$

$$H(A|B)_\rho = \log d_A - D(\rho_{AB} \| 1_A \otimes \rho_B)$$

③ $\rho_{ABC} \in D(H_A \otimes H_B \otimes H_C)$

$$I(A:B|C) = I(A:BC) - I(A:C)$$
$$= D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) - D(\rho_{AC} \| \rho_A \otimes \rho_C) \geq 0$$

Data Process Inequality for quantum Entropy

For any quantum channel $\mathcal{E}$:

$$D(\rho\|\sigma) \geq D(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) \qquad \forall \rho, \sigma \in D(H).$$

- Lindblad '75 & Ulmann '77
- Super important in QIT ("The" inequality in my opinion)

- Later many simplied alternative proof. We consider two for final propecs
  ① Operator montone/convex function
  ② Complex interpolation

Many Corollaries

Cor. ① $D(\rho\|\sigma)\geq 0$ with equality iff $\rho=\sigma$

② Joint convexity. $D(t\rho_1+(1-t)\rho_2\|t\sigma_1+(1-t)\sigma_2)\leq tD(\rho_1\|\sigma_1)+(1-t)D(\rho_2\|\sigma_2)$

③ $I(A:B|C)\geq 0$.

Pf ① Consider $tr: B(H)\to\mathbb{C}$

$$D(\rho\|\sigma)\geq D(tr\rho\|tr\sigma)= D(1\|1)=0$$

② Consider channel $id\otimes tr_2\otimes B(H)\otimes M_2\to B(H)$

$$tD(\rho_1\|\sigma_1)+(1-t)D(\rho_2\|\sigma_2)= D\left(\begin{bmatrix} t\rho_1 & \\ & (1-t)\rho_2\end{bmatrix}\middle\|\begin{bmatrix} t\sigma_1 & \\ & (1-t)\sigma_2\end{bmatrix}\right)$$

$$\left(\text{DPI of } id\otimes tr_2\right)\geq D(t\rho_1+(1-t)\rho_2\|t\sigma_1+(1-t)\sigma_2)$$

③ $I(A:B|C)=I(A:BC)-I(A:C)$

$$=D(\rho_{ABC}\|\rho_A\otimes\rho_{BC})-D(\rho_{AC}\|\rho_A\otimes\rho_C)\geq 0$$

$id_A\otimes tr_B\otimes id_C$

Cor 2

Given quantum Channel $\mathcal{E}: B(H_B) \to B(H_{B'})$

① $I(A:B)_\rho \geq I(A:B')_{id_A \otimes \mathcal{E}(\rho)}$

② $H(A|B)_\rho \leq H(A|B')_{id_A \otimes \mathcal{E}(\rho)}$

Pf: $I(A:B)_\rho = D(\rho_{AB} \| \rho_A \otimes \rho_B) \geq D(id \otimes \mathcal{E}(\rho_{AB}) \| \rho_A \otimes \mathcal{E}(\rho_B))$

$= I(A:B')_{Id \otimes \mathcal{E}(\rho)}$

$I(A:B)_\rho = H(A)_\rho - H(A|B)_\rho$

$\vee$

$I(A:B')_{id \otimes \mathcal{E}(\rho)} = H(A)_\rho - H(A|B')_{\mathcal{E}(\rho)}$

Cor 3

If $H_B \cong H_{B'}$, and $\mathcal{E}(I) = I$ unital

③ $H(\rho) \leq H(\mathcal{E}(\rho))$

④ $H(B|A)_\rho \geq H(B'|A)_{\mathcal{E} \otimes id(\rho)}$

Pf: ③ $H(\rho) = \log d - D(\rho \| \frac{1}{d})$

$\leq \log d - D(\mathcal{E}(\rho) \| \mathcal{E}(\frac{1}{d}))$

$= \log d - D(\mathcal{E}(\rho) \| \frac{1}{d})$ b/c $\mathcal{E}(1) = 1$

④ is similar

Operational meaning of $D(\cdot\|\cdot)$

Given $\rho, \sigma \in D(H)$. We want distinguish ρ from σ by an measurement.

Ideal case: $\rho = |0\rangle\langle 0|$ $\sigma = |1\rangle\langle 1|$,

we do $E_0 = |0\rangle\langle 0|$ $E_1 = |1\rangle\langle 1|$ POVM

Then $\mathrm{tr}(\rho E_0) = 1$ $\mathrm{tr}(\rho E_0) = 0$

$\mathrm{tr}(\rho E_1) = 0$ $\mathrm{tr}(\sigma E_1) = 1$

In general, such perfect tese is no avaliable. For a general $\{T, I-T\}$ $T \geq 0$

we want $\mathrm{tr}(\rho T)$ large $\beta(T) = \mathrm{tr}(\sigma T)$ small

$\alpha(T) = \mathrm{tr}(\rho(I-T))$ small $\mathrm{tr}(\sigma(I-T))$ large

$\alpha(T)$ type I error $\beta(T)$ type II error

We can consider $p_e^* = \min_T \alpha(T) + \beta(T)$

or $\beta^*(\varepsilon) = \min_T \beta(T)$ given $\alpha(T) \leq \varepsilon$

For general (ρ, σ), $\beta^*(\varepsilon) \neq 0$ for $0 < \varepsilon < 1$.

In the i.i.d setting. $\rho^{\otimes n}$ and $\sigma^{\otimes n} \in B(H^{\otimes n})$ $\subset B(H^{\otimes n})$

$$\beta_n^*(\varepsilon) = \min_{T_n} \left\{ \mathrm{tr}(\sigma^{\otimes n} T_n) \;\middle|\; 0 \leq T_n \leq I,\ \mathrm{tr}(\rho^{\otimes n} T_n) \geq 1-\varepsilon \right\}$$

Intuitively, $\beta_n^*(\varepsilon) \to 0$, but how?

Theorem (Quantum Stein's Lemma)

$$\beta_n^*(\varepsilon) \approx e^{-nD(\rho\|\sigma)}$$

More precisely, $\limsup_{n\to\infty} \frac{1}{n} \log \beta_n^*(\varepsilon) = -D(\rho\|\sigma)$

Final project: give an proof of this.

So far we talked about

① P_X – single distribution. information source data compression

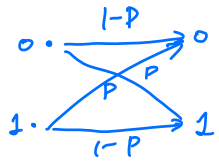
encoder Alice Bob decoder

$P_X^n \xrightarrow{f} \{0,1\}^{nR} \xrightarrow{Id} \{0,1\}^{nR} \longrightarrow P_{\tilde{X}}^n$

Rate: $\min R$ s.t. $\lim_{n\to\infty} P(X^n \neq \tilde{X}^n) \to 0$

② A channel: $N = P_{Y|X} : X \to Y$

N_p



p - error probility $0 \le p \le 1$ noisy channel

p=0 or 1 pefect channel

channel input channel output

original message $W \xrightarrow{f} X \xrightarrow{N=P_{Y|X}} Y \xrightarrow{g} \hat{W}$ decoded message

f encoder g decoder

Example: English text $\xrightarrow{f}$ morse code (telegraph station) $\xrightarrow{\text{telegraph}}$ morsecode (telegraph station) $\xrightarrow{g}$ English text

Picture $\xrightarrow{f}$ binary bits (smart phone) $\longrightarrow$ (wireless communication) binary bits (Computer) $\xrightarrow{g}$ picture

Of course, we want to minimize error. $\varepsilon = P(W \neq \hat{W})$

but as the size of message ↗, ε ↗.

So what is the **largest size of file** that can be sent through N faithfully?

Channel capacity

Mathematical Model

(Determinsetic encoder decoder): $W \xrightarrow{f} \underline{X} \xrightarrow{N=P_{Y|X}} \underline{Y} \xrightarrow{g} \hat{W}$

$m \longrightarrow f(m) \longrightarrow P_{Y|X=f(m)} \longrightarrow P_{\hat{W}|X=f(m)}$

$$P_{\hat{W}|W}(\hat{m}|m) = \sum_{g(y)=\hat{m}} P_{Y|X}(y|f(m))$$

$P_W(m) = \frac{1}{M}$ $\quad P_{\hat{W}W} = P_{\hat{W}W}(\hat{m}, m) = \frac{1}{M} \sum_{g(y)=\hat{m},} P_{Y|X}(y|f(m))$

(f,g) encoder-decoder $\quad P_e(f,g) = P(\hat{W} \neq W)$ $\quad \varepsilon^*(M) = \inf_{(f,g)} P(\hat{W} \neq W)$

(Probablistic Coding): $W \xrightarrow{E=P_{X|W}} \underline{X} \xrightarrow{N=P_{Y|X}} \underline{Y} \xrightarrow{D=P_{\hat{W}|Y}} \hat{W}$

E D classical channel

$$P_{\hat{W}W}(\hat{m}, m) = \frac{1}{M} \sum_{x,y} P_{\hat{W}|Y}(\hat{m}|y) P_{Y|X}(y|x) P_{X|W}(a|m)$$

It turns out probalistic coding is not doing better than deterministic

$$\varepsilon^*(M) = \inf_{(f,g)} P(\hat{W} \neq W) = \inf_{(E,D)} P(\hat{W} \neq W)$$

Why choose $m \sim$ uniform on W.?

By source coding. $P_X^n \sim$ uniform $\{0,1\}^{nH(X)}$, so asymptotically it suffices to consider uniform distribution, whose amount information is $\log|W| = \log M$.

$$P_e \simeq \| Id_W - E \circ N \circ D \|$$

Definition: (1) An M-code for $N = P_{Y|X}$ is an encoder/decoder pair (f,g) such that

(1) $f: \underbrace{[M]}_{\{1, \cdots, M\}} \to \underline{X}$

(2) $g: \underline{Y} \to [M]$ (or $[M] \cup \{e\}$) $\quad$ decodable error

② We say (f,g) is an (M,ε) code if (f,g) is an M-code with (for $N = P_{Y|X}$)

$$P_e = P(W \neq \hat{W}) \leq \varepsilon.$$

We interested in $M^*(P_{Y|X}, \varepsilon) = \max\{M : \exists\ (M,\varepsilon)\text{-code}\}$

$\log_2 M^*(\varepsilon)$ largese # of bits can be send through N with error $\leq \varepsilon$.

I.I.d setting. $\frac{\log_2 M^*(P_{Y|X}^n, \varepsilon)}{n}$ largese # of bits per use of channel ----- error $\leq \varepsilon$

Definition (Channel Capacity)

The Shannon capacity $C(N)$ of $N = P_{Y|X}$ is

$$C_\varepsilon(N) := \lim_{n\to\infty} \inf \frac{1}{n} \log M^*(n,\varepsilon) \qquad C(N) = \lim_{\varepsilon \to 0^+}$$

Theorem (Shannon's Noisy Channel Coding, 1948)

$$C = \sup_{P_X} I(X;Y) \qquad (P_{XY} = P_{Y|X} P_X)$$

Alternative angle: $\varepsilon^*(M,N) = \inf_{(f,g)\text{-}M \text{ code of } N} P(W \neq \hat{W})$

Optimal error, when sending $\log_2 M$ bits

I.I.d setting. $\varepsilon^*(2^{nR}, N^n)$ optimal error sending nR bits over n use of N.

Theorem: $\lim_{n\to\infty} \varepsilon^*(2^{nR}, N^n) = \begin{cases} 1 & \text{if } R > I(X;Y) \quad \text{(Strong converse)} \\ 0 & \text{if } R < I(X;Y) \quad \text{(Direct coding} \end{cases}$

Recall $I(X:Y) = D(P_{XY} \| P_X \times P_Y)$ $\quad I(X:Y|Z) = I(XZ:Y) - I(Y:Z)$

$D(P_{XYZ} \| P_{XZ} \times P_Y) - D(P_{XYZ} \| P_X \times P_{YZ})$

Lemma: ① If $Y \xrightarrow{N} Z$, $\quad I(X:Y) \geq I(X:Z)$

② $I(X:Z|Y) \geq 0$, equality iff $X \to Y \to Z$

Pf: $P_{XYZ} = P_{Z|Y} P_{Y|X} P_X$ $\quad P_{XYZ}(x,y,z) = \sum_{y} P_{Z|Y}(z|y) P_{Y|X}(y|x) P_X(x)$

$P_{XZ|Y} = \frac{\sum_x P_{Z|Y}(z|y) P_{Y|X}(y|x) P_X(x)}{P_Y(y)} = P_{Z|Y} P_{X|Y}$

③ If $X \to Y \to Z$, $\quad I(X:Z) \leq I(Y:Z)$

$I(XY:Z) = I(Y:Z) + \boxed{I(X:Z|Y)}^{=0}$

$= I(X:Z) + I(X:Y|Z)$

③ If $X \to Y \to Z \to W$

$I(Y:Z) \geq I(X:Z) \geq I(X:W)$

④ Chain Rule: $I(X_1 \cdots X_n : Y) = \sum_{k=1}^{n} I(X_k : Y | X_1 \cdots X_{k-1})$

Pf: $I(X_1 \cdots X_n : Z) = I(X_n : Y | X_1 \cdots X_{n-1}) + I(X_1 \cdots X_{n-1} : Y)$

$= \cdots = \sum_{k=1}^{n} I(X_k : Y | X_1 \cdots X_{k-1})$

Lemma: If $X_1 \to Y_1$ & $X_2 \to Y_2$,

$I(X_1X_2 : Y_1Y_2) \leq I(X_1:Y_1) + I(X_2:Y_2)$

Pf: $I(X_1X_2 : Y_1Y_2) - I(X_1:Y_1) - I(X_2:Y_2)$

$= H(X_1X_2) + H(Y_1Y_2) - H(X_1X_2Y_1Y_2) - H(X_1) - H(Y_1) + H(X_1Y_1)$

$- H(X_2) - H(Y_2) + H(X_2Y_2)$ $\quad X_1 \to X_1Y_1 \to X_1$

$= \boxed{I(X_2Y_2 : X_1Y_1) - I(X_1:X_2)} - I(Y_1:Y_2) = -I(Y_1:Y_2) \leq 0$

Theorem (Weak Converse by DPI)

Any M-code satisfies

$$\log M \le \frac{\sup_X I(X;Y) + h(P_e)}{1-P_e}$$

Pf: $W \xrightarrow{E} X \xrightarrow{N} Y \xrightarrow{D} \hat{W}$

$$\sup_{P_X} I(X;Y) \ge I(X;Y) \ge I(W;\hat{W}) \ge d(P(W\neq\hat{W}) \| \tfrac{1}{M})$$

$$h: W\hat{W} \to \{0,1\} \qquad \ge -h(P_e) + (1-P_e)\log M$$

$$h(m,\hat{m}) = \begin{cases} 1 & m=\hat{m} \\ 0 & \text{otherwise} \end{cases}$$

Lemma: $\sup_{P_{X^n}} I(X^n;Y^n) = n \sup_{P_X} I(X;Y)$

Now Apply it for N^n

If $R > I(X;Y)$, $\exists\, \varepsilon > 0$ s.t. $(1-\varepsilon)R > \sup_{P_X} I(X;Y)$

$$M = 2^{nR} \qquad \log 2^{nR} \le \frac{\sup I(X^n;Y^n) + h(\varepsilon)}{1-\varepsilon}$$

$$nR \le \frac{n \sup_{P_X} I(X;Y) + h(\varepsilon)}{1-\varepsilon}$$

$n \to \infty$ $\qquad R \le \dfrac{\sup_{P_X} I(X;Y)}{1-\varepsilon}$ (contradiction.)

So If $R > I(X;Y)$, $\qquad \lim_{n\to\infty} \varepsilon^*(2^{nR}, N^n) \neq 0$

Shannon's Achievability bound.

Theorem: Given $N = P_{Y|X}$, $\forall P_X$, $\forall \tau > 0$ $\exists (M, \varepsilon)$-code s.t.

$$\varepsilon \le P\left[\log \frac{P_{XY}}{P_X \times P_Y} \le \log M + \tau\right] + e^{-\tau}$$

where $P_{XY} = P_{Y|X} P_X$.

Denote $i(X;Y) = \log \frac{P_{XY}}{P_X \times P_Y} = \log \frac{P_{X|Y}}{P_X}$

$$\boxed{M} \xrightarrow{f} \boxed{X} \xrightarrow{N} \boxed{Y} \xrightarrow{g} \boxed{M}$$

Pf: Define $c_m = f(m)$, codeword for $m \in [M] = \{1 \cdots M\}$

We need to find good c_m and decoder g.

Define $g(y) = \begin{cases} m, & \exists!\ c_m \text{ s.t. } i(c_m; y) \ge \log M + \tau \\ e, & \text{o.w.} \end{cases}$

Interpretation: $i(c_m; y) \ge \log M + \tau \iff P_{X|Y}(c_m|y) \ge M e^{\tau} P_X(c_m)$

there is a unique m s.t. the probability m being sent given y received is above certain threshold.

Given a code book $\{c_1 \cdots c_M\}$,

$$P(W \ne \hat{W} | W = m) = P\left(\{i(c_m, Y) \ge \log M + \tau\} \cap \{\exists \bar{m} \ne m, \text{ s.t. } i(c_{\bar{m}}, Y) > \log M + \tau\}\right)$$

$$P_e(c_1 \cdots c_M) = P(W \ne \hat{W}) = \frac{1}{M} \sum_{m=1}^{M} P\left(\{i(c_m, Y) < \log M + \tau\} \cup \{\exists \bar{m} \ne m, \text{ s.t. } i(c_{\bar{m}}, Y) > \log M + \tau\}\right)$$

W is uniform on $[M]$

Random coding: we choose $C_m \sim P_X$, i.i.d. By symmetry

$$\mathbb{E}_{C_m \sim P_X}[P_e(C_1 \cdots C_M)]$$
$$= \mathbb{E}[P_e(C_1 \cdots C_M) | W=1]$$
$$= \mathbb{P}[\{i(C_1;Y) \le \log M + \tau\} \cup \{\exists\, \bar{m} \neq 1, i(C_{\bar{m}};Y) > \log M + \tau\} | W=1]$$
$$\le \mathbb{P}[\{i(C_1;Y) \le \log M + \tau | W=1\}] + \sum_{m=2}^{M} \mathbb{P}(i(C_m;Y) > \log M + \tau | W=1\}$$
$$\le \mathbb{P}[i(X;Y) \le \log M + \tau] + (M-1)\,\mathbb{P}[i(\bar{X};Y) > \log M + \tau]$$
$$\le \mathbb{P}[i(X;Y) \le \log M + \tau] + (M-1)\exp(-\log M - \tau)$$
$$\le \mathbb{P}[\text{---}] + e^{-\tau}$$

where we used $\forall x$,

$$\mathbb{P}[i(x,Y) > \tau] = \mathbb{P}\left[\log \frac{P_{Y|X=x}}{P_Y} > \tau\right] \le e^{-\tau}$$

Indeed, $Q\left[\log \frac{P}{Q} \ge t\right] = \sum_{\log \frac{P(x)}{Q(x)} \ge t} Q(x) \le \sum e^{-t} P(x) \le e^{-t}$.

Since for $C_1 \cdots C_M \sim P_X$ i.i.d $\exists$ (M, ε) code as desired, there exists some deterministic code make this happen.

Proof of Shannon's Theorem, achievability. For $M_n = 2^{nR}$, $R < I$, $\exists \delta$ s.t. $R + 2\delta < I$, Choose $\tau = \delta n$

$$\varepsilon_n^*(M) \le \mathbb{P}[i(X^n;Y^n) \le \log M_n + \tau_n] + e^{-n\delta}$$
$$= \mathbb{P}[i(X^n;Y^n) \le nR + n\delta] + e^{-n\delta}$$
$$= \mathbb{P}\left[\log \frac{P_{Y^n|X^n}}{P_{Y^n}} \le nR + n\delta\right]$$
$$= \mathbb{P}\left[\sum_{j=1}^{n} \log \frac{P_{Y_j|X_j}}{P_{Y_j}} \le nR + n\delta\right] + e^{-n\delta}$$

$$= P\left[\sum_{k=1}^{n} i(X_k;Y_k) \le n I(X;Y) - \delta n\right] + \exp(-\delta n) \to 0$$

by W.L.L.N.

Note that the above argument holds for $\forall P_X$. $\forall \varepsilon > 0$

so $C_\varepsilon = \lim_{n\to\infty} \frac{1}{n} \log M^*(n,\varepsilon) \ge \sup_{P_X} I(X;Y) - 2\delta \quad \forall \delta > 0$

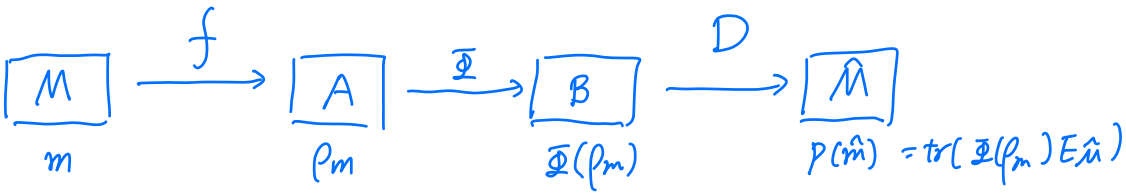
Quantum channel $\mathcal{I}: B(H_A) \to B(H_B)$

How we transmit information through quantum channel?

classical bits　　quantum qubits

Classical communication over quantum channel

$A = B(H_A)$　　$B = B(H_A)$　　$M = \{1, \cdots, M\} = \hat{M}$



Encoder: $f: M \to D(H_A)$.　$m \mapsto \rho_m$　state preparation

Decoder: POVM measurement $\{E_{\hat{m}}\}_{\hat{m}=1}^{M}$

$D: D(H_B) \longrightarrow C(M)$

$\rho \longrightarrow p(\hat{m}) = tr(\rho E_m)$

Error probability: Given input m, $P(\hat{M} = m | M = m) = tr(\mathcal{I}(\rho_m) E_m)$

$P(\hat{M} \neq M | M = m) = 1 - tr(\mathcal{I}(\rho_m) E_m)$

Averaged error: $P_e(f, D) = P(M \neq \hat{M}) = 1 - \frac{1}{M} \sum_{m=1}^{M} tr(\mathcal{I}(\rho_m) E_m)$

Max error: $P_{e,\max}(f, D) = \max_m P(M \neq \hat{M} | M = m) = \max_m 1 - tr(\mathcal{I}(\rho_m) E_m)$

(f, D) is an M code if $f: M \to A$, $D: B \to M$; (M, ε) code if $P_e(f, D) \leq \varepsilon$

Optimal error : $\varepsilon^*(\mathcal{E}, M) = \inf_{\substack{(f,D) \\ M\text{-code}}} P_e(f,D)$

Optimal code size : $M^*(\mathcal{E}, \varepsilon) = \sup\{M \mid \exists (f,D).\ (M,\varepsilon)\text{-code},\ P_e(f,D) \le \varepsilon\}$

In the i.i.d setting:

$$M \xrightarrow{f_n} A^n \xrightarrow{\mathcal{E}^{\otimes n}} B^n \xrightarrow{D_n} \hat{M}$$

Def: (Classical Capacity of Quantum Channel.)

$$C(\mathcal{E}) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\log M(\mathcal{E}^n, \varepsilon)}{n}$$

$$= \{R \mid \forall \varepsilon > 0 \ \exists\ (2^{nR}, \varepsilon)\text{-code for } \mathcal{E}^{\otimes n} \text{ for some } n\}$$

Holevo information

Def: Let $\mathcal{Y} = \{p_x, \rho_x\}$ be ensemble of quan states where $(p_x) \in \mathcal{P}(X)$ $\{\rho_x\} \subseteq D(\mathcal{H}_B)$.

$$\chi(\{p_x, \rho_x\}) = H\left(\sum_x p_x \rho_x\right) - \sum_x p_x H(\rho_x)$$

- $\chi(\{p_x, \rho_x\}) \ge 0$ by concavity of H
- Define c-q state : $\rho_{XB} = \sum p_x |x\rangle\langle x| \otimes \rho_x^B \longleftrightarrow \mathcal{Y} = \{p_x, \rho_x\}$

  $\chi(\{p_x, \rho_x\}) = I(X:B)_{\rho_{XB}}$ $\qquad \rho_X = \sum p_x |x\rangle\langle x|$ $\rho_B = \sum p_x \rho_x^B$

Indeed, $I(X{:}B)_\rho = H(X) + H(B) - H(XB)$

$= H(B) - H(X|B)$

$= H(\sum_x p_x \rho_x) - \sum_x p_x H(\rho_x)$

Def (Holevo Information of a Channel)

$$\chi(\Phi) = \sup_{\substack{\{p_x, \rho_x\} \\ \rho_x \in D(H_A)}} \chi(\{p_x, \Phi(\rho_x)\})$$

$$= \sup_{\rho_{XB} = \mathrm{Id}_X \otimes \Phi(\rho_{XX'})} I(X:B)_\rho$$

$$\rho_{XX'} = \sum_x p_x(x) |x\rangle\langle x| \otimes |x\rangle\langle x|$$

Theorem (Holevo - Schumacher - Westmore, 1997)

$$C(\Phi) = \lim_{n\to\infty} \frac{\chi(\Phi^{\otimes n})}{n}$$

① Acheivability: $\chi(\Phi) \le C(\Phi)$

② By definition, $C(\Phi^{\otimes n}) = n\, C(\Phi)^n$

(i) $C(\Phi^{\otimes n}) \ge n\, C(\Phi)$, R is an achievable rate for Φ

$\forall \varepsilon \ \exists (f_\varepsilon, D_\varepsilon)$ $(2^{kR}, \varepsilon)$ code for $\Phi^{\otimes k}$ some k)

$\Rightarrow$ $(f_\varepsilon^{\otimes n}, D_\varepsilon^{\otimes n})$ $(2^{nkR}, n\varepsilon)$ code for $\Phi^{\otimes nk}$ $\cong (\Phi^{\otimes n})^{\otimes k}$

($n\varepsilon$: union bound)

$\Rightarrow nR$ is achievable for $C(\Phi^{\otimes n})$

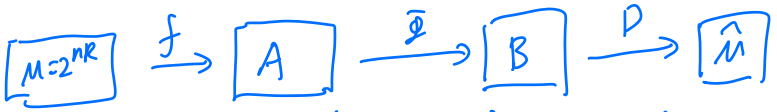
(ii) $C(\Phi^n) \le n\, C(\Phi)$, if nR ------ for $\Phi^{\otimes n}$

$\forall \varepsilon \ \exists (f_\varepsilon, D_\varepsilon)$. $(2^{nRk}, \varepsilon)$ code for $\Phi^{\otimes nk}$ - ~

$\Rightarrow R$ achievable for $\mathcal{E}$.

① + ② $\Rightarrow$ $\frac{\chi(\mathcal{E}^{\otimes n})}{n} \le C(\mathcal{E})$

③ (Weak Converse): Suppose (f, D) is (M, ε) code for $\mathcal{E}$



$$\chi(\mathcal{E}) = \sup_{P_X} I(X;B) \ge I(M;B) \ge I(M;\hat{M}) \ge D(B_\varepsilon \| B_{\frac{1}{M}})$$

$$= \varepsilon \log \frac{\varepsilon}{\frac{M-1}{M}} + (1-\varepsilon) \log \frac{(1-\varepsilon)}{\frac{1}{M}}$$

$$= \log M - \varepsilon \log(M-1) - h(\varepsilon)$$

$$\log M \le \frac{\chi(\mathcal{E}) + h(\varepsilon)}{1-\varepsilon}$$

I.i.d Suppose $R > \lim_{n\to\infty} \frac{\chi(\mathcal{E}^{\otimes n})}{n}$. $\exists\, \varepsilon > 0$ s.t. $(1-\varepsilon)R > \lim_{n\to\infty} \frac{\chi(\mathcal{E}^{\otimes n})}{n}$

For any $(2^{nR}, \varepsilon)$ code for $\mathcal{E}^{\otimes n}$

$$nR = \log 2^{nR} \le \frac{\chi(\mathcal{E}^{\otimes n}) + h(\varepsilon)}{1-\varepsilon}$$

$$(1-\varepsilon)R \le \frac{\chi(\mathcal{E}^{\otimes n})}{n} + \frac{h(\varepsilon)}{n} \to \lim_{n\to\infty} \frac{\chi(\mathcal{E}^{\otimes n})}{n}$$ Contradiction

In terms of optimal error ε^*

Theorem: $\lim_{n\to\infty} \varepsilon^*(2^{nR}, \mathcal{I}^n) = \begin{cases} 0 & \text{if } R < C(\mathcal{I}) = \lim_{n\to\infty} \frac{\chi(\mathcal{I}^{\otimes n})}{n} \\ 1 & \text{if } R > C(\mathcal{I}) \end{cases}$

$\lim_{n\to\infty} \frac{\chi(\mathcal{I}^{\otimes n})}{n}$ called regularization of $\chi(\mathcal{I})$.

In general, $\chi(\mathcal{I}^n) \geq n\,\chi(\mathcal{I})$

(Hastings '07) $\exists\, \mathcal{I}$ s.t. $\chi(\mathcal{I}\otimes\mathcal{I}) > 2\chi(\mathcal{I})$

probabilistically proof. No explicit construction.

Why does this result mean?

$M \longrightarrow A \xrightarrow{\mathcal{I}} B \longrightarrow \hat{M}$

$m \qquad \rho_m \in D(\mathcal{H}_A)$

$M \longrightarrow A^n \xrightarrow{\mathcal{I}^{\otimes n}} B^n \longrightarrow \hat{M}$

$\rho_m \in B(\mathcal{H}_A^{\otimes n}) \cong B(\mathcal{H}_A)\otimes B(\mathcal{H}_A) \cdots B(\mathcal{H}_A)$

$M^*(\mathcal{I}^n, \varepsilon) = \max\{M \mid \exists\ (M,\varepsilon) \text{ code for } \mathcal{I}^n\}$

$M^*_{pr}(\mathcal{I}^n, \varepsilon) = \max\{M \mid \exists\ (M,\varepsilon) \text{ code for } \mathcal{I}^n \ \&\ f(m) = \rho_{1,m}\otimes\rho_{2,m}\otimes\cdots\rho_{n,m}\}$ (product state)

(or $f(m) = \sum_\lambda p_\lambda \rho_{1,\lambda}\otimes\rho_{2,\lambda}\otimes\cdots\rho_{n,\lambda}$ separate state)

$\chi(\mathcal{I}) = \lim_{\varepsilon\to 0}\lim_{n\to\infty} \frac{\log M^*_{pr}(\mathcal{I}^n, \varepsilon)}{n}$ capacity if we only use product code state

$C(\mathcal{I})$ capacity use all possible code state (including entangled state)

Will entanglement help? Yes, by (Hastings '05)

$$C(\Phi) \geq \frac{\chi(\Phi \otimes \Phi)}{2} > \chi(\Phi)$$

$C(\Phi)$: Capacity using all possible code words.

$\frac{\chi(\Phi \otimes \Phi)}{2}$: capacity of allowing using entanglement between two input system

$\rho_1 \otimes \rho_2 \otimes \cdots \rho_{\frac{n}{2}}$

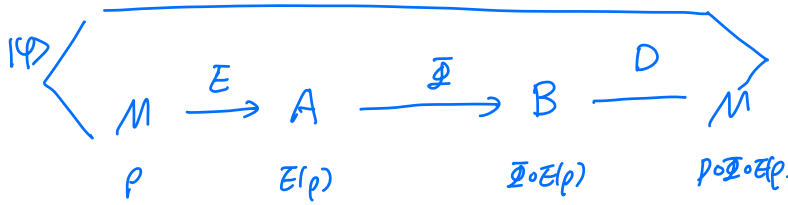
$\rho_1 \in B(H_A \otimes H_A)$, $\rho_2 \in B(H_A \otimes H_A)$

$\chi(\Phi)$: Capacity using only product code word

How to send qubits over a quantum channel ?

$M = B(H_M)$ $\quad A = B(H_A)$ $\quad B = B(H_B)$

$\dim H_M = M$



Encoder: quantum channel $E : B(H_M) \to B(H_A)$

Decoder: " " $D = B(H_B) \to B(H_M)$

Error: quantify how $D \circ \Phi \circ E \approx Id_M$

Trace Distance: for $\rho, \sigma \in D(H)$, $\|\rho - \sigma\|_1 = tr(|\rho - \sigma|)$

Channel distance: for $\Phi, \Psi : D(H) \to D(K)$ quantum channel,

$$\|\Phi - \Psi\| = \sup_{tr(|x|)=1} \frac{1}{2} \|\Phi(x) - \Psi(x)\|_1$$

$$\sup_{\rho \in D(H)} \|\Phi(\rho) - \Psi(\rho)\|_1 \le \|\Phi - \Psi\| \le 2 \sup_{\rho \in D(H)} \|\Phi(\rho) - \Psi(\rho)\|_1$$

max error over all possible input

$$\varepsilon(E, D) = \|D \circ \Phi \circ E - Id_M\|$$

$$\varepsilon^*(\Phi, M) = \inf_{(E,D)} \|D \circ \Phi \circ E - Id_M\|$$

$$M^*(\Phi, \varepsilon) = \sup \{ M \mid \exists (M, \varepsilon)\text{-code for } \Phi \}$$

Def (Quantum Capacity of Φ)

$$Q(\Phi) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\log M^*(\Phi^{\otimes n}, \varepsilon)}{n}$$

Coherent Information

Given $\rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$.

$$I(A>B)_\rho = H(B)_\rho - H(AB)_\rho = -H(A|B)_\rho$$

Lemma. $I(A>B)_\rho$ is convex over ρ

Pf: $H(A|B)_\rho = \log d_A - D(\rho_{AB} \| \frac{1}{|A|} \otimes \rho_B)$

D is jointly convex $\Rightarrow$ $H(A|B)_\rho$ is concave

Corollary: $I(A>B)_\rho \leq 0$ for all seperable state.

For a quantum channel

$$I_c(\mathcal{E}) = \sup_{\substack{\rho_{AA} \\ \in \\ \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_A)}} I(A>B)_{Id_A \otimes \mathcal{E}(\rho)}$$

Theorem: (Lloyd - Shor - Devetak)

$$Q(\mathcal{E}) = \lim_{n\to\infty} \frac{I_c(\mathcal{E}^{\otimes n})}{n}$$

- $\lim_{n\to\infty} \frac{I_c(\mathcal{E}^{\otimes n})}{n}$ is the regularization of $I_c(\mathcal{E})$

Define $M_{sep}(\mathcal{E}^n, \varepsilon) = \sup\{M \mid \exists\ (M, \varepsilon)\text{-code for } \mathcal{E}^n\}$

s.t. Range(E) are all seperable states

In general $I_c(\mathcal{E}) = \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{\log M_{sep}(\mathcal{E}^n, \varepsilon)}{n} \leq Q(\mathcal{E}) = \sup_n \frac{I_c(\mathcal{E}^{\otimes n})}{n}$

capacity for seperable coding — capacity for general coding

Example: $\mathcal{E}: M_2 \to M_2$ $\quad \mathcal{E}(\rho) = \lambda\rho + (1-\lambda)\rho$

For $\lambda \leq 0.23$, $\quad \frac{I_c(\mathcal{E}^{\otimes 5})}{5} > I_c(\mathcal{E})$

- Weak converse again by data processing inequality

- A channel $\mathcal{E}$ is called entanglement breaking if

  $\forall\ \rho_{AR} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_R)$

  $\mathcal{E} \otimes id_R(\rho_{AR})$ is seperable

  $\mathcal{E}$ is entanglement-break iff $C_{\mathcal{E}}$ is seperable

  $\mathcal{E}_\lambda(\rho) = \lambda\rho + (1-\lambda)\frac{1}{2}$ is entanglement breaking iff $0 \leq \lambda \leq \frac{1}{3}$

  For entanglement breaking $\mathcal{E}$

  $I_c(\mathcal{E}) = Q(\mathcal{E}) = 0$ No qubits can be sent through

- $Q(N)$ is really hard to computed ingeneral

  No control for $\lim_{n\to\infty} \frac{I_c(\mathcal{E})^{\otimes n}}{n}$

$\forall n \quad \exists\ \mathcal{E}$ s.t. $I_c(\mathcal{E}^{\otimes n}) = 0$ but $I_c(\mathcal{E}^{\otimes(n+1)}) = 0$

- Snperactivation (Smith & Yard)

  $\exists\ \mathcal{E}_1, \mathcal{E}_2$ s.t. $Q(\mathcal{E}_1) = Q(\mathcal{E}_2) = 0$, $Q(\mathcal{E}_1 \otimes \mathcal{E}_2) > 0$

  (Note that $Q(\mathcal{E} \otimes \mathcal{E}) = 2Q(\mathcal{E})$)

- Good case. $\mathcal{I}(\rho) = \mathrm{tr}_E(V\rho V^\dagger)$ $\quad V: H_A \to H_B \otimes H_E$

$\mathcal{I}^c(\rho) = \mathrm{tr}_A(V\rho V^\dagger)$ complementary channel

Note that for a pure state $|\varphi\rangle_{AR}$ $\quad H(A)_\varphi = H(R)_\varphi$

$$I_c(\mathcal{I}) = \sup_{|\varphi_{AR}\rangle} H(B) - H(BR) \qquad |\psi_{BER}\rangle = V|\varphi_{AR}\rangle$$
$$= \sup_\rho H(B)_{\mathcal{I}(\rho)} - H(E)_{\mathcal{I}^c(\rho)}$$

$\mathcal{I}$ is called degradable if $\mathcal{I}^c = \Xi \circ \mathcal{I}$ for some channel Ξ

Theorem: If $\mathcal{I}_1$ & $\mathcal{I}_2$ degradable,

$$I_c(\mathcal{I}_1 \otimes \mathcal{I}_2) = I_c(\mathcal{I}_1) + I_c(\mathcal{I}_2)$$

Pf: $\quad I_c(\mathcal{I}_1 \otimes \mathcal{I}_2) = I_c(R\rangle B_1B_2)_\rho \qquad |\psi\rangle_{RB_1B_2E_1E_2} = V_1 \otimes V_2 |\varphi_{A_1A_2R}\rangle$

$$= H(B_1B_2) - H(RB_1B_2)$$
$$= H(B_1B_2) - H(E_1E_2)$$
$$= H(B_1) + H(B_2) - H(E_1) - H(E_2)$$
$$- [I(B_1 : B_2) - I(E_1 : E_2)] \qquad (DPI)$$
$$\le H(B_1) - H(E_1) + H(B_2) - H(E_1)$$
$$= I_c(\mathcal{I}_1) + I_c(\mathcal{I}_2)$$

Cor If $\mathcal{I}$ degrable, $Q(\mathcal{I}) = I_c(\mathcal{I})$

Pf: $\mathcal{I}$ degrable $\Rightarrow$ $\mathcal{I}^n$ degrable

Example: (Bit flip) $\mathcal{I}_\lambda(\rho) = \lambda X\rho X + (1-\lambda)\rho$ degrable $\quad \lambda \in [0, \frac{1}{2}]$

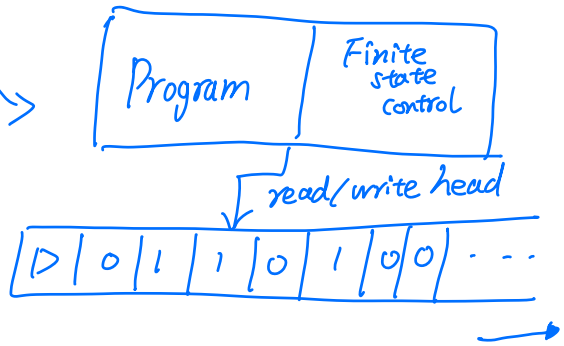
$$\mathcal{I}_\lambda^c(\rho) = \begin{bmatrix} \lambda & \sqrt{\lambda(1-\lambda)}\,\mathrm{tr}(X\rho) \\ \sqrt{\lambda(1-\lambda)}\,\mathrm{tr}(X\rho) & 1-\lambda \end{bmatrix} \qquad I_c(\mathcal{I}_\lambda) = (1-2\lambda)$$

dephasing

Example: (Schur Multiplier) $\mathcal{I}_a(\rho) = [\rho_{ij} \cdot a_{ij}]_{i,j=1}^n \qquad [a_{ij}] \ge 0$

$a_{ii} = 1 \ \forall i$

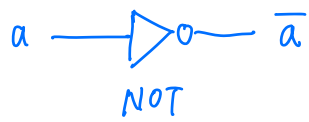
Computation model { Turing Machine, Circuits }

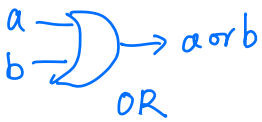


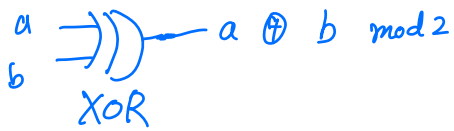
wires : carry information (binary bits)

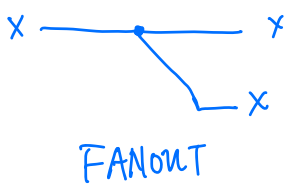
gates : simple computational tasks

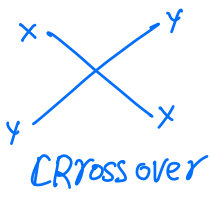
e.g.

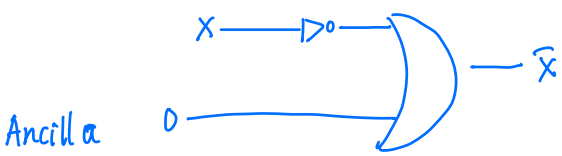












Circuit $\Longleftrightarrow$ Turing model

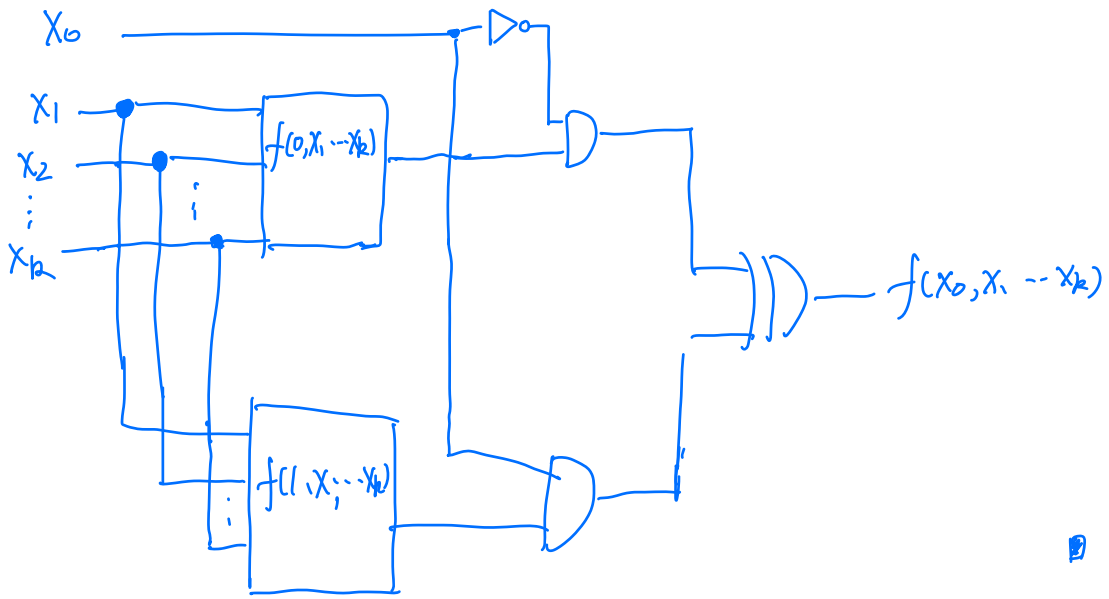
Any logic gate $f: \{0,1\}^n \longrightarrow \{0,1\}^m$ can be realized as circuit

Pf: Sufficient to consider $f: \{0,1\}^n \to \{0,1\}$ Boolean Function.

Induction on n. ① $n=1$: $f\equiv 1$ $f\equiv 0$, $f(x)=x$ $f(x)=\bar{x}$

② Assume $n=k$. For $n=k+1$,

$$f(x_0 \cdots x_k) = \begin{cases} f(0, x_1 \cdots x_k) & \text{if } x_0 = 0 \\ f(1, x_1 \cdots x_k) & \text{if } x_0 = 1 \end{cases}$$



Universal circuit construction:

① Wires ② ancilla ③ FANOUT ④ crossover ⑤ AND OR NOT

Quantum circuit

State space $H = (\mathbb{C}^2)^{\otimes n}$ n qubits

computational basis $\{|x_1 \cdots x_n\rangle \mid x_i \in \{0,1\}\} \cong \{0,1\}^n$

$|0101\rangle \longleftrightarrow 0101$

Wire $|\psi\rangle$ ——— $|\psi\rangle$

Ancilla $|0\rangle$ ———

Unitary Operations: $|\psi\rangle$ -|U|- $U|\psi\rangle$

(In computational model, we assume our gate are ideal, no noise, given by unitary)

① Single qubit gates

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

$$H=\frac{1}{\sqrt{2}}\begin{bmatrix}1&1\\1&-1\end{bmatrix} \quad S=\begin{bmatrix}1&0\\0&i\end{bmatrix} \quad T=\begin{bmatrix}1&0\\0&e^{i\pi/4}\end{bmatrix}$$

Theorem: $\forall$ unitary $U\in M_2$, $\exists\ \alpha,\beta,\gamma,\delta\in\mathbb{R}$ s.t.

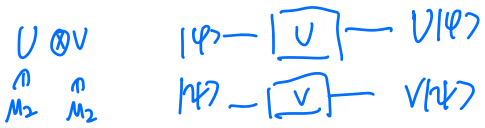
$$U=e^{i\alpha}e^{i\beta X}e^{i\gamma Y}e^{i\delta Z}$$

Pf:

$$U=e^{i\alpha}\begin{bmatrix}e^{-i(\beta+\delta)}\cos\gamma & -e^{i(\delta-\beta)}\sin\gamma\\ e^{i(\beta-\delta)}\sin\gamma & e^{i(\beta+\delta)}\cos\gamma\end{bmatrix}$$

Multiple qubits gate

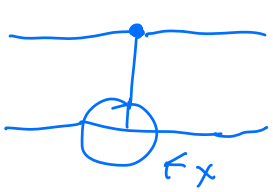
Product gate $U\otimes V$ (M_2 M_2)



Swap gate

$$F=\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}$$

Controlled-Not (non entangling)

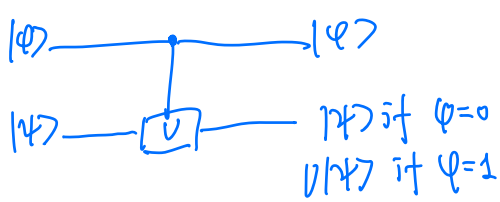


00	00
01	01
10	11
11	10

$$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix} \quad |+\rangle|0\rangle=\frac{1}{\sqrt{2}}(|00\rangle+|10\rangle)\longrightarrow\frac{1}{\sqrt{2}}|00\rangle+|11\rangle=|\Phi^+\rangle$$

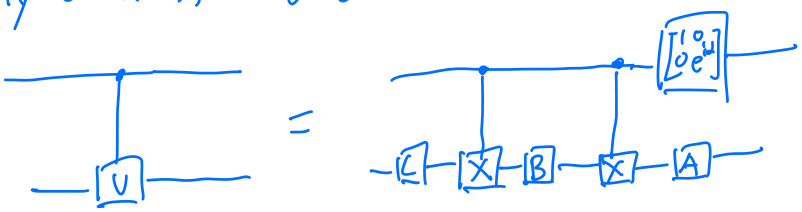
entangling

Controlled $-U$

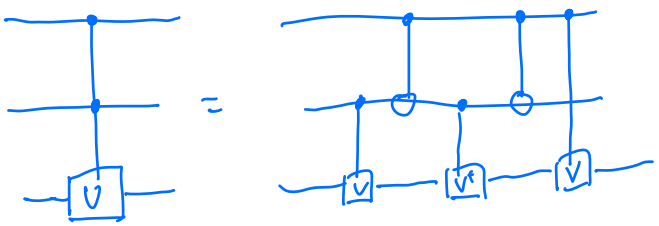


Controlled phase shift

Any $U \in U(M_2)$, $U = e^{i\alpha} AXBXC$ so



Conditional on More qubits



$V^2 = U$

No-cloning Theorem ① There is no quantum gate cloning a qubit $U|\varphi\rangle = |\varphi\rangle|\varphi\rangle$ for $\forall |\varphi\rangle \in \mathbb{C}^2$.

Pf: Suppose $U|0\rangle = |0\rangle\otimes|0\rangle$ $U|1\rangle = |1\rangle\otimes|1\rangle$

$U|+\rangle = U\left(\frac{1}{\sqrt{2}}|0\rangle + |1\rangle\right) = \frac{1}{\sqrt{2}}|00\rangle + |11\rangle \neq |+\rangle|+\rangle$

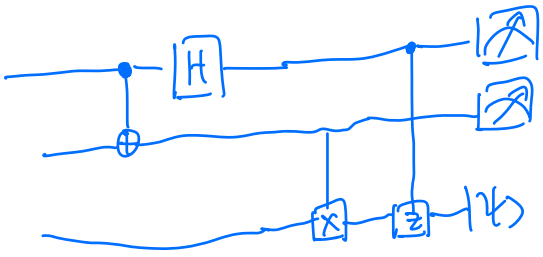
② There is no quantum channel $\mathcal{I}(\rho) = \rho\otimes\rho$ $\forall$ $\rho \in M_n$

Pf: Suppose $\mathcal{I}(\rho) = \rho\otimes\rho$ $\mathcal{I}(\sigma) = \sigma\otimes\sigma$ $\mathcal{I}\left(\frac{1}{2}\rho + \frac{1}{2}\sigma\right) = \frac{1}{2}\rho\otimes\rho + \frac{1}{2}\sigma\otimes\sigma$

Quantum Theory is linear!

Principle 1: Classical control operation $\rightsquigarrow$ quantum control operation

Principle 2: Any unterminated wires at the end of circuit can be assumed to be measure.



Since there is no-cloning, a measurement will destroy the quantum states.

Universal gate set

A set of unitary gate $S \subseteq \bigcup_{n=1}^{\infty} U(\mathbb{C}^{2^n})$ is universal if

$$\overline{\langle S \rangle} = \bigcup_{n=1}^{\infty} U(\mathbb{C}^{2^n})$$

where $\langle S \rangle = \{U_1 U_2 \cdots U_m \mid \forall m \in \mathbb{N},\ U_i \in S\}$

Namly, $\forall$ n qubit gate $U \in U(\mathbb{C}^{2^n})$, $\forall \varepsilon > 0$, $\exists$

$$\| U_1 \cdots U_m - U \| < \varepsilon, \quad U_i \in S$$

Here $\|X\| = \sup_{|h\rangle \in H} \frac{\|X|h\rangle\|}{\||h\rangle\|}$.

Universal gate set

① All 2-level unitarys : $\begin{bmatrix} 1 & & & \\ & a & b & \\ & c & d & \\ & & & 1 \end{bmatrix}$

Fact: $\forall\ U \in U(\mathbb{C}^n)$

$U = U_1 \cdots U_n$ $\quad U_i$ 2-level unitary

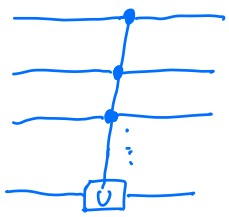
Proof: Gaussian Elimination

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \xrightarrow{U_1} \begin{bmatrix} a & 0 & c \\ d & e & f \\ g & h & i \end{bmatrix} \xrightarrow{U_2} \begin{bmatrix} a & 0 & 0 \\ d & e & f \\ g & h & i \end{bmatrix}$$

$\| $ b/c unitary

$$\begin{bmatrix} a & 0 & 0 \\ 0 & e & f \\ 0 & h & i \end{bmatrix}$$

② Single qubit gates + CNOT.

$\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & a & b \\ & & c & d \end{bmatrix}$ =  is achievable

In general, $\begin{bmatrix} 1 & & & \\ & a & \cdots & b \\ & \vdots & 1 & \vdots \\ & c & \cdots & d \end{bmatrix}$ can be converted to $\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & a & b \\ & & c & d \end{bmatrix}$

using a sequence of Control^2-Not

For example $\begin{bmatrix} a & & & b \\ & 1 & & \\ & & 1 & \\ c & & & d \end{bmatrix}$ 8×8

$|000\rangle \to |100\rangle \to |110\rangle \to U|110\rangle$

$|111\rangle \to |111\rangle \to |111\rangle \to U|111\rangle$

③ $\left\{ \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{H}, \underbrace{\begin{bmatrix} e^{i\frac{\pi}{8}} & \\ & e^{-i\frac{\pi}{8}} \end{bmatrix}}_{T}, \underbrace{\begin{bmatrix} 1 & \\ & i \end{bmatrix}}_{S}, \text{CNOT} \right\}$ discrete

Sufficient to show $\overline{\langle H, T, S \rangle} = U(M_2)$

$T = e^{-i\frac{\pi}{8}Z}$, $HTH = e^{-i\frac{\pi}{8}X}$

$e^{-i\frac{\pi}{8}Z} e^{-i\frac{\pi}{8}X} = R_{\vec{n}}(\theta)$ for $n = (\cos\frac{\pi}{8}, \sin\frac{\pi}{8}, \cos\frac{\pi}{8})$

$R_{\vec{n}}(\theta) = e^{-i\theta \vec{n}\cdot\vec{\sigma}/2}$

$\cos(\theta/2) = \cos^2\frac{\pi}{8}$

$\frac{\theta}{2\pi}$ irrational

$\vec{n}\cdot\vec{\sigma} = n_1 X + n_2 Y + n_3 Z$

Since θ is irrational. $\{\theta^n \bmod 2\pi \mid n \in \mathbb{Z}\}$ dense in $[0, 2\pi]$

So we can approximate $R_{\vec{n}}(\alpha)$ for $\forall \alpha \in [0, 2\pi]$.

$H R_{\hat{n}}(\alpha) H = R_{\hat{m}}(\alpha)$ $\quad \vec{m} = \langle \cos\frac{\pi}{8}, -\sin\frac{\pi}{8}, \cos\frac{\pi}{8} \rangle$

$\forall U \in U(M_2)$, $U = R_{\hat{n}}(\beta) R_{\hat{m}}(\gamma) R_{\hat{n}}(\delta)$ for some β, γ, δ

How many gates needed to approximate a generic $U \in U(\mathbb{C}^{2^n})$

Exponentially many!

Suppose we have g many gates, each on f qubits, in total we have $\binom{n}{f}^g$ gates

For a circuit of m gates, we have $\binom{n}{f}^{gm} = O(n^{fgm})$ many different unitary

All pure states in $\mathbb{C}^{2^n}$ = unit complex sphere $S^{2^n} = \{(z_1, \cdots z_{2^n}) \mid |z_1 \cdots z_{2^n}| = 1\}$

A ε-neighborhood in S^{2^n} $B_\varepsilon(\varphi) = \{|\psi\rangle \in S^{2^n} \mid d(\varphi, \psi) \leq \varepsilon\}$

$\forall |\varphi\rangle \in S^{2^n}$, one need at lease one unitary U s.t. $d(\varphi, \psi) \leq \varepsilon$

$$\frac{A(S^{2^n})}{A(B_\varepsilon(\varphi))} = \frac{\sqrt{\pi}\,\Gamma(2^n - \frac{1}{2})(2^{n+1}-1)}{\Gamma(2^n)\varepsilon^{2^{n+1}-1}}$$

Since $\Gamma(2^n - \frac{1}{2}) \geq \Gamma(2^n)/2^n$, we need

$$\Omega\left(\frac{1}{\varepsilon^{2^{n+1}-1}}\right)$$

unitary. If $O(n^{fgm}) \geq \Omega\left(\frac{1}{\varepsilon^{2^{n+1}-1}}\right) \Rightarrow m = \Omega\left(\frac{2^n \log(1/\varepsilon)}{\log n}\right)$ many gates

Summary of quantum circuit

① Classical Resource: classical register and computer

② Quantum State space: $(\mathbb{C}^2)^{\otimes n}$ n qubits
(register)

③ State preparation: $x_1 \cdots x_n \longmapsto |x_1 \cdots x_n\rangle$
bit string computational basis

encode classical data into quantum register

④ Quantum gate: arbitrary U on $(\mathbb{C}^2)^{\otimes n}$
via universal gate set

⑤ Measurement: Usually in computational basis
read-off computational result to classical message.