# Note

Zanqiu Shen

March 5, 2023

# Contents

I Mathematical Fundamentals 4

1 Advent of Mathematical Symbols 5

2 Measure Theory 9

2.1 Sigma algebra . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2.2 What is a measure? . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

2.3 Not everything is lebesgue measurable . . . . . . . . . . . . . . . . . . 10

2.4 Measurable maps . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

2.5 Lebesgue integral . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2.6 Fatou' lemma . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

2.7 Lebesgue's dominated convergence theorem . . . . . . . . . . . . . . . 15

3 Linear algebra 17

3.1 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

II Machine Learning Fundamentals 18

3.2 Computational learning theory . . . . . . . . . . . . . . . . . . . . . . 19

III Quantum Fundamentals 20

4 Prerequisites 21

4.1 Hilbert Spaces and Linear Operators . . . . . . . . . . . . . . . . . . . 21

4.2 Quantum States . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22

4.3 Composite systems, partial trace, entanglement . . . . . . . . . . . . . 22

4.4 Measurements . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

4.5 Entropies . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

5 Prerequisites 24

5.1 Hilbert Spaces and Linear Operators . . . . . 24
5.2 Quantum States . . . . . 25
5.3 Composite systems, partial trace, entanglement . . . . . 25
5.4 Measurements . . . . . 26
5.5 Entropies . . . . . 26

6 Classes of Quantum Channels 27
6.1 Qubit channels . . . . . 27
6.2 Generalized dephasing channels . . . . . 30
6.2.1 Detour: Holevo information ($\mathcal{X}$-quantity) . . . . . 31
6.2.2 Some comments on last lecture . . . . . 32
6.3 Entanglement-breaking channels . . . . . 33
6.4 PPT-channels . . . . . 35
6.5 Anti-degradable channels . . . . . 36

IV Quantum Information 41

7 Estimation 42
7.1 DFE . . . . . 42

8 Sensing 43
8.1 Joint measurement of TFE via SFG . . . . . 43
8.2 Introduction . . . . . 43
8.3 Terminology . . . . . 43
8.4 Problem formulation . . . . . 43
8.5 Protocol . . . . . 43
8.6 Performance evaluation . . . . . 43

9 Imaging 44
9.1 Quantum and non-local effects offer over 40 dB noise resilience advantage towards quanutm lidar . . . . . 44

# Part I

# Mathematical Fundamentals

# Chapter 1

# Advent of Mathematical Symbols

- Kronecker delta:

$$\delta_{ij} := \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{1.1}$$

- Levi-Civita symbol:

$$\varepsilon_{ijk} := \begin{cases} 1, & (i,j,k) = (1,2,3) \text{ or } (2,3,1) \text{ or } (3,1,2) \\ -1, & (i,j,k) = (3,2,1) \text{ or } (2,1,3) \text{ or } (1,3,2) \\ 0, & \text{else.} \end{cases} \tag{1.2}$$

**Example 1.0.1.**

$$(a \times b)_i = \sum_{j,k=1}^{3} \varepsilon_{ijk} a_j b_k, \tag{1.3}$$

*where $a, b$ are three dimensional vectors and " $\times$ " denotes cross product.*

- Nabla symbol:

$$\nabla := \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{pmatrix}. \tag{1.4}$$

- Factorial: $n! := n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$.

  Recursive definition: $0! := 1$, $n! := n \cdot (n-1)!,\ n \in \mathbb{N}$.

- Gamma function:

$$\Gamma(z) := \int_0^\infty x^{z-1} \cdot \mathrm{e}^{-x} \mathrm{d}x,\ \mathrm{Re}(z) \geq 0. \tag{1.5}$$

**Property 1.0.1.**

$$\Gamma(n) = (n-1)!,\ n \in \mathbb{N};\ \Gamma(z+1) = z \cdot \Gamma(z). \tag{1.6}$$

- Composition: $(g \circ f)(x) := g(f(x))$.

- Sum symbol: $\sum_{k=1}^{n} a_k := a_1 + a_2 + \cdots + a_n$.

  Recursive defintion: $\sum_{k=1}^{0} a_k := 0$, $\sum_{k=1}^{n} a_k := \left(\sum_{k=1}^{n-1} a_k\right) + a_n$.

- Product: $\Pi_{k=1} n a_k := a_1 \cdot a_2 \cdot \cdots \cdot a_n$.

  Recursive defintion: $\Pi_{k=1}^{0} a_k := 1$, $\Pi_{k=1}^{n} a_k := \left(\Pi_{k=1}^{n-1} a_k\right) \cdot a_n$

- Restriction: $f|_A : A \to Y$. For $f : X \to Y$ and $A \subseteq X$, we define $f|_A(x) = f(x)$ for all $x \in A$.

- Pauli matrices:

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} . \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} . \sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} . \tag{1.7}$$

  **Property 1.0.2.** *We have $\sigma_k^2 = I$ and $\sigma_j \sigma_k - \sigma_k \sigma_j = 2i\varepsilon_{jkl}\sigma_l$.*

- Set brackets: $\{f(x)|x \in A\}$.

  **Example 1.0.2.**

$$\{2x + 1|x \in \{0, 1, 2, 3\}\} = \{1, 3, 5, 7\} . \tag{1.8}$$

- Big $O$: $f(x) = O(g(x))$, $(x \to a)$, which means that $|f(x)| \leq M \cdot |g(x)|$, i.e., $\limsup_{x \to a} \frac{f(x)}{g(x)} < \infty$.

$$x^2 + x + 2 = O(x^2), \ (x \to \infty) \tag{1.9}$$

$$x^2 + x + 2 = O(x^3), \ (x \to \infty). \tag{1.10}$$

- Binomial coefficient:

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdots (n-k-1)}{k!} \tag{1.11}$$

$$= \frac{n!}{k!(n-k)!}. \tag{1.12}$$

- Modulo: $x \bmod n := r \in [0, n)$ with $x = n \cdot q + r$ where $q$ is the integer.

  **Example 1.0.3.**

$$5 \bmod 3 = 2 \tag{1.13}$$

$$6 \bmod 3 = 0 \tag{1.14}$$

$$7.1 \bmod 3 = 1.1 \tag{1.15}$$

$$9.7 \bmod 2.1 = 1.3. \tag{1.16}$$

- Beta function:

$$\beta(x, y) := \int_0^1 t^{x-1}(1-t) \, \mathrm{d}t, \tag{1.17}$$

  where $x, y \in \mathbb{C}$, $\mathrm{Re}(x) > 0$ and $\mathrm{Re}(y) > 0$.

**Lemma 1.0.1** (Identity between $\beta$ func. and $\Gamma$ func.)**.**

$$\beta(x, y) = \frac{\Gamma(x) \cdot \Gamma(y)}{\Gamma(x+y)}, \tag{1.18}$$

*where* $\Gamma(\cdot)$ *is related to factorial and* $\beta(x, y)$ *is related to binomial coefficient.*

- Map arrows: $f : \ X \to Y$ where $X$ is the domain and $Y$ is the codomain. This map can also be denoted as elementwise-mapping as $x \longmapsto f(x)$.

  **Example 1.0.4.**

$$f := \mathbb{R} \to \mathbb{R} \tag{1.19}$$

$$x \longmapsto x^2. \tag{1.20}$$

- Little $o$: $f(x) = o(g(x))$, $(x \to a)$, which means $\lim_{x \to a} |\frac{f(x)}{g(x)}| = 0$.

  **Example 1.0.5.**

$$8 \cdot x^2 \neq o(x^2), \ (x \to \infty) \tag{1.21}$$

$$8 \cdot x^2 \neq o(x^3), \ (x \to \infty). \tag{1.22}$$

- Outer product (Kronecker product for vectors):

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \otimes \begin{pmatrix} w_1 & w_2 & w_3 \end{pmatrix} = \begin{pmatrix} v_1 w_1 & v_1 w_2 & v_1 w_3 \\ v_2 w_1 & v_2 w_2 & v_2 w_3 \end{pmatrix}, \tag{1.23}$$

  i.e. matrix entries $(V \otimes W)_{ij} = v_i \cdot w_j$.

- Euler's phi function: $\phi : \mathbb{N} \to \mathbb{N}$ defined as

$$\phi(n) = \text{count numbers } a \in \mathbb{N} \text{ with} \tag{1.24}$$

$$(1) \ a \leq n \tag{1.25}$$

$$(2) \ \gcd(a, n) = 1(\text{mutually prime}). \tag{1.26}$$

  **Example 1.0.6.**

$$\phi(4) = 2 \tag{1.27}$$

$$\phi(5) = 4 \tag{1.28}$$

$$\phi(p) = p - 1 \ for \ p \ prime. \tag{1.29}$$

- Laplace operator (Laplacian):

$$\Delta f(x) = \frac{\partial^2 f}{\partial x_1^2}(x) + \frac{\partial^2 f}{\partial x_2^2}(x) + \frac{\partial^2 f}{\partial x_3^2}(x), \tag{1.30}$$

  where $f : \mathbb{R}^3 \to \mathbb{R}$.

- Convolution: $(f * g)(x) := \int_{-\infty}^{\infty} f(\tau) \cdot g(x - \tau) \mathrm{d}\tau$, where $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$ and $f * g : \mathbb{R} \to \mathbb{R}$.

- Heaviside function:

  $$H(x) := \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1.31}$$

  **Property 1.0.3.**

  $$H' = \delta. \tag{1.32}$$

- Quaternions:

  $$\mathbb{H} \supseteq \mathbb{C}, \tag{1.33}$$

  where $a, b, c, d \in \mathbb{R}$, the element in $\mathbb{H}$ is $a + i \cdot b + j \cdot c + k \cdot d$ with $i^2 = -1, j^2 = -1, k^2 = -1, ijk = -1$. $\mathbb{H}$ is not commutative in multiplication, i.e., $i \cdot j = -j \cdot i$.

- Infinity: $\infty$.

  **Example 1.0.7.** *In measure theory:* $[0, \infty]$*. We have*

  $$a + \infty = \infty + a = \infty \ for \ a \in [a, \infty] \tag{1.34}$$

  $$\tag{1.35}$$

- means equivalence reltation. For example, $x$ $y$ means $x$ is equivalent to $y$ for some conditions.

# Chapter 2

# Measure Theory

## 2.1 Sigma algebra

**Example 2.1.1.** *We define $\mathcal{P}(X)$ as the power set of set $X$. Assume that set $X = \{a, b\}$, the power set $P(X)$ would be $\{\emptyset, X, \{a\}, \{b\}\}$*

**Definition 2.1.1** (Sigma algebra)**.** $\mathcal{A} \subseteq P(X)$ *is called a* $\sigma -$ algebra*:*

$$(a)\ \emptyset, X \in \mathcal{A} \tag{2.1}$$

$$(b)\ A \in \mathcal{A} \Longrightarrow A^c := X \setminus A \in \mathcal{A} \tag{2.2}$$

$$(c)\ A_i \in \mathcal{A},\ i \in \mathcal{N} \Longrightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{A}. \tag{2.3}$$

**Definition 2.1.2** (Measurable sets)**.** *$A \in \mathcal{A}$ is called a $\mathcal{A}$-measurable set.*

**Example 2.1.2.**

$$(1)\ \mathcal{A} = \{\emptyset, X\} \tag{2.4}$$

$$(2)\ \mathcal{A} = \{P(X)\}. \tag{2.5}$$

**Lemma 2.1.1.** *Assume $A_i$ is $\sigma$-algebra on $X$, $i \in I$(index set). Then, we have $\cap_{i \in I} \mathcal{A}_i$ is also a $\sigma$-algebra on $X$.*

**Definition 2.1.3** (Sigma algebra generated by $\mathcal{M}$)**.** *For $\mathcal{M} \subseteq P(X)$, there is a smallest $\sigma$-algebra that contains $\mathcal{M}$:*

$$\sigma(\mathcal{M}) := \cap_{\mathcal{A} \supseteq \mathcal{M},\ A\ \sigma-algebra} \mathcal{A}. \tag{2.6}$$

**Example 2.1.3.** *We define $X = \{a, b, c, d\}$ and $\mathcal{M} = \{\{a\}, \{b\}\}$. Then we have*

$$\sigma(\mathcal{M}) = \{\emptyset, X, \{a\}, \{b\}, \{a, b\}, \{b, c, d\}, \{a, c, d\}, \{c, d\}\}. \tag{2.7}$$

**Definition 2.1.4** (Borel sigma algebra)**.** *Let $(X, \mathcal{T})$ be a topological space (Let $X$ be a metric space/Let $X$ be a subset of $\mathbb{R}^n$; We need "open sets".). We then define $\mathcal{B}(X)$ is the borel $\sigma$-algebra on $X$ as*

$$\mathcal{B}(X) := \sigma(\mathcal{T}), \tag{2.8}$$

*which is the $\sigma$-algebra generated by the open sets $\mathcal{T}$.*

## 2.2 What is a measure?

**Definition 2.2.1** (Measure)**.** *$(X, \mathcal{A})$ is called a measurable space, where $X$ is a set and $\mathcal{A}$ is a $\sigma$-algebra on $X$. A map $\mu : A \to [0, \infty] := [0, \infty) + \{\infty\}$ is called a measure if it satisfies:*

$$(a)\ \mu(\emptyset) = 0 \tag{2.9}$$

$$(b)\ \mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i) \ with\ A_i \cap A_j = \emptyset,\ i \neq j\ for\ all\ A_i \in \mathcal{A}.(\sigma - additive) \tag{2.10}$$

**Definition 2.2.2.** *$(X, \mathcal{A}, \mu)$ is called a measure space.*

**Example 2.2.1.** *Given $X$ and $\mathcal{A} = \mathcal{P}(X)$.*

- *Counting measure ($A \in \mathcal{A}$) is defined as*

  $$\mu(A) := \begin{cases} \#A, & A\ has\ finitely\ many\ elements \\ \infty & else \end{cases} \tag{2.11}$$

  *where $\#A$ means the number of elements in $A$.*

  *Calculation rules in $[0, \infty]$:*

  $$x + \infty := \infty\ for\ all\ x \in [0, \infty] \tag{2.12}$$
  $$x \cdot \infty := \infty\ for\ all\ x \in (0, \infty] \tag{2.13}$$
  $$0 \cdot \infty := 0\ (only\ true\ in\ most\ cases\ in\ measure\ theory!) \tag{2.14}$$

- *Dirac measure for $p \in X$ is defined as*

  $$\delta_p(A) := \begin{cases} 1, & p \in A \\ 0, & else \end{cases} \tag{2.15}$$

- *We search a measure on $X \in \mathcal{R}^n$ satisfying:*

  $$(1)\ \mu([0, 1]^n) = 1 \tag{2.16}$$
  $$(2)\ \mu(x + A) = \mu(A)\ for\ all\ x \in \mathcal{R}^n, \tag{2.17}$$

  *which is known as Lebesgue measure where the $\sigma$-algebra is not equal to power set.*

## 2.3 Not everything is lebesgue measurable

**Measure problem:** search measure $\mu$ on $\mathcal{P}(\mathbb{R})$ with:

- (1) $\mu([a, b]) = b - a,\ b > a,$
- (2) $\mu(x + A) = \mu(A),\ A \in \mathcal{P}(\mathbb{R}),\ x \in \mathbb{R}.$

$\implies \mu$ does not exist.

**Claim:** Let $\mu$ be a measure on $\mathcal{P}(\mathbb{R})$ with $\mu((0, 1]) < \infty$ and (2). $\implies \mu = 0$.

*Proof.* (a) Definitions: $I \in (0,1]$ with equivalence relation on $I$: $x\ y \iff x-y \in \mathbb{Q}$ i.e., $[x] := \{x+r|r \in \mathbb{Q},\ x+r \in I\}$. Following this definition, we have a disjoint decomposition of $I$ into boxes, possibly uncontable many of them! We then pick one element $a_n$ from each box $[x_n]$ and form a set $A \in I$, i.e., $\{a_1, a_2, \cdots\} = A$. We have $A \in I$ with prperty:

- (1) For each $[x]$, there is an $a \in A$ with $a \in [x]$.
- (2) For all $a, b \in A:\ a, b \in [x] \Longrightarrow a = b$.

In uncountable case, the existence of $A \in I$ with the above property is guaranted by the axiom of choice of set theory.

We define $A_n := r_n + A$, where $(r_n)_{n\in\mathbb{N}}$ enumeration of $\mathbb{Q}_n(-1,1]$.

(b) We then claim that $A_n \cap A_m = \emptyset \Longleftarrow n \neq m$. The proof is as follows: $x \in A_n \cap A_m \Longrightarrow x = r_n + a_n,\ a_n \in A$ and $x = r_m + a_m,\ a_m \in A. \Longrightarrow r_n + a_n = r_m + a_m \Longrightarrow a_n - a_m = r_n - r_m \in \mathbb{Q} \Longrightarrow a_n\ a_m \Longrightarrow a_m, a_n \in [a_m] \Longrightarrow a_n = a_m \Longrightarrow r_n = r_m \Longrightarrow n = m$.

(c) We claim that $(0,1] \subseteq \cup_{n\in\mathbb{N}} A_n \subseteq (-1,2]$. The proof is as follows:

Assume now: $\mu$ measure on $\mathcal{P}(\mathbb{R})$ with $\mu((0,1]) < \infty$ and (2).

By (2): $\mu(1+A) = \mu(A)$ for all $n \in \mathbb{N}$.

By (c): we have

$$\mu((0,1]) \leq \mu(\cup_{n\in\mathbb{N}}) \leq \mu((-1,2]) \tag{2.18}$$

We know: $\mu((0,1]) =: C < \infty$. By using (2) and $\sigma$-additivity, we get $\mu((-1,2]) = \mu\left((-1,0] \cup (0,1] \cup (1,2] = 3C\right)$. $\Longrightarrow_{2.18,(b)} C \leq \sum_{n=1}^{\infty} \mu(A_n) \leq 3C \Longrightarrow C \leq \sum_{n=1}^{\infty} \mu(A) \leq 3C \Longrightarrow \mu(A) = 0 \Longrightarrow C = 0(\text{henceL } \mu\left((0,1]\right) = 0) \Longrightarrow \mu(\mathbb{R}) = \mu\left(\cup_{n\in\mathbb{Z}}(m, m+1]\right) = 0 \Longrightarrow \mu = 0$. □

## 2.4 Measurable maps

**Definition 2.4.1** (Measurable maps). *$(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ are measurable spaces. $f : \Omega_1 \to \Omega_2$ is a measurable map w.r.t. $\mathcal{A}_1$ and $\mathcal{A}_2$ if $f^{-1}(A_2) \in \mathcal{A}_1$ for all $A_2 \in \mathcal{A}_2$.*

**Example 2.4.1.**

- *$(\Omega, \mathcal{A})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are two measurable spaces. We define characteristic fucntion (aksi indicator function) as $\chi_A : \Omega \to \mathbb{R}$, where*

$$\chi_A(w) := \begin{cases} 1, & w \in A \\ 0, & w \notin A \end{cases} \tag{2.19}$$

*For all measurable $A \in \mathcal{A}$, $\chi_A$ is a measurable map. We have*

$$\chi_A^{-1}(\emptyset) = \emptyset \in \mathcal{A},\ \chi_A^{-1}(\mathbb{R}) = \Omega \in \mathcal{A} \tag{2.20}$$

$$\chi_A^{-1}(\{A\}) = A,\ \chi_A^{-1}(\{0\}) = A^c \in \mathcal{A}. \tag{2.21}$$

- *Composition of measurable maps.*

**Lemma 2.4.1.** $(\Omega_1, \mathcal{A}_\infty)$, $(\Omega_2, \mathcal{A}_\in)$, $(\Omega_3, \mathcal{A}_\ni)$ *are measurable space. We define* $\Omega_1 \xrightarrow{f} \Omega_2 \xrightarrow{g} \Omega_3$*. Then* $f, g$ *are measurable implies* $g \circ f$ *is measurable.*

*Proof.*

$$\begin{aligned} (g \circ f)^{-1}(A_3) &= f^{-1}(g^{-1}(A_3)) && (2.22)\\ &\in \mathcal{A}_1 && (2.23) \end{aligned}$$

□

**Important measurable maps**

**Lemma 2.4.2.** $(\Omega, \mathcal{A})$ *and* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *are measurable spaces.* $f, g : \Omega \to \mathbb{R}$ *are measurable maps indicates that* $f + g,\ f - g,\ f \cdot g,\ |f|$ *are measurable maps.*

## 2.5 Lebesgue integral

**Example 2.5.1.** *Define Characteristic function* $\chi_A : X \to \mathbb{R},\ A \in \mathcal{A}$*. We define* $I(A) := \mu(A)$*. Surprisingly,* $I(A)$ *is nothing but the integral of* $\chi_A$ *over* $A$*.*

**Definition 2.5.1** (Simple/Step/Staircsae functions,...)**.** *For* $A_1, A_2, \ldots, A_n \in \mathcal{A}$*, and* $c_1, c_2, \cdots, c_n \in \mathbb{R}$*. We define*

$$f(x) := \sum_{i=1}^{n} c_i \cdot \chi_{A_i}(x). \tag{2.24}$$

*We then have* $f(x)$ *is measurable and the integraal of* $f$ *is defined as* $I(f) := \sum_{i=1}^{n} c_i \mu(A_i)$*.*

**Remark 2.5.1.** *The problem of the integral* $I(f)$ *is that it is undefined when* $\mu(A_i) = \infty$*. The problem can be solved by exclude* $\infty$ *by defintion or the following way.*

**Definition 2.5.2** (Lebesgue integral)**.** *Define* $S^+ := \{f : X \to \mathbb{R} | f\ simple\ function,\ f \geq 0\}$*.* $f \in S^+$ *and choose representation* $f(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x),\ c_i \geq 0$*. The lebesgue integral of* $f$ *w.r.t.* $\mu$ *is defined as*

$$\begin{aligned} \int_X f(x)\ \mathrm{d}\mu(x) &= \int_X f\ \mathrm{d}\mu && (2.25)\\ &= I(f) && (2.26)\\ &= \sum_{i=1}^{n} c_i \cdot \mu(A_i) && (2.27)\\ &= [0, \infty]. && (2.28) \end{aligned}$$

**Property 2.5.1.**

- $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g),\ \alpha, \beta \geq 0.$
- $f \leq g \Longrightarrow I(f) \leq I(g)$ *(monotomicity)*

**Definition 2.5.3.** *Define a measurable map* $f : X \to [0, \infty)$*.* $h = \sum_{i=1}^{n} c_i \cdot \chi_{A_i}$*. The lebesgue integral of* $f$ *w.r.t.* $\mu$ *is defined as*

$$\begin{aligned} \int_X f\ \mathrm{d}\mu &:= \sup\left\{I(h) | h \in S^+,\ h \leq f\right\} && (2.29)\\ &\in [0, \infty]. && (2.30) \end{aligned}$$

$f$ *is called* $\mu$*-integrable if* $\int_X f\ \mathrm{d}\mu < \infty$*.*

**Property 2.5.2.** *Define measurable maps $f, g : X \to [0, \infty)$, we have*

- *1. $f = g$ for $\mu$-almost everywhere(a.e.), which satisfies $\mu\left(\{x \in X | f(x) \neq g(x)\}\right) = \Longrightarrow \int_X f \ \mathrm{d}\mu = \int_X g \ \mathrm{d}\mu$.*
- *2. $f \leq g$ for $\mu$ a.e. $\Longrightarrow \int_X f \ \mathrm{d}\mu \leq \int_X g \ \mathrm{d}\mu$*
- *3. $f = 0$ for $\mu$-a.e. $\Longleftrightarrow \int_X f \ \mathrm{d}\mu = 0$.*

*Proof of 2.: monotonicity.* Let $h := X \to [0, \infty)$ be a simple function, i.e.,

$$h(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x) \tag{2.31}$$

$$= \sum_{t \in h(X)} t \cdot \chi_{\{x \in X | h(x) = t\}}. \tag{2.32}$$

Let $X = \tilde{X}^c \cup \tilde{X}$ with $\mu(\tilde{X}^c) = 0$,

$$\tilde{h}(x) := \begin{cases} h(x), & x \in \tilde{X} \\ a, & x \in \tilde{X}^c \end{cases} \tag{2.33}$$

$$\tilde{h}(x) = \sum_{t \in h(X)} t \cdot \chi_{\left\{x \in \tilde{X} | h(x) = t\right\}} + a \cdot \chi_{\tilde{X}^c} \tag{2.34}$$

$$I(\tilde{h}) = \sum_{t \in h(X)} t \cdot \mu(\left\{x \in \tilde{X} | h(x) = t\right\}) + a \cdot \mu(\tilde{X}^c) \tag{2.35}$$

$$= \sum_{t \in h(X)} t \left[\mu\left(\left\{x \in \tilde{X} | h(x) = t\right\}\right) + \mu\left(\left\{x \in \tilde{X}^c | h(x) = t\right\}\right)\right] \tag{2.36}$$

$$= \sum_{t \in h(X)} t \left[\mu\left(\left\{x \in \tilde{X} | h(x) = t\right\} \cup \left\{x \in \tilde{X}^c | h(x) = t\right\}\right)\right] \tag{2.37}$$

$$I(h) = \sum_{t \in h(X) \setminus \{0\}} t \cdot \mu\left(\{x \in X | h(x) = t\}\right). \tag{2.38}$$

We define

$$\tilde{X} := \{x \in X | f(x) \leq g(x)\}, \tag{2.39}$$

$$\mu(\tilde{X}^c) = 0 \tag{2.40}$$

$$\int_X f \ \mathrm{d}\mu = \sup\left\{I(h) | h \in S^+, h \leq f\right\} \tag{2.41}$$

$$= \sup\{I(\tilde{h}) | \tilde{h} \in S^+, \tilde{h} \leq f \ on \ \tilde{X}\} \tag{2.42}$$

$$\leq \sup\{I(\tilde{h}) | \tilde{h} \in S^+, h \leq g \ on \ \tilde{X}\} \tag{2.43}$$

$$= \sup\{I(h) | h \in S^+, h \leq g\} \tag{2.44}$$

$$= \int_X g \ \mathrm{d}\mu. \tag{2.45}$$

□

**Theorem 1** (Monotone convergence theorem)**.** $(X, \mathcal{A}, \mu)$ *measurable spaces,* $f_n : X \to [0, \infty]$*,* $(f : X \to [0, \infty])$ *measurable for all* $n \in \mathbb{N}$ *with*

$$f_1 \leq f_2 \leq f_3 \leq \cdots \quad \mu - a.e. \tag{2.46}$$

$$(\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu \quad \mu - a.e.(x \in X)) \tag{2.47}$$

*This implies that*

$$\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu = \int_X \lim_{n\to} f_n \, \mathrm{d}\mu. \tag{2.48}$$

*Proof.* $\int_X f_1 \, \mathrm{d}\mu \leq \int_X f_2 \, \mathrm{d}\mu \leq \cdots$ and $\int_X f_n \, \mathrm{d}\mu \leq \int_X f \, \mathrm{d}\mu$ for $n \in \mathbb{N}$. Then we have

$$\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu \leq \int_X f \, \mathrm{d}\mu, \tag{2.49}$$

which is the first part of 2.48.

Let $h$ be a simple function $0 \leq f \leq f$ and $\varepsilon > 0$. We define

$$X_n := \{x \in X | f_n(x) \geq (1 - \varepsilon)h(x)\} \tag{2.50}$$

with $\cup_{n=1}^{\infty} X_n = \tilde{X}$, and $\mu(\tilde{X}^c) = 0$. We have

$$\int_X f_n \, \mathrm{d}\mu \geq \int_{X_n} f_n \, \mathrm{d}\mu \geq \int_{X_n} (1 - \varepsilon)h \, \mathrm{d}\mu \tag{2.51}$$

$$\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu \geq \lim_{n\to\infty} \int_{X_n} (1 - \varepsilon)h \, \mathrm{d}\mu \tag{2.52}$$

$$= \int_{\tilde{X}} (1 - \varepsilon)h \, \mathrm{d}\mu \tag{2.53}$$

$$= \int_X (1 - \varepsilon)h \, \mathrm{d}\mu. \tag{2.54}$$

This implies

$$\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu \geq \int_X h \, \mathrm{d}\mu, \tag{2.55}$$

since $\varepsilon > 0$ arbitrarily. Then we have

$$\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu \geq \int_X f \mathrm{d}\mu, \tag{2.56}$$

since $h$ is arbitrary and $h \leq f$, which is second part of 2.48. □

**Applictions** Given a series $(g_n)_{n\in\mathbb{N}}$, $g_n : X \to [0, \infty]$ measurable for all $n$. Then we have $\sum_{n=1}^{\infty} g_n : X \to [0, \infty]$ measurable and

$$\int_X \sum_{n=1}^{\infty} g_n \, \mathrm{d}\mu = \sum_{n=1}^{\infty} \int_X g_n \, \mathrm{d}\mu, \tag{2.57}$$

which means the integral and sum can exchange.

## 2.6 Fatou' lemma

**Lemma 2.6.1.** *Given $(X, \mathcal{A}, \mu)$ measurable space, $f_n : X \to [0, \infty]$ measurable for all $n \in \mathbb{N}$. Then we have*

$$\int_X \liminf_{n\to\infty} f_n \ \mathrm{d}\mu \le \liminf_{n\to\infty} \int_X f_n \ \mathrm{d}\mu. \tag{2.58}$$

**Remark 2.6.1.** $\liminf_{n\to\infty} f_n : X \to [0, \infty]$ *is a function. This is*

$$\begin{aligned} g(x) &:= \left(\liminf_{n\to\infty} f_n\right)(x) && (2.59)\\ &:= \lim_{n\to\infty} \left(\inf_{k\ge n} f_k(x)\right) && (2.60)\\ &\in [0, \infty] && (2.61)\\ g_n(x) &:= \inf_{k\ge n} f_k(x). && (2.62) \end{aligned}$$

*We have*

$$g_1 \le g_2 \le g_3 \le \cdots, \tag{2.63}$$

*which is monotonically increasing. All these functions are measurable.*

*Proof.*

Since (1),

$$\begin{aligned} \int_X \lim_{n\to\infty} g_n \ \mathrm{d}\mu &= \lim_{n\to\infty} \int_X g_n \ \mathrm{d}\mu && (2.64)\\ &= \liminf_{n\to\infty} \int_X g_n \ \mathrm{d}\mu. && (2.65) \end{aligned}$$

We know that $g_n \le f_n$ for all $n \in \mathbb{N}$. By (2.5.2), we have

$$\int_X g_n \ \mathrm{d}\mu \le \int_X f_n \ \mathrm{d}\mu, \tag{2.66}$$

for all $n \in \mathbb{N}$. Then we have

$$\begin{aligned} \int_X \liminf_{n\to\infty} f_n \ \mathrm{d}\mu &= \liminf_{n\to\infty} \int_X g_n \ \mathrm{d}\mu && (2.67)\\ &\le \liminf_{n\to\infty} \int_X f_n \ \mathrm{d}\mu. && (2.68) \end{aligned}$$

□

## 2.7 Lebesgue's dominated convergence theorem

$(X, \mathcal{A}, \mu)$, $\mathcal{L}^1 := \left\{f : X \to \mathbb{R} \text{ } \textit{measurable} \middle| \int_X |f|^1 \ \mathrm{d}\mu < \infty\right\}$. For $f \in \mathcal{L}^1(\mu)$, write $f = f^+ - f^-$, where $f^+, f^- \ge 0$. Define $\int_X f \ \mathrm{d}\mu := \int_X f^+ \ \mathrm{d}\mu - \int_X f^- \ \mathrm{d}\mu$.

**Theorem 2** (Lebesgue's dominated convergence theorem)**.** *$f_n : X \to \mathbb{R}$ measurable for all $n \in \mathbb{N}$. $f : X \to \mathbb{R}$ with $\overset{n\to\infty}{f(x)}$ for $x \in X$ ($\mu$-a.e.) and $|f_n| \leq g$ with $g \in \mathcal{L}^1(\mu)$ for all $n \in \mathbb{N}$, where $g$ is called integral majorant. Then: $f_1, f_2, \dots \in \mathcal{L}^1(\mu)$, $f \in \mathcal{L}^1(\mu)$ and*

$$\lim_{n\to\infty} \int_X f_n \;\mathrm{d}\mu = \int_X f \;\mathrm{d}\mu. \tag{2.69}$$

*Proof.* □

# Chapter 3

# Linear algebra

## 3.1 Exercises

- 3.C.6: Are the construced $v_i$ form a basis of vector space $V$?
-

# Part II

# Machine Learning Fundamentals

## 3.2 Computational learning theory

# Part III

# Quantum Fundamentals

# Chapter 4

# Prerequisites

## 4.1 Hilbert Spaces and Linear Operators

Throughout this course, $\mathcal{H}$ denotes a finite-dimensional Hilbert space (complex vector space with an associated inner product). Using Dirac's "bra-ket" notation we denote elements of the Hilbert space (called kets) as

$$|\psi\rangle \in \mathcal{H}. \tag{4.1}$$

The elements of the dual Hilbert space are called bras and are denoted

$$\langle\psi| \in \mathcal{H}^*, \tag{4.2}$$

where $\langle\psi| = (|\psi\rangle)^\dagger$. Here, $X^\dagger := \bar{X}^T$ denotes the Hermitian adjoint (also called the conjugate transpose). We denote

$$B(\mathcal{H}_1, \mathcal{H}_2) := \{\text{linear maps from } \mathcal{H}_1 \text{ to } \mathcal{H}_2\} \tag{4.3}$$

and the set of all linear maps to and from the same space will be denoted $B(\mathcal{H}) = B(\mathcal{H}, \mathcal{H})$. An operator $X \in B(\mathcal{H})$ is *normal* if $XX^T = X^TX$. Every normal operator has a *spectral decomposition.* That is, there exists a unitary $U$ and a diagonal matrix $D$ whose entries are the eigenvalues $\lambda_1, \ldots, \lambda_d \in \mathbb{C}$ of $X$ such that

$$X = UDU^\dagger. \tag{4.4}$$

In other words,

$$X = \sum_{i=1}^{d} \lambda_i |\psi_i\rangle \langle\psi_i| \tag{4.5}$$

where $X|\psi_i\rangle = \lambda_i |\psi_i\rangle$ and $U = (|\psi_i\rangle, \ldots, |\psi_d\rangle)$. If $X$ is Hermitian, $X = X^\dagger$, then $\lambda_i \in \mathbb{R}$. An operator $X$ is positive semi-definite (PSD) if

$$\langle\varphi| X |\varphi\rangle \geq 0 \qquad \forall |\varphi\rangle \in \mathcal{H}. \tag{4.6}$$

As a consequence, $X \geq 0$ and $\lambda_i \geq 0$. It holds that PSD $\implies$ Hermitian $\implies$ normal. Unless otherwise stated, we will always assume we are working in an orthonormal basis.

## 4.2 Quantum States

A quantum state $\rho$ in a Hilbert space $\mathcal{H}$ is a PSD linear operator with

$$\rho \in B(\mathcal{H}), \quad \rho \geq 0, \quad \mathrm{tr}\rho = 1. \tag{4.7}$$

This means that the state has eigenvalues $\{\lambda_i\}_{i=1}^d$ satisfying $\lambda_i \geq 0$ and $\sum_{i=1}^d \lambda_i = 1$. Thus, $\{\lambda_i\}_{i=1}^d$ forms a probability distribution.

A *pure quantum state* $\psi$ is a quantum state with rank 1. We can find $|\psi\rangle \in \mathcal{H}$ such that $\psi = |\psi\rangle\langle\psi|$. In this case, $\psi$ is called a *projector.* A *mixed state* is a quantum state with rank $> 1$. Mixed states are convex combinations of pure states. That is, for every quantum state $\rho$ with $r = \mathrm{rank}(\rho)$ there are pure states $|\psi_i\rangle_{i=1}^k \quad (k \geq r)$ and a probability distribution $\{p_i\}_{i=1}^k$ such that

$$\rho = \sum_{i=1}^k p_i |\psi_i\rangle\langle\psi_i|. \tag{4.8}$$

The spectral decomposition of $\rho$ is a special case of this property.

## 4.3 Composite systems, partial trace, entanglement

Let $A$ and $B$ be two quantum systems with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. The *joint system* $AB$ is described by the Hilbert space $\mathcal{H}_{AB} := \mathcal{H}_A \otimes \mathcal{H}_B$. We denote quantum states of the joint system as $\rho_{AB} \in \mathcal{H}_{AB}$. The marginal of the bipartite state, denoted $\rho_A$, is uniquely defined as the operator satisfying

$$\rho_A := \mathrm{tr}_B \rho_{AB}, \tag{4.9}$$

which is defined via $\mathrm{tr}(\rho_{AB}(X_A \otimes \mathbb{I}_B)) = \mathrm{tr}\rho_A X_A \quad \forall X_A \in B(\mathcal{H}_A)$. For a Hilbert space with $|B| := \dim \mathcal{H}_B$, the explicit form of the partial trace is

$$\mathrm{tr}_B \rho_{AB} = \sum_{i=1}^{|B|} (\mathbb{I}_A \otimes \langle i|_B) \rho_{AB} (\mathbb{I}_A \otimes |i\rangle_B), \tag{4.10}$$

for some basis $\{|i\rangle_B\}_{i=1}^{|B|}$ of $\mathcal{H}_B$.

A *product state* on $AB$ is a state of the form $\rho_A \otimes \sigma_B$. The state is called *separable* if it lies in the convex hull of product states:

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \sigma_B^i \tag{4.11}$$

for some states $\{\rho_A^i\}_i$ and $\{\sigma_B^i\}_i$ and probability distribution $\{p_i\}_i$. A state is called *entangled*, if it is not separable. An entangled state of particular interest is the *maximally entangled state.* Let $d = \dim \mathcal{H}$, $\{|i\rangle\}_{i=1}^d$ be a basis for $\mathcal{H}$. A maximally entangled state is expressed as

$$|\phi^+\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^d |i\rangle \otimes |i\rangle \quad \in \mathcal{H} \otimes \mathcal{H} \tag{4.12}$$

## 4.4 Measurements

The most general measurement is given by a *positive operator-valued measure* (POVM) $E = \{E_i\}_i$ where $E_i \geq 0 \quad \forall i$ and $\sum_i E_i = \mathbb{I}$. Then, for a quantum system $\mathcal{H}$ in state $\rho$, the probability of obtaining measurement outcome $i$ is given by $p_i = \mathrm{tr}[\rho E_i]$. So, we have

$$\sum_i p_i = \sum_i \mathrm{tr}[\rho E_i] = \mathrm{tr}\left[\rho \sum_i E_i\right] = \mathrm{tr}[\rho \mathbb{I}] = \mathrm{tr}\rho = 1, \tag{4.13}$$

for all normalized quantum states. A *projective measurement* $\Pi = \{\Pi_i\}$ is a POVM with the added property of orthogonality, which for projectors means

$$\Pi_i \Pi_j = \delta_{ij} \Pi_i. \tag{4.14}$$

Any basis $\{|e_i\rangle\}_{i=1}^{\dim \mathcal{H}}$ gives rise to a projective measurement $\Pi = \{|e_i\rangle \langle e_i|\}_{i=1}^{\dim \mathcal{H}}$.

## 4.5 Entropies

The *Shannon entropy* $H(p)$ of a probability distribution $p = \{p_i, \ldots, p_d\}$ is defined as $H(p) = -\sum_{i=1}^d p_i \log p_i$, where the logarithm is base 2 unless otherwise specified. Note that when the logarithm is base 2, the entropy has units of *bits*. The *von Neumann entropy* $S(\rho)$ of a quantum state $\rho$ is defined as

$$S(\rho) = -\mathrm{tr}\left[\rho \log \rho\right] = H(\{\lambda_i, \ldots, \lambda_d\}), \tag{4.15}$$

where $\rho = \sum_i \lambda_i |\psi_i\rangle \langle \psi_i|$ is a spectral decomposition of $\rho$ and where the logarithm of an operator is obtained by first diagonalizing the matrix representing the operator and then taking the logarithm of the diagonal elements. That is,

$$\log \rho = \sum_{i:\lambda_i > 0} \log(\lambda_i) |\psi_i\rangle \langle \psi_i|. \tag{4.16}$$

# Chapter 5

# Prerequisites

## 5.1 Hilbert Spaces and Linear Operators

Throughout this course, $\mathcal{H}$ denotes a finite-dimensional Hilbert space (complex vector space with an associated inner product). Using Dirac's "bra-ket" notation we denote elements of the Hilbert space (called kets) as

$$|\psi\rangle \in \mathcal{H}. \tag{5.1}$$

The elements of the dual Hilbert space are called bras and are denoted

$$\langle\psi| \in \mathcal{H}^*, \tag{5.2}$$

where $\langle\psi| = (|\psi\rangle)^\dagger$. Here, $X^\dagger := \bar{X}^T$ denotes the Hermitian adjoint (also called the conjugate transpose). We denote

$$B(\mathcal{H}_1, \mathcal{H}_2) := \{\text{linear maps from } \mathcal{H}_1 \text{ to } \mathcal{H}_2\} \tag{5.3}$$

and the set of all linear maps to and from the same space will be denoted $B(\mathcal{H}) = B(\mathcal{H}, \mathcal{H})$. An operator $X \in B(\mathcal{H})$ is *normal* if $XX^T = X^TX$. Every normal operator has a *spectral decomposition*. That is, there exists a unitary $U$ and a diagonal matrix $D$ whose entries are the eigenvalues $\lambda_1, \dots, \lambda_d \in \mathbb{C}$ of $X$ such that

$$X = UDU^\dagger. \tag{5.4}$$

In other words,

$$X = \sum_{i=1}^{d} \lambda_i |\psi_i\rangle \langle\psi_i| \tag{5.5}$$

where $X|\psi_i\rangle = \lambda_i |\psi_i\rangle$ and $U = (|\psi_i\rangle, \dots, |\psi_d\rangle)$. If $X$ is Hermitian, $X = X^\dagger$, then $\lambda_i \in \mathbb{R}$. An operator $X$ is positive semi-definite (PSD) if

$$\langle\varphi| X |\varphi\rangle \geq 0 \qquad \forall |\varphi\rangle \in \mathcal{H}. \tag{5.6}$$

As a consequence, $X \geq 0$ and $\lambda_i \geq 0$. It holds that PSD $\implies$ Hermitian $\implies$ normal. Unless otherwise stated, we will always assume we are working in an orthonormal basis.

## 5.2 Quantum States

A quantum state $\rho$ in a Hilbert space $\mathcal{H}$ is a PSD linear operator with

$$\rho \in B(\mathcal{H}), \quad \rho \geq 0, \quad \mathrm{tr}\rho = 1. \tag{5.7}$$

This means that the state has eigenvalues $\{\lambda_i\}_{i=1}^d$ satisfying $\lambda_i \geq 0$ and $\sum_{i=1}^d \lambda_i = 1$. Thus, $\{\lambda_i\}_{i=1}^d$ forms a probability distribution.

A *pure quantum state* $\psi$ is a quantum state with rank 1. We can find $|\psi\rangle \in \mathcal{H}$ such that $\psi = |\psi\rangle\langle\psi|$. In this case, $\psi$ is called a *projector*. A *mixed state* is a quantum state with rank $> 1$. Mixed states are convex combinations of pure states. That is, for every quantum state $\rho$ with $r = \mathrm{rank}(\rho)$ there are pure states $|\psi_i\rangle_{i=1}^k \quad (k \geq r)$ and a probability distribution $\{p_i\}_{i=1}^k$ such that

$$\rho = \sum_{i=1}^k p_i |\psi_i\rangle\langle\psi_i| \,. \tag{5.8}$$

The spectral decomposition of $\rho$ is a special case of this property.

## 5.3 Composite systems, partial trace, entanglement

Let $A$ and $B$ be two quantum systems with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. The *joint system* $AB$ is described by the Hilbert space $\mathcal{H}_{AB} := \mathcal{H}_A \otimes \mathcal{H}_B$. We denote quantum states of the joint system as $\rho_{AB} \in \mathcal{H}_{AB}$. The marginal of the bipartite state, denoted $\rho_A$, is uniquely defined as the operator satisfying

$$\rho_A := \mathrm{tr}_B \rho_{AB}, \tag{5.9}$$

which is defined via $\mathrm{tr}(\rho_{AB}(X_A \otimes \mathbb{I}_B)) = \mathrm{tr}\rho_A X_A \quad \forall X_A \in B(\mathcal{H}_A)$. For a Hilbert space with $|B| := \dim \mathcal{H}_B$, the explicit form of the partial trace is

$$\mathrm{tr}_B \rho_{AB} = \sum_{i=1}^{|B|} (\mathbb{I}_A \otimes \langle i|_B) \rho_{AB} (\mathbb{I}_A \otimes |i\rangle_B), \tag{5.10}$$

for some basis $\{|i\rangle_B\}_{i=1}^{|B|}$ of $\mathcal{H}_B$.

A *product state* on $AB$ is a state of the form $\rho_A \otimes \sigma_B$. The state is called *separable* if it lies in the convex hull of product states:

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \sigma_B^i \tag{5.11}$$

for some states $\{\rho_A^i\}_i$ and $\{\sigma_B^i\}_i$ and probability distribution $\{p_i\}_i$. A state is called *entangled*, if it is not separable. An entangled state of particular interest is the *maximally entangled state*. Let $d = \dim \mathcal{H}$, $\{|i\rangle\}_{i=1}^d$ be a basis for $\mathcal{H}$. A maximally entangled state is expressed as

$$|\phi^+\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^d |i\rangle \otimes |i\rangle \quad \in \mathcal{H} \otimes \mathcal{H} \tag{5.12}$$

## 5.4 Measurements

The most general measurement is given by a *positive operator-valued measure* (POVM) $E = \{E_i\}_i$ where $E_i \geq 0 \quad \forall i$ and $\sum_i E_i = \mathbb{I}$. Then, for a quantum system $\mathcal{H}$ in state $\rho$, the probability of obtaining measurement outcome $i$ is given by $p_i = \mathrm{tr}[\rho E_i]$. So, we have

$$\sum_i p_i = \sum_i \mathrm{tr}[\rho E_i] = \mathrm{tr}\left[\rho \sum_i E_i\right] = \mathrm{tr}[\rho \mathbb{I}] = \mathrm{tr}\rho = 1, \tag{5.13}$$

for all normalized quantum states. A *projective measurement* $\Pi = \{\Pi_i\}$ is a POVM with the added property of orthogonality, which for projectors means

$$\Pi_i \Pi_j = \delta_{ij} \Pi_i. \tag{5.14}$$

Any basis $\{|e_i\rangle\}_{i=1}^{\dim \mathcal{H}}$ gives rise to a projective measurement $\Pi = \{|e_i\rangle\langle e_i|\}_{i=1}^{\dim \mathcal{H}}$.

## 5.5 Entropies

The *Shannon entropy* $H(p)$ of a probability distribution $p = \{p_i, \ldots, p_d\}$ is defined as $H(p) = -\sum_{i=1}^{d} p_i \log p_i$, where the logarithm is base 2 unless otherwise specified. Note that when the logarithm is base 2, the entropy has units of *bits*. The *von Neumann entropy* $S(\rho)$ of a quantum state $\rho$ is defined as

$$S(\rho) = -\mathrm{tr}\left[\rho \log \rho\right] = H(\{\lambda_i, \ldots, \lambda_d\}), \tag{5.15}$$

where $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ is a spectral decomposition of $\rho$ and where the logarithm of an operator is obtained by first diagonalizing the matrix representing the operator and then taking the logarithm of the diagonal elements. That is,

$$\log \rho = \sum_{i:\lambda_i > 0} \log(\lambda_i) |\psi_i\rangle\langle\psi_i|. \tag{5.16}$$

# Chapter 6

# Classes of Quantum Channels

## 6.1 Qubit channels

Note: please see Nielsen and Chuang for figures depicting these channels' actions on the Bloch sphere. I may make nice ones at some point but for now, I am too lazy.

1. **Bit flip channel.** First, we look at the classical bit flip channel, the properties of which were established by Claude Shannon in his seminal 1948 paper. The quantum version is simply given by the Pauli $X$ operator,

   $$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{6.1}$$

   which acts as

   $$X\left|0\right\rangle = \left|1\right\rangle, \tag{6.2}$$
   $$X\left|1\right\rangle = \left|0\right\rangle. \tag{6.3}$$

   It's eigenbasis is $\left|\pm\right\rangle := \frac{1}{\sqrt{2}}(\left|0\right\rangle \pm \left|1\right\rangle)$. So, $X\left|+\right\rangle\left|+\right\rangle$ and $X\left|-\right\rangle = -\left|-\right\rangle$. The channel acts as

   $$\mathcal{F}_p^X : \rho \mapsto (1-p)\rho + pX\rho X \tag{6.4}$$

   So, the Kraus operators are $\sqrt{1-p}\mathbb{I}$, $\sqrt{p}X$

2. **Phase-flip/Z-dephasing channel**

   $$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{6.5}$$
   $$Z\left|0\right\rangle = \left|0\right\rangle \tag{6.6}$$
   $$Z\left|1\right\rangle = -\left|1\right\rangle \tag{6.7}$$
   $$Z\left|+\right\rangle = \left|-\right\rangle \tag{6.8}$$
   $$Z\left|-\right\rangle = \left|+\right\rangle \tag{6.9}$$

   The channel is given as

   $$\mathcal{F}_p^Z : \rho \mapsto (1-p)\rho + pZ\rho Z \tag{6.10}$$

3. **Bit-phase flip/Y-dephasing channel**

$$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{6.11}$$

$$|\pm i\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i\,|1\rangle) \tag{6.12}$$

$$Y\,|i\rangle = |i\rangle \tag{6.13}$$

$$Y\,|-i\rangle = -\,|-i\rangle \tag{6.14}$$

$$Y\,|0\rangle = i\,|1\rangle \tag{6.15}$$

$$Y\,|1\rangle = -i\,|0\rangle \tag{6.16}$$

The channel is then

$$\mathcal{F}_p^Y : \rho \mapsto (1-p)\rho + pY\rho Y \tag{6.17}$$

$$= (1-p)\rho + pXZ\rho ZX \tag{6.18}$$

4. **Depolarizing channel.** Corresponds to an error model in which each Pauli error occurs with equal probability. The channel is

$$\mathcal{D}_p : \rho \mapsto (1-p)\rho + \frac{p}{3}\left(X\rho X + Y\rho Y + Z\rho Z\right) \tag{6.19}$$

with equal probability $p/3$. The Kraus operators are then $\sqrt{1-p}\mathbb{I}$, $\sqrt{p/3}X$, $\sqrt{p/3}Y$, $\sqrt{p/3}Z$. Thus, the Kraus rank is 4 for this channel when $p \in (0,1)$. An alternative representation is given by

$$\rho \mapsto (1-q)\rho + q\mathrm{tr}(\rho)\frac{\mathbb{I}}{2} \tag{6.20}$$

This represents replacing the input state with the maximally mixed state with "probability" $q$ ($q$ can be greater that 1).

What about the relation $p \leftrightarrow q$? To deduce this, we can use the identity

$$\frac{1}{2}\mathbb{I}_2 = \frac{1}{4}\left(\rho + X\rho X + Y\rho Y + Z\rho Z\right) \forall \rho \in B(\mathbb{C}^2) \tag{6.21}$$

We can then derive the relationship between $p$ and $q$. We have

$$\frac{1}{2}\mathbb{I}_2 = \frac{1}{4}\left(\rho + X\rho X + Y\rho Y + Z\rho Z\right) \quad \forall \rho \in B(\mathbb{C}^2) \tag{6.22}$$

$$(1-p)\rho + p\frac{\mathbb{I}}{2} = (1-q)\rho + \frac{q}{4}\left(\rho + X\rho X + Y\rho Y + Z\rho Z\right) \tag{6.23}$$

$$\implies \boxed{q = \frac{4}{3}p} \tag{6.24}$$

5. **Generalized Pauli channel.** Let $\vec{p} = (p_0, p_1, p_2, p_3)$ be a probability distribution. The generalized Pauli channel is

$$\mathcal{N}_{\vec{p}}(\rho) = p_0\rho + p_1 X\rho X + p_2 Y\rho Y + p_3 Z\rho Z \tag{6.25}$$

where we recover the depolarizing channel by setting $p_0 = 1-p$ and $p_i = p/3$. Note that for any $\vec{p}$, this channel is unital.

Pauli channels are interesting from an information theoretic standpoint. Classically, they are very easily understood. Quantum mechanically, they very much are not understood (except in the case of flip or dephasing channels).

In what sense are these flip channels de-phasing? Let us look at the example of the phase flip channel

$$\mathcal{F}_p^Z : \rho \mapsto (1-p)\rho + pZ\rho\rho Z. \tag{6.26}$$

We can understand the justification of the term de-phasing by looking at the action on the density matrix

$$\begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} \mapsto \begin{pmatrix} \rho_{11} & (1-2p)\rho_{12} \\ (1-2p)\rho_{21} & \rho_{22} \end{pmatrix} \tag{6.27}$$

where $\rho \geq 0, \mathrm{tr}(\rho) = 1 \implies \rho_{11} + \rho_{22} = 1$. We can make the following observations

1. If $\rho = x\,|0\rangle\langle 0| + (1-x)\,|1\rangle\langle 1|$, then

$$\mathcal{F}_p^Z(\rho) = \rho \quad \forall p \in [0,1] \tag{6.28}$$

2. $p = \frac{1}{2}$: the channel is diagonal in the z-basis for all input states

We can also think about sending classical information through this channel. Let us encode 0 as $|0\rangle\langle 0|$ and 1 as $|1\rangle\langle 1|$. That is, we are encoding one bit in one qubit. We can do this reliably with this channel because

$$\mathcal{F}_p^Z(|0\rangle\langle 0|) = |0\rangle\langle 0|, \tag{6.29}$$

$$\mathcal{F}_p^Z(|1\rangle\langle 1|) = |1\rangle\langle 1|. \tag{6.30}$$

This means that we can send one bit through the channel. The key idea is: this channel preserves classical information because it preserves the orthogonality of the basis states. The same would hold for the $X$ channel but you would have to encode info in the $|\pm\rangle$ basis.

6. **Amplitude damping channel.** Physical model: 2-level system (e.g. an atom with a ground state $|0\rangle$ and excited state $|1\rangle$. If the system is in an excited state $|1\rangle$, it decays with a certain probability, $\gamma$, emitting a photon to the environment. How can we capture this decay process mathematically? Consider the isometry

$$|0\rangle_A \mapsto |0\rangle_B\,|0\rangle_E\,, \tag{6.31}$$

$$|1\rangle \mapsto \sqrt{1-\gamma}\,|1\rangle_B\,|0\rangle_E + \sqrt{\gamma}\,|0\rangle_B\,|1\rangle_E\,. \tag{6.32}$$

Compactly we can write

The Kraus operators for this channel are

$$K_0 = \langle 0|_E\, V = |0\rangle\langle 0|_B + \sqrt{1-\gamma}\,|1\rangle\langle 1|)_B = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \tag{6.33}$$

$$K_1 = \langle 1|_E\, V = \sqrt{\gamma}\,|0\rangle_B\,\langle 1|_A = \begin{pmatrix} 0 & \gamma \\ 0 & 0 \end{pmatrix}. \tag{6.34}$$

In the Kraus representation, then, we can write the amplitude damping channel

$$\mathcal{A}_\gamma : \rho \mapsto K_0\rho K_0^\dagger + K_1\rho K_1^\dagger. \tag{6.35}$$

We note that the amplitude damping channel is not unital. Therefore, it is not a Pauli (or even a mixed-unitary) channel. Interestingly, it is completely understood how to send *quantum information* through the amplitude damping channel; however, it is not known how much classical information can be sent through this channel.

7. **Erasure channel.** Classical model:

   Quantum version: $\mathcal{E}_p : B(\mathcal{H}) \to B(\mathcal{H} \otimes \mathbb{C})$. The channel acts as

   $$\rho \mapsto (1-p)\rho + p\mathrm{tr}(\rho) \left|e\right\rangle \left\langle e\right| . \tag{6.36}$$

   This channel has Kraus operators given by $K_0\sqrt{1-p}\mathbb{I}, K_1\sqrt{p}\left|0\right\rangle\left\langle 0\right|, K_2\sqrt{p}\left|e\right\rangle\left\langle 1\right|$. Note that $\left|e\right\rangle$ must be orthogonal to all $\rho \in B(\mathcal{H})$. That is, $\langle \left|e\right\rangle\left\langle e\right|, \rho\rangle = 0 \quad \forall \rho \in B(\mathcal{H})$. Bob can always tell whether erasure happened by performing a measurement! This is a very well-understood quantum channel.

## 6.2 Generalized dephasing channels

A generalized dephasing channel leaves a *fixed* orthonormal basis invariant and it dephases off-diagonal elements with respect to that fixed basis. That is, we lose quantum coherences under the action of a generalized dephasing channel.

Construction: $\mathcal{H} = \mathbb{C}^d$ with orthonormal basis $\{\left|i\right\rangle\}_{i=1}^d$. Choose an environment $\mathcal{H}_E$ with $\mathrm{dim}\mathcal{H}_E := |E| \geq 2$, and let $\{\varphi_i\}_{i=1}^{|E|}$ be *some* set of pure states on $E$. This set is normalized but not orthogonal. Then, define the isometry

$$V : \left|i\right\rangle_A \mapsto \left|i\right\rangle_B \otimes \left|\varphi_i\right\rangle_E . \tag{6.37}$$

The Stinespring extension is

$$\begin{aligned} \mathcal{N}(\rho_A) &= \mathrm{tr}_E V \rho_A V^\dagger, & (6.38)\\ &= \sum_{i,j} \left|i\right\rangle \rho \left\langle i\right| \left|i\right\rangle \left\langle j\right|_B \mathrm{tr}\left(\left|\varphi_i\right\rangle\left\langle\varphi_j\right|\right), & (6.39)\\ &= \sum_{i,j} \langle\varphi_i|\varphi_j\rangle \langle i|\rho|j\rangle \left|i\right\rangle\left\langle i\right|_B . & (6.40) \end{aligned}$$

Let us confirm this channel acts as expect.

$$\begin{aligned} [\mathcal{N}(\rho)]_{kk} &= \langle\varphi_k|\varphi_k\rangle \left\langle k\right| \rho \left|k\right\rangle = \rho_{kk}, & (6.41)\\ [\mathcal{N}(\rho)]_{jk} &= \langle\varphi_j|\varphi_k\rangle \left\langle j\right| \rho \left|k\right\rangle , & (6.42) \end{aligned}$$

but $0 < \langle\varphi_j|\varphi_k\rangle \leq 1$. Thus, we see that the diagonals are preserved but the off-diagonals are dephased.

Higher-dimensional example: Let $d \geq 2$, and define two unitaries

$$X\left|i\right\rangle = \left|i+1 \bmod d\right\rangle \qquad \text{(shift operator)} \tag{6.43}$$

$$Z\left|j\right\rangle = \omega^j\left|j\right\rangle, \text{ where } \quad \omega = \exp\left(\frac{2\pi i}{d}\right) \qquad \text{(clock operator)} \tag{6.44}$$

This generalizes the Pauli operators: $X^d = Z^d = \mathbb{I}$. These are the generators of the Heisenberg-Weyl group:

$$\{\omega^j Z^k X^l : j, k, l \in [d]\}. \tag{6.45}$$

So, for $\rho \in B(\mathbb{C}^d)$ we maps like

$$\rho \mapsto (1-p)\rho + pX\rho X^\dagger, \tag{6.46}$$

$$\rho \mapsto (1-p)\rho + \frac{p}{3}Z\rho Z^\dagger + \frac{2p}{3}Z^2\rho(Z^\dagger)^2. \tag{6.47}$$

The generalized de-phasing channels have full classical capacity

$$C(\mathcal{N}) = \log d \quad (\text{ in general, } C(\mathcal{N}) \leq \log d) \tag{6.48}$$

Sketch of a proof: we know that for a generalized dephasing channel, there exists an orthonormal basis $\{|i\rangle\}_i$ such that

$$\mathcal{N}(|i\rangle\langle i|) = |i\rangle\langle i| \tag{6.49}$$

of classical signals/messages $x_i, \dots, x_d$. We can use the encoding

$$x_i \mapsto |i\rangle\langle i| \tag{6.50}$$

where $\langle i|j\rangle = \delta_{ij}$. Thus, $d$ messages can be sent perfectly. So $\log d$ bits of classical info can be sent through $N$. Bob can measure the output to retrieve the classical message.

### 6.2.1 Detour: Holevo information ($\mathcal{X}$-quantity)

Recall: the von Neumann entropy is defined as

$$S(\rho) = -\mathrm{tr}\rho \log \rho. \tag{6.51}$$

When one has the spectral decomposition of $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$, the von Neumann entropy is just the Shannon entropy for the eigenvalue distribution

$$S(\rho) = -\sum_i \lambda_i \log \lambda_i. \tag{6.52}$$

If we then have an ensemble of quantum states $E = \{p_x, \rho_x\}$, we have

$$\mathcal{X}(E, \mathcal{N}) = S(\sum_x p_x \mathcal{N}(\rho_x)) - \sum_x p_x S(\mathcal{N}(\rho_x)) \tag{6.53}$$

The Holevo information is then defined as

$$\boxed{\text{Holevo Information, } \mathcal{X}(\mathcal{N}) = \max_E \mathcal{X}(E, \mathcal{N})} \tag{6.54}$$

A fundamental result due to Holevo, Schumacher, and Westmoreland is

$$\boxed{C(\mathcal{N}) \geq \mathcal{X}(\mathcal{N})} \tag{6.55}$$

If $\mathcal{N}$ is a generalized dephasing channel, there exists an orthonormal basis $\{|i\rangle\}$ such that $\mathcal{N}(|i\rangle\langle i|) = |i\rangle\langle i|$ for all $i$. Let our ensemble be

$$\rho_i = |i\rangle\langle i| \tag{6.56}$$

$$p_i = \frac{1}{d}, \tag{6.57}$$

which implies

$$\bar{\rho} = \sum_i p_i \rho_i = \frac{1}{d}\mathbb{I} \tag{6.58}$$

Then, $E_n = \{\frac{1}{d}, \rho_i\}$, so the Holevo quantity is

$$\begin{aligned} \mathcal{X}(E_n, \mathcal{N}) &= S(\sum_i p_i \mathcal{N}(\rho_i)) - \sum_i p_i S(\mathcal{N}(\rho_i)) && (6.59)\\ &= S(\sum_i p_i \rho_i) - \sum_i p_i S(\rho_i) && (6.60)\\ &= \log d - 0 && (6.61)\\ &= \log d && (6.62) \end{aligned}$$

Thus, we have the following chain of inequalities

$$\log d \leq \mathcal{X}(\mathcal{N}) \leq C(\mathcal{N}) \leq \log d \implies \boxed{C(\mathcal{N}) = \log d} \tag{6.63}$$

as desired.

### 6.2.2 Some comments on last lecture

- The rank of the Kraus operators is not unitarily invariant. We can see this by examining the 50-50 dephasing channel.

$$\rho \mapsto \frac{1}{2}\rho + \frac{1}{2}Z\rho Z \implies K_0 = \frac{1}{\sqrt{2}}\mathbb{I}, \quad K_1 = \frac{1}{\sqrt{2}}Z \tag{6.64}$$

  We can apply a Hadamard transform to the operators to obtain new Kraus operators

$$\begin{aligned} L_i &= \sum_j H_{ij} K_j, && (6.65)\\ \implies L_0 &= |0\rangle\langle 0|, L_1 = |1\rangle\langle 1|, && (6.66) \end{aligned}$$

  which are rank-1 operators now when the previous Kraus operators were full rank.

- Every unital qubit channel is unitarily equivalent to a Pauli channel. That is, if $\mathcal{N} : B(\mathbb{C}^2) \to B(\mathbb{C}^2)$ is a unital channel, then there are unitaries $U, V$ such that

$$M(\rho) = U\mathcal{N}(V\rho V^\dagger)U^\dagger \tag{6.67}$$

  is Pauli.

- mixed unitary channels

$$\rho \mapsto \sum_i p_i U_i \rho U_i^\dagger, \quad U_i \text{ unitary} \tag{6.68}$$

$$\mathbb{I} \mapsto \sum_i p_i U_i U_i^\dagger = \mathbb{I} \quad \text{ unital!} \tag{6.69}$$

Every unital qubit channel is mixed unitary. Watrous' book provides a nice example of a unital channel that is *not* mixed unitary:

$$X \in B(\mathbb{C}^3), \quad \rho \mapsto \frac{1}{2}\mathrm{tr}(X)\mathbb{I} - \frac{1}{2}X^T \tag{6.70}$$

## 6.3 Entanglement-breaking channels

Reminder: A bipartite $\rho_{AB}$ is called separable if it lies in the convex hull of product states. That is,

$$\rho_{AB} \in \mathrm{conv}\{\omega_A \otimes \sigma_B : \omega_{A(B)} \text{ state on } \mathcal{H}_{A(B)}\} \tag{6.71}$$

Explicitly,

$$\rho_{AB} = \sum_i p_i \omega_A^i \otimes \sigma_B^i \tag{6.72}$$

**Definition 6.3.1.** *A channel $\mathcal{N} : A \to B$ is entanglement-breaking if*

$$(\mathbb{I}_R \otimes \mathcal{N})(\rho_{RA}) \tag{6.73}$$

*is separable for any $\rho_{RA}$.*

**Proposition 6.3.1.** *The following are all equivalent*

- *$\mathcal{N} : A \to B$ is entanglement breaking*
- *$\tau_{AB}^{\mathcal{N}} = (\mathbb{I}_A \otimes \mathcal{N})(\gamma_{AA'})$ is separable*
- *$\mathcal{N}$ has a Kraus representation with rank-1 Kraus operators*
- *$\mathcal{N}$ is a measure-and-prepare channel: there exists POVM $E = \{E_i\}_i$ and states $\{\sigma_i\}_i$ such that*

$$\mathcal{N}(\rho) = \sum_i tr(\rho E_i)\sigma_i \tag{6.74}$$

*where we recall that a POVM must satisfy $E_i \geq 0, \sum_i E_i = \mathbb{I}$.*

*Proof.* (a $\implies$ b): $(\mathbb{I}_R \otimes \mathcal{N})(\rho_{RA})$ is separable for all $\rho_{RA}$. In particular, for $\gamma_{RA}$, $\quad (|\gamma\rangle_{RA}) = \sum_i |i\rangle_R |i\rangle_A)$.

(b $\implies$ c): $\tau^{\mathcal{N}}_{AB}$ is separable: there exist pure states $\psi_i = |\psi_i\rangle\langle\psi_i|_A$ and $\varphi_i = |\varphi_i\rangle\langle\varphi_i|_B$ such that $\frac{1}{d}\tau^{\mathcal{N}}_{AB} = \sum_i p_i \psi_i \otimes \varphi_i$, where $d = |A|$. Then, set $K_i = \sqrt{p_i d}\,|\varphi_i\rangle_B \langle\bar{\psi}_i|_A$. We can then check the action of these operators

$$\begin{aligned}
\frac{1}{d}\sum_i (\mathbb{I}\otimes K_i)(\gamma_{AA'})(\mathbb{I}_A\otimes K_i)^\dagger &= \frac{1}{d}\sum_{i,j,k} |j\rangle\langle k|_A \otimes K_i\,|j\rangle\langle k|_{A'}\,K_i^\dagger \qquad (6.75)\\
&= \sum_{i,j,k} dp_i\,|j\rangle\langle k|_A \otimes \langle\bar{\psi}_i|\,|j\rangle\langle k|\,|\bar{\psi}_i\rangle\,|\varphi_i\rangle\langle\varphi_i|_B \qquad (6.76)\\
&= \sum_{i,j,k} p_i\,|j\rangle\langle k|_A \otimes \langle\bar{\psi}_i|\,|j\rangle\langle k|\,|\bar{\psi}_i\rangle\,|\varphi_i\rangle\langle\varphi_i|_B \qquad (6.77)\\
&= \sum_i p_i \left(\sum_{j,k} \langle k|\bar{\psi}_i|j\rangle\,|j\rangle\langle k|\right) \otimes |\varphi_i\rangle\langle\varphi_i|_B \qquad (6.78)\\
&= \sum_i p_i(\psi_i)\otimes\varphi_i \qquad (6.79)\\
&= \frac{1}{d}\tau^{\mathcal{N}}_{AB} \qquad (6.80)
\end{aligned}$$

$$\begin{aligned}
\sum_i K_i^\dagger K_i &= \mathbb{I} \qquad (6.81)\\
&= d\sum_i p_i\,|\bar{\psi}_i\rangle\langle\varphi_i|\varphi_i\rangle\langle\bar{\psi}_i| \qquad (6.82)\\
&= d\sum_i p_i\bar{\psi}_i \qquad (6.83)\\
&= d(\sum_i p_i\bar{\psi}_i) \qquad (6.84)\\
&= d\frac{1}{d}\bar{\mathbb{I}}_A \qquad (6.85)\\
&= \mathbb{I}_A \qquad (6.86)
\end{aligned}$$

(c $\implies$ d): rank($K_i$)=1: $K_i = |\mathcal{X}_i\rangle_B\langle\omega_i|_A$ for some vectors $|\mathcal{X}_i\rangle_B, |\omega_i\rangle_A, \langle\mathcal{X}_i|\mathcal{X}_i\rangle =$ 1.

$$\begin{aligned}
\mathcal{N}(\rho) &= \sum_i K_i\rho K_i^\dagger \qquad (6.87)\\
&= \sum_i \langle\omega_i|\rho|\omega_i\rangle\,|\mathcal{X}_i\rangle\langle\mathcal{X}_i|_B \qquad (6.88)
\end{aligned}$$

POVM: $\omega = \{\omega_i\}$, states: $\mathcal{X}_i$

$$\begin{aligned}
\mathbb{I} &= \sum_i K_i^\dagger K_i \qquad (6.89)\\
&= \sum_i |\omega_i\rangle\langle\mathcal{X}_i|\mathcal{X}_i\rangle\langle\omega_i| \qquad (6.90)\\
&= \sum_i |\omega_i\rangle\langle\omega_i| \qquad (6.91)
\end{aligned}$$

(d $\implies$ a): Let $\rho_{RA}$ be arbitrary:

$$(\mathbb{I}_R \otimes \mathcal{N})(\rho_{RA}) = \sum_i \mathrm{tr}_B \left[(\mathbb{I}_A \otimes E_i)\rho_{RA}\right] \otimes \sigma_i \tag{6.92}$$

$$= \sum_i \mathrm{tr}_B \left[(\mathbb{I}_R \otimes \sqrt{E_i})\rho_{RA}(\mathbb{I}_R \otimes \sqrt{E_i})\right] \otimes \sigma_i \tag{6.93}$$

$$= \sum_i p_i \omega_i \otimes \sigma_i \tag{6.94}$$

where $p_i = \mathrm{tr}(E_i \rho_{RA}), \omega_i = \frac{1}{p_i}\mathrm{tr}_B \left[(\mathbb{I}_R \otimes \sqrt{E_i})\rho_{RA}(\mathbb{I}_A \otimes \sqrt{E_i})\right] \geq 0.$ □

We note that there are *some* channel capacities of entanglement-breaking channels that are understood.

- Quantum information transmission is equivalent to generating entanglement. Because entanglement-breaking channels cannot do that, their quantum capacity is zero.
- $C(\mathcal{N}) \geq \mathcal{N}, C(\mathcal{N}) = \sup_{n\in\mathbb{N}} \frac{1}{n}\mathcal{X}(\mathcal{N}^{\otimes n})$. Entanglement-breaking channels destroy entanglement between different inputs. This implies that

  $$C(\mathcal{N}) = \mathcal{X}(\mathcal{N}) \tag{6.95}$$

  BUT: $\mathcal{X}(\mathcal{N})$ is NP-hard to compute, so this relationship doesn't actually get us much.

## 6.4 PPT-channels

Checking separability is NP-hard. So, is there some easier criterion? The standard relaxed criterion is following.

**Definition 6.4.1.** *Peres-Horodecki criterion: if $\rho_{AB}$ is separable if $\rho_{AB}$ has a positive partial transpose with respect to either party.*

Let us see why this is a reasonable criterion for separability. If $\rho_{AB}$ is separable,

$$\rho_{AB} = \sum_i p_i \omega_A^i \otimes \sigma_B^i \tag{6.96}$$

$$\implies \rho_{AB}^{T_B} = \sum_i p_i \omega_A^i \otimes (\sigma_B^i)^T \geq 0 \tag{6.97}$$

- if $\rho_{AB}$ is NPT $\implies \rho_{AB} \notin$SEP.
- if $|A| \cdot |B| \leq 6$, then $\rho_{AB} \in$ SEP $\Leftrightarrow \rho_{AB} \in$ PPT.

**Definition 6.4.2.** *A channel $\mathcal{N} : A \to B$ is called PPT if $(\mathcal{I}_R \otimes \mathcal{N})(\rho_{RA})$ is PPT for all $\rho_{RA}$.*

**Proposition 6.4.1.** *The following are all equivalent*

- *$\mathcal{N} : A \to B$ is PPT*
- *$\tau_{AB}^{\mathcal{N}}$ is PPT*
- *$\vartheta \circ \mathcal{N}$ is CP (Recall that when $\vartheta : X \mapsto X^T : (\mathbb{I}_R \otimes \vartheta)(\gamma) = \mathbb{F} \ngeq 0$)*

*Proof.* (a ⇒ b): True by definition. We know

$$(\mathbb{I}_A \otimes \mathcal{N})(\rho_{AA'}) \tag{6.98}$$

is PPT, in particular for $\rho = \gamma_{AA'}$.
(b ⇒ c):

$$\begin{aligned}(\mathbb{I}_A \otimes \vartheta \circ \mathcal{N})(\gamma) &= (\mathbb{I} \otimes \vartheta)(\mathbb{I} \otimes \mathcal{N})(\gamma_{AA'}) &(6.99)\\ &= (\mathbb{I}_A \otimes \vartheta)(\tau^{\mathcal{N}}_{AB}) &(6.100)\\ &= (\tau^{\mathcal{N}}_{AB})^{T_B} \geq 0 &(6.101)\end{aligned}$$

(c ⇒ a) $\vartheta \circ \mathcal{N}$ is CP:

$$(\mathbb{I}_R \otimes \vartheta \circ \mathcal{N})(\rho_{RA}) \geq 0 \quad \forall \rho_{RA} \geq 0 \tag{6.102}$$

This implies $\mathcal{N}$ is PPT. □

**Remarks**: In general, $\vartheta \circ \mathcal{N}$ is *not* CP. However, for every CP map $\mathcal{N}$, the map $\vartheta \circ \mathcal{N} \circ \vartheta$ is CP (see exercises).

What about the capacities of PPT channels?

- Horodeckis: PPT states are undistillable. Entanglement distillation: given iid copies of a state $\rho_{AB}$, the goal is to convert these copies into a smaller number of maximally entangled states. If there is an entanglement distillation protocol (LOCC) such that

  1. error of the protocol tends to 0 as $n \Rightarrow \infty$
  2. rate c:= $\frac{1}{n} \log m_n \to c > 0$ as $n$ goes to infinity

  then $\rho_{AB}$ is *distillable.* We conclude that the quantum capacity for PPT channels is zero. This result even holds if two-way classical communication is allowed!

We note that the set of protocols allowing only one-way communication is a strict subset of the set of protocols allowing two-way communication. Even still, the Horodecki's showed that the quantum capacity of all PPT channels is zero in both cases. Also note that the classical capacity is generally unknown.

## 6.5 Anti-degradable channels

Let $\mathcal{N}$ be a quantum channel from $A$ to $B$ with an isometry $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ such that $\mathcal{N}(\rho) = \mathrm{tr}_E V \rho V^\dagger$. Recall that the completmentary channel is $\mathcal{N}^c(\rho) = \mathrm{tr}_B V \rho V^\dagger$.

**Definition 6.5.1.** *$\mathcal{N}$ as above is called antidegradable if there exists a channel $\mathcal{A} : E \to B$ such that*

$$\mathcal{N} = \mathcal{A} \circ \mathcal{N}^c \tag{6.103}$$

Intuition: Eve (environment) can locally obtain Bob's output via the channel $\mathcal{A}$. Antidegradable channels cannot transmit quantum information and thus have zero quantum capacity, $Q(\mathcal{N}) = 0$.

A "proof" of this is as follows. Assume that this channel has non-zero quantum capacity. This means that Alice can faithfully send qubits to Bob at a positive rate. But there is a protocol based on the channel $\mathcal{A}$ that lets Eve implement the same protocol that Alice and Bob use. This violates no-cloning and is thus prohibited.

**Examples of anti-degradable channels**:

1. erasure channel $\mathcal{E}_p : \rho \mapsto (1-p)\rho + p\text{tr}(\rho) |e\rangle \langle e|$ for $p \geq \frac{1}{2}$.

   Let $\mathcal{H}_1 = \mathbb{C}^2$ be the input space. Let the $\mathcal{H}_2 = \mathbb{C}$ be an erasure flag. Let $\mathcal{E}_p : B(\mathcal{H}_1) \to B(\mathcal{H}_1 \oplus \mathcal{H}_2)$. That is, an input state maps as

$$\rho \mapsto (1-p)\tilde{\rho} + p\text{tr}(\rho) |e\rangle \langle e| \tag{6.104}$$

   So, how do we embed our input state in the larger space? We use a pretty trivial embedding that is represented as:

$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} & 0 \\ \rho_{10} & \rho_{11} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{6.105}$$

   and the erasure flag is simply

$$|e\rangle \langle e| = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.106}$$

   Note that the complementary channel of the erasure channel has the same form but with the probabilities flipped:

$$\mathcal{E}_p^c : B(\mathcal{H}_1) \to B(\mathcal{H}_1 \oplus \mathcal{H}_2), \rho \mapsto p\tilde{\rho} + (1-p)\text{tr}(\rho) |e\rangle \langle e|. \tag{6.107}$$

   Now consider $p \geq \frac{1}{2}$ and define $q = \frac{2p-1}{p}$. the idea is that we erase $\tilde{\rho}$ with probability $q$ and do nothing with $|e\rangle \langle e|$.

$$\begin{aligned} \mathcal{E}_q(p\rho) &= p(1-q)\tilde{\rho} + pq\text{tr}(\rho) |e\rangle \langle e| & (6.108) \\ &= (1-p)\tilde{\rho} + (2p-1) |e\rangle \langle e| & (6.109) \\ &= (1-p) |e\rangle \langle e| & (6.110) \end{aligned}$$

   **almost certainly messed up p's and q's here.**

   We need to extend the action of $\mathcal{E}_q$ to $B(\mathcal{H}_1 \otimes \mathcal{H}_2)$. The solution is to define $\mathcal{A}$ in the Kraus representation. The Kraus operators are

$$K_0 = \sqrt{1-q}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix} \tag{6.111}$$

$$K_1 = \sqrt{q}\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 1 & 0 & 0\end{pmatrix} \tag{6.112}$$

$$K_2 = \sqrt{q}\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 1 & 0\end{pmatrix} \tag{6.113}$$

$$K_3 = \sqrt{q}\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix} \tag{6.114}$$

where the first three operators correspond to erasure on $B(\mathcal{H}_1 \oplus 0)$ and the last to doing nothing on $B(0 \oplus \mathcal{H}_2)$. This definition ensures that $\mathcal{E}_p = \mathcal{A} \circ \mathcal{E}_p^c$ for $p \geq \frac{1}{2}$. We have $B(\mathcal{H}_1 \oplus 0), B(0 \oplus \mathcal{H}_2) : \mathcal{A}_p = \mathcal{E}_p \oplus \mathbb{I}$.

For $\rho \in B(\mathcal{H}_1)$, $\sigma = (1-\lambda)\tilde{\rho} + \lambda \left|e\right\rangle \left\langle e\right|$. This has the block matrix:

$$\begin{pmatrix}(1-\lambda)\rho & 0\\ 0 & \lambda\end{pmatrix} \tag{6.115}$$

So

$$(\mathcal{E}_p \oplus \mathbb{I})(\sigma) = \begin{pmatrix}\mathcal{E}_p((1-\lambda)\rho) & 0\\ 0 & \mathbb{I}(\lambda)\end{pmatrix} \tag{6.116}$$

Then for $\omega \in B(\mathcal{H}_1 \oplus \mathcal{H}_2)$ we have

$$\omega = \begin{pmatrix}\omega_{\mathcal{H}_1} & *\\ * & \omega_{\mathcal{H}_2}\end{pmatrix} \tag{6.117}$$

which has off-diagonal elements. We can get rid of these by measuring with respect to $\mathcal{H}_1 \oplus \mathcal{H}_2 = \mathcal{H}$. Let $P_2 = \left|e\right\rangle\left\langle e\right|, P_1 = \mathbb{I} - \left|e\right\rangle\left\langle e\right|$. Then $P_1 \omega P_1 + P_2 \omega P_2$ is diagonal. Finally, $\mathcal{A}_M = P_1 \cdot P_1 + P_2 \cdot P_2$ which implies $\mathcal{A} - \mathcal{A}_p \circ \mathcal{A}_M$ where $\mathcal{A}_p = \mathcal{E}_p \oplus \mathbb{I}$.

2. amplitude damping channel $\mathcal{A}_\gamma$ for $\gamma \geq \frac{1}{2}$.
3. depolarizing channel $\mathcal{D}_p$ for $p \geq \frac{1}{4}$

**Definition 6.5.2.** *Let $\mathcal{N} : A \to B$ be a quantum channel and let $G$ be a group with unitary representations $U_g$ on $\mathcal{H}_A$ and $V_g$ on $\mathcal{H}_B$. Then $\mathcal{N}$ is called covariant with respect to $(G, U_g, V_g)$ if*

$$V_g \mathcal{N}(\cdot) V_g^\dagger = \mathcal{N}(U_g \cdot U_g^\dagger) \tag{6.118}$$

*for all $g \in G$.*

**Representation theory basics:**

- $(\varphi, V)$ is a representation of $G$: $\varphi : G \to GL(V), \varphi(gh) = \varphi(g)\varphi(h)$. Subspace $W \leq V$ is called $G$-invariant, if $\varphi(g)w \in W \quad \forall w \in W, \forall g \in G$.
- $\{0\}, V$ are always $G$-invariant
- $(\varphi, V)$ is called irreducible if $\{0\}, V$ are the only $G$-inv subspaces.
- Schur's lemma: $(\varphi, V), (\psi, W)$ representations of $G$. $G$-linear map $f : f \circ \varphi(g) = \psi(g) \circ f$. If $\varphi, \psi$ are irreducible representations, then either $V \ncong W$ and $f = 0$ or $V \cong W$ and $f = \lambda \mathbb{I}_{V \to W}$ for some $\lambda \in \mathbb{C}$.

**Some examples:**

- Pauli channels: $\rho \mapsto p_0\rho + p_1 X\rho X + p_2 Y\rho Y + p_3 Z\rho Z$

  Covariance group: Pauli group, $P = \{\pm 1, \pm i\} \bigcup \{\mathbb{I}, X, Y, Z\}$

  For $\theta_1, \vartheta_2 \in P$: $\vartheta_1\theta_2 \cdot \theta_2^\dagger \vartheta_1^\dagger = \vartheta_2\theta_1 \cdot \theta_1^\dagger \vartheta_2^\dagger$
- depolarizing channel: $\rho \mapsto (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$. The covariance group is $\mathcal{U}(2)$. This is easy to see in the "$q$-representation": $\rho \mapsto (1-q)\rho + q\mathrm{tr}\rho\frac{1}{2}\mathbb{I}$

  $$\mathcal{D}_q(U\rho U^\dagger) = (1-q)U\rho U^\dagger + q\mathrm{tr}(U\rho U^\dagger)\frac{1}{2}\mathbb{I} = U\mathcal{D}_p(\rho)U^\dagger \tag{6.119}$$

  In $d \geq$ dimensions: $\rho \mapsto (1-q)\rho + q\mathrm{tr}\rho\frac{1}{d}\mathbb{I}$. The covariance group is $\mathcal{U}(d)$.
- Amplitude damping channel $A_\gamma = \{K_0, K_1\}$ as normal. Covariance group: $\{\mathbb{I}, Z\} \cong \mathbb{Z}_2$.
- Erasure channel $\mathcal{E}_p(\rho) = (1-p)\rho + p\mathrm{tr}\rho\, |e\rangle\, \langle e|$. The covariance group is again $\mathcal{U}(2)$. Similarly, in $d \geq 2$ dimensions, the $d$-dimensional erasure channel with covariance group $\mathcal{U}(d)$.

**Proposition 6.5.1.** *A channel $\mathcal{N} : A \to B$ is $(G, U_g, V_g)$-covariant iff $\tau^{\mathcal{N}_{AB}} = (\bar{U}_g \otimes V_g)\tau_{AB}^{\mathcal{N}}(\bar{U}_g \otimes V_g)^\dagger$ for all $g \in G$.*

*Proof.* $(\Rightarrow)$: $V_g\mathcal{N}(\cdot)V_g^\dagger = \mathcal{N}(U_g \cdot U_g^\dagger)$ for all $g \in G$. This is equivalent to $\mathcal{N}(\cdot) = V_g^\dagger\mathcal{N}(U_g \cdot U_g^\dagger)V_g$ for all $g \in G$.

$$\tau_{AB}^{\mathcal{N}} = (\mathbb{I} \otimes \mathcal{N})(\gamma) \tag{6.120}$$
$$= (\mathbb{I}_1 \otimes V_g^\dagger)(\mathbb{I} \otimes \mathcal{N})\left((\mathbb{I}_1 \otimes U_g)\gamma(\mathbb{I} \otimes U_g^\dagger)\right)(\mathbb{I}_1 \otimes V_g) \tag{6.121}$$
$$= (\mathbb{I} \otimes V_g)(\mathbb{I} \otimes \mathcal{N})\left((U_g^T \otimes \mathbb{I}_2)\gamma(U_g^T \otimes \mathbb{I}_2)^\dagger\right)(\mathbb{I} \otimes V_g) \tag{6.122}$$
$$= (U_g^T \otimes V_g^\dagger)(\mathbb{I} \otimes \mathcal{N})(\gamma)(\bar{U}_g \otimes V_g) \tag{6.123}$$
$$\Longrightarrow \tau_{AB}^{\mathcal{N}} = (U_g^T \otimes V_g^\dagger)\tau_{AB}^{\mathcal{N}}(\bar{U}_g \otimes V_g) \quad \text{for all} \quad g \in G \tag{6.124}$$

$(\Leftarrow)$: $\tau_{AB}^{\mathcal{N}} = (\bar{U}_g \otimes V_g)\tau_{AB}^{\mathcal{N}}(\bar{U}_g \otimes V_g)^\dagger$ for all $g \in G$. This implies $(U_g^T \otimes \mathbb{I})\tau_{AB}^{\mathcal{N}}(U_g^T \otimes \mathbb{I})^\dagger = (\mathbb{I} \otimes V_g)\tau_{AB}^{\mathcal{N}}(\mathbb{I} \otimes V_g)^\dagger$.

Recall the choi isomorphism: $\mathcal{N}(X) = \mathrm{tr}_1 \tau^{\mathcal{N}}_{AB}(X^T \otimes \mathbb{I})$.

$$\begin{aligned}
\mathcal{N}(U_g X U_g^\dagger) &= \mathrm{tr}_1 \left[\tau^{\mathcal{N}}_{AB}((U_g X U_g^\dagger)^T \otimes \mathbb{I})\right] && (6.125)\\
&= \mathrm{tr}_1 \left[(U_g^T \otimes \mathbb{I})\tau_{AB}(U_g^T \otimes \mathbb{I})^\dagger (X^T \otimes \mathbb{I})\right] && (6.126)\\
&= \mathrm{tr}_1 \left[(\mathbb{I} \otimes V_g)\tau_{AB}(\mathbb{I} \otimes V_g)^\dagger (X^T \otimes \mathbb{I})\right] && (6.127)\\
&= V_g \mathrm{tr}_1 \left[\tau^{\mathcal{N}_{AB}}(X^T \otimes \mathbb{I})\right] V_g^\dagger && (6.128)\\
&= V_g \mathcal{N}(X) V_g^\dagger && (6.129)
\end{aligned}$$

□

**Problem:** with the above proposition is that it is basis-dependent (via the Choi operator).

**Solution:** is that there is also a basis independent version based on the Jamiolkowski operator $J^{\mathcal{N}}_{AB} = (\mathbb{I} \otimes \mathcal{N})(\mathbb{F}_{AA'})$ where $\mathbb{F}_{AB} = |\gamma\rangle \langle\gamma|^{T_A}$. In terms of the Jamialkowski operator, $(G, U_g, V_g)$-covariance of $\mathbb{N}$ is equivalent to $(U_g \otimes V_g) J^{\mathcal{N}}_{AB} (U_g \otimes V_g)^\dagger = J^{\mathcal{N}}_{AB}$ for all $g \in G$.

We note that a $d$-dimensional depolarizing channel; $\rho \mapsto (1-q)\rho + q\mathrm{tr}\rho\frac{1}{d}\mathbb{I}_d$ has $\mathcal{U}(d)$ as its covariance group.

**Proposition 6.5.2.** *Let $\mathcal{N} : A \to B$ be a channel with input and output spaces of equal dimension $d$. If $U\mathcal{N}(\cdot)U^\dagger = \mathcal{N}(U \cdot U^\dagger)$ for all $U \in mathcalU(d)$, then*

$$\mathcal{N} = (1-q) \cdot + q tr(\cdot)\frac{1}{d}\mathbb{I}_d \tag{6.130}$$

*with $q = (1-f)/(1-d)$ where $f = \langle\gamma| \tau^{\mathcal{N}} |\gamma\rangle$.*

# Part IV

# Quantum Information

# Chapter 7

# Estimation

## 7.1 DFE

# Chapter 8

# Sensing

## 8.1 Joint measurement of TFE via SFG

## 8.2 Introduction

## 8.3 Terminology

## 8.4 Problem formulation

## 8.5 Protocol

## 8.6 Performance evaluation

**Lemma 8.6.1.**

$$a = b \tag{8.1}$$

# Chapter 9

# Imaging

## 9.1 Quantum and non-local effects offer over 40 dB noise resilience advantage towards quanutm lidar