Felix Leditzky

Quantum Channels

THEORY AND APPLICATIONS (WORKING TITLE)

Compiled by Jacob L. Beckey & Louis Schatzki

COMPILED BY JACOB L. BECKEY & LOUIS SCHATZKI

TUFTE-LATEX.GOOGLECODE.COM

Last Update: October 2025

Contents

1
Representations

1.1 Unitary evolution of closed quantum systems

Quantum mechanics

We start by stating some broadly-accepted facts or axioms about quantum mechanics.

- quantum systems are represented using (in our case) finite dimensional Hilbert spaces, denoted $\mathcal{H}$.
- observables are measurable quantities represented by Hermitian operators $A \in B(\mathcal{H})$,[1] satisfying $A = A^\dagger$ and $A = \sum_i a_i |a_i\rangle \langle a_i|$, where the $|a_i\rangle$ are orthonormal eigenvectors ($\langle a_i|a_j\rangle = \delta_{ij}$) corresponding to real eigenvalues $a_i \in \mathbb{R}$.
- eigenvalues, $a_i \in \mathbb{R}$, are the possible measurement outcomes of the observable A.
- quantum states on $\mathcal{H}$ are linear ($\rho \in B(\mathcal{H})$), positive semi-definite ($\rho \geq 0$), unit trace ($\operatorname{tr}\rho = 1$) operators. They always have a spectral decomposition $\rho = \sum_i \lambda_i |\psi_i\rangle \langle \psi_i|$, where $\rho |\psi_i\rangle = \lambda_i |\psi_i\rangle$ and $\lambda \in [0, 1]$.
- pure quantum states have rank one[2], i.e. $\exists |\psi\rangle \in \mathcal{H}$ s.t. $\rho = |\psi\rangle \langle \psi|$
- given a state ρ and an observable A on the Hilbert space, $\mathcal{H}$, the probability of a specific outcome is given by $p_i = \operatorname{tr}(\rho |a_i\rangle \langle a_i|) = \langle a_i|\rho|a_i\rangle$.[3]
- expected outcomes of measurement: $\langle A \rangle_\rho = \sum_i p_i a_i = \operatorname{tr}(\rho A)$

[1] $B(\mathcal{H})$ is denotes the set of all *bounded* linear operators acting on the Hilbert space $\mathcal{H}$. Results would need to be modified or handled with care in the infinite-dimensional case. Throughout this course, we will assume $\dim \mathcal{H} < \infty$.

[2] One λ_i in the spectral decomposition is one, the rest are zero.

[3] The second equality follows from cyclicity of the trace.

Evolution of quantum systems

There are essentially three *pictures* of quantum evolution:

1. Schrödinger picture: states evolve

2. Heisenberg picture: observables evolve
3. Interaction picture: both evolve

We primarily focus on the Schrödinger picture in this course. Now, what do we require of a formalism that describes quantum evolution? Well, any evolution operation must be linear and should preserve the Hilbert space structure. Mathematically, we require the evolution map $T\colon B(\mathcal{H}) \to B(\mathcal{H})$ to satisfy

$$\operatorname{tr}(\psi\phi) = \operatorname{tr}(T(\psi)T(\phi)) \quad \forall \quad \psi, \phi \in B(\mathcal{H}) \tag{1.1}$$

The above requirements lead to the following theorem of fundamental importance in quantum mechanics.[4]

[4] Though outside the scope of the lecture, Wigner's Theorem is worth reading more about.

Wigner's Theorem: The above requirements imply that $T(X)$ must be either unitary or anti-unitary. That is,

$$T(X) = UXU^\dagger \quad \text{or} \quad UX^TU^\dagger \tag{1.2}$$

for some unitary U.

1.2 *Open systems and noisy evolution*

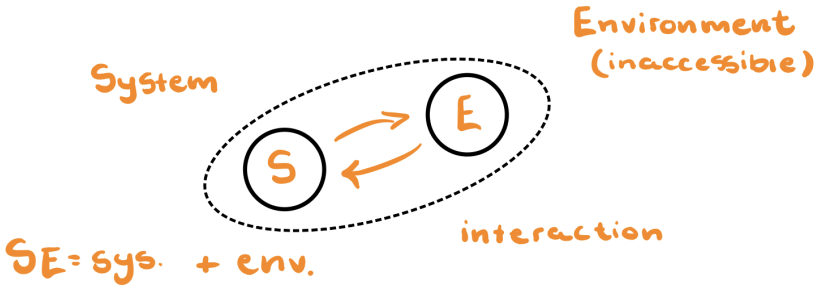

Figure 1.1: The closed system assumption is **not** realistic in practice! Sadly, no matter how hard we try to isolate them, systems interact with the environment.

Because interaction with the environment (some external, inaccessible system) is unavoidable, the closed system assumption is not realistic. But given a quantum system S of interest, we may assume the existence of an environment E that is sufficiently large so that the closed system assumption does apply. In this case, our system S together with the inaccessible environment E evolve unitarily as[5]

$$X_{SE} \mapsto UX_{SE}U^\dagger, \quad \text{where } U \in \mathcal{U}(\mathcal{H}_S \otimes \mathcal{H}_E). \tag{1.3}$$

[5] Note that $\mathcal{U}(\mathcal{H}_S \otimes \mathcal{H}_E)$ denotes the unitary group on the system and environment composite space.

We can then trace out the environment to recover the evolved system of interest via

$$X_S = \operatorname{tr}_E(UX_{SE}U^\dagger) \tag{1.4}$$

This partial trace over the environment generally leads to a *noisy and irreversible* evolution of S.

Now, recall that we are focusing on the Schrodinger picture of quantum mechanics. This means we evolve quantum states with maps $T : B(\mathcal{H}_1) \to B(\mathcal{H}_2)$. Maps that evolve quantum states must satisfy the following requirements:

1. Linearity
2. Map states to states, which involves:
   (a) T should preserve trace: $\mathrm{tr}(T(X)) = \mathrm{tr}\, X$
   (b) T should be positive: If $X \geq 0$ then also $T(X) \geq 0$. (short-hand notation: $T \geq 0$)
   (b′) T should be *completely* positive: $T \otimes \mathrm{id}_n \geq 0$ for all $n \in \mathbb{N}$. Here, id_n denotes the identity map on the space $B(\mathbb{C}^n)$.

The condition (b′) ensures that a quantum channel maps states to states also when they are only acting on part of a system. Obviously, (b′) implies the weaker criterion (b) upon choosing $n = 1$. At the end of this section we will see an example of a map that is positive but not completely positive.

The requirements stated above lead us to our first definition.[6]

Definition 1. A quantum channel is a linear, completely positive (CP), trace-preserving (TP) map $T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$.

Given a map $T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$, the *adjoint map* is $T^\dagger : B(\mathcal{H}_2) \to B(\mathcal{H}_1)$ defined via

$$\langle T^\dagger(X), Y\rangle = \langle X, T(Y)\rangle \tag{1.5}$$

for all $X \in B(\mathcal{H}_2)$ and $Y \in B(\mathcal{H}_1)$. Note that a map $T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ is:

- CP iff $T^\dagger$ is CP.
- TP iff $T^\dagger$ is unital, i.e., $T^\dagger(\mathbb{1}_2) = \mathbb{1}_1$:

$$\langle T^\dagger(\mathbb{1}_2), Y\rangle = \langle \mathbb{1}_2, T(Y)\rangle = \mathrm{tr}(T(Y)) = \mathrm{tr}\, Y = \langle \mathbb{1}_1, Y\rangle, \tag{1.6}$$

for all $Y \in B(\mathcal{H}_1)$. This chain of equalities implies that $T^\dagger(\mathbb{1}_2) = \mathbb{1}_1$. Note that unital quantum channels are both TP and unital.

[6] Quantum channels are often also referred to as CPTP maps or quantum operations.

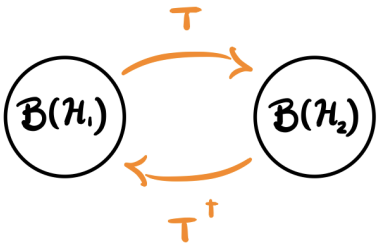

Figure 1.2: If a map T takes you from input space to output space, the adjoint map takes you from the output space to the input space such that Eq. (1.5) is satisfied.

1.3 *Choi–Jamiołkowski isomorphism*

We now define a very useful tool for studying quantum channels.

Definition 2. Let $T : B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ be a linear map. The Choi operator $\tau \in B(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is defined as

$$\tau := (\mathrm{id}_1 \otimes T)(\gamma) \tag{1.7}$$

where $\gamma := |\gamma\rangle\langle\gamma|$ and $|\gamma\rangle = \sum_{i=1}^{\dim \mathcal{H}_1} |i\rangle \otimes |i\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_1$. Note that $\{|i\rangle\}_{i=1}^{\dim \mathcal{H}_1}$ is an orthonormal basis for $\mathcal{H}_1$.

The explicit form of this operator is then

$$\tau = \sum_{i,j} |i\rangle\langle j| \otimes T(|i\rangle\langle j|) \tag{1.8}$$

Example. If $\mathcal{H}_1 = \mathcal{H}_2 = \mathbb{C}^2$, we can express the Choi operator as the block matrix

$$\tau = \begin{pmatrix} T(|0\rangle\langle 0|) & T(|1\rangle\langle 0|) \\ T(|0\rangle\langle 1|) & T(|1\rangle\langle 1|) \end{pmatrix} \tag{1.9}$$

where the elements of this matrix are themselves operators $T(|i\rangle\langle j|)$.

Proposition 3. *Let $T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$. The map*

$$T \mapsto \tau = (\mathrm{id}_1 \otimes T)(\gamma)$$

is a bijection between $\{T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)\}$ and $B(\mathcal{H}_1 \otimes \mathcal{H}_2)$ with inverse mapping

$$\tau \mapsto T(X) := \mathrm{tr}_1(\tau(X^T \otimes \mathbb{1}))),$$

where the transpose is taken with respect to the basis used to define the maximally entangled vector $|\gamma\rangle$.

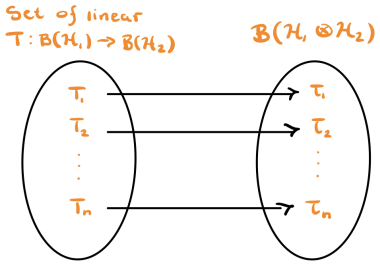

Figure 1.3: The Choi-Jamiołkowski isomorphism is a one-to-one mapping between the set of linear operators acting from $B(\mathcal{H}_1)$ to $B(\mathcal{H}_2)$ and the set of bounded operators on $\mathcal{H}_1 \otimes \mathcal{H}_2$, denoted $B(\mathcal{H}_1 \otimes \mathcal{H}_2)$.

Proof. Let $\tau = (\mathrm{id} \otimes T)(\gamma)$, where as usual $\gamma = |\gamma\rangle\langle\gamma|$. Tracing over

the first system, we have

$$\begin{aligned}
\mathrm{tr}_1(\tau(X^T \otimes \mathbb{1})) &= \mathrm{tr}_1\left(\left(\sum_{i,j} |i\rangle\langle j| \otimes T(|i\rangle\langle j|)\right)(X^T \otimes \mathbb{1})\right) && (1.10)\\
&= \sum_{i,j} \mathrm{tr}(|i\rangle\langle j| X^T) T(|i\rangle\langle j|) && (1.11)\\
&= \sum_{i,j}\left(\sum_k \langle k|(|i\rangle\langle j| X^T)|k\rangle T(|i\rangle\langle j|)\right) && (1.12)\\
&= \sum_{i,j} \langle j| X^T |i\rangle T(|i\rangle\langle j|) && (1.13)\\
&= \sum_{i,j} x_{ij} T(|i\rangle\langle j|), && (1.14)\\
&= T\left(\sum_{i,j} x_{ij} |i\rangle\langle j|\right), \quad \text{by linearity of } T && (1.15)\\
&= T(X), && (1.16)
\end{aligned}$$

which shows that $T \mapsto (\mathrm{id} \otimes T)(\gamma)$ is injective.

It remains to be shown that $T \mapsto (\mathrm{id} \otimes T)(\gamma)$ is surjective. We note that there exists $|\psi_i\rangle, |\phi_i\rangle$ such that $\tau = \sum_i |\psi_i\rangle\langle\phi_i| \in B(\mathcal{H}_1 \otimes \mathcal{H}_2)$ where $|\psi_i\rangle \neq |\phi_i\rangle$.

Claim: For every vector $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ there exists $V \in B(\mathcal{H}_1, \mathcal{H}_2)$ such that

$$|\psi\rangle = (\mathbb{1}_1 \otimes V)|\gamma\rangle. \tag{1.17}$$

To see this, let $|\psi\rangle = \sum_{i,j} p_{ij} |i\rangle \otimes |e_j\rangle$ where $\{|i\rangle\}$ is γ's basis and $\{e_j\}$ is an arbitrary basis. Then we can construct V as

$$V = \sum_{i,j} p_{ij} |e_j\rangle\langle i|, \tag{1.18}$$

which proves the claim.

We can write $\tau = \sum_i |\psi_i\rangle\langle\phi_i| \in B(\mathcal{H}_1 \otimes \mathcal{H}_2)$. The claim proved above shows that there exist L_i, K_i such that

$$\begin{aligned}
|\psi_i\rangle &= (\mathbb{1} \otimes K_i)|\gamma\rangle, && (1.19)\\
|\phi_i\rangle &= (\mathbb{1} \otimes L_i)|\gamma\rangle. && (1.20)
\end{aligned}$$

This implies

$$\begin{aligned}
\tau &= \sum_i |\psi_i\rangle\langle\phi_i| && (1.21)\\
&= \sum_i (\mathbb{1} \otimes K_i)|\gamma\rangle\langle\gamma|(\mathbb{1} \otimes L_i)^\dagger && (1.22)\\
&= (\mathrm{id} \otimes T)(\gamma), && (1.23)
\end{aligned}$$

where we have identified

$$T(X) = \sum_i K_i X L_i^\dagger \tag{1.24}$$

as the linear map we sought. This completes the proof of the proposition. □

Remark 4. The first part of the above proof shows that the mapping $T \mapsto (\mathbb{1} \otimes T)(\gamma)$ is injective. Noting that $\{T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ and $B(\mathcal{H}_1 \otimes \mathcal{H}_2)$ have the same dimension $d^2 d'^2$ (where $d = \dim \mathcal{H}_1$ and $d' = \dim \mathcal{H}_2$) then shows that it is also surjective, hence an isomorphism.

Proposition 5. *Let $T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ be a linear map with associated Choi operator $\tau = (\mathrm{id} \otimes T)(\gamma)$, where $|\gamma\rangle = \sum_i |i\rangle \otimes |i\rangle$. Then the following hold:*

1. *$T(X)^\dagger = T(X^\dagger)$ iff $\tau = \tau^\dagger$.*
2. *T is CP iff $\tau \geq 0$.*
3. *T is TP iff $\mathrm{tr}_2\, \tau = \mathbb{1}_1$.*
4. *T is unital iff $\mathrm{tr}_1\, \tau = \mathbb{1}_2$.*

Proof. 1. ($\Rightarrow$) First we prove if $T(X)^\dagger = T(X^\dagger)$, then $\tau = \tau^\dagger$.

$$\begin{aligned}
\tau^\dagger &= \left(\sum_{i,j} |i\rangle \langle j| \otimes T(|i\rangle \langle j|) \right)^\dagger && (1.25)\\
&= \sum_{i,j} |j\rangle \langle i| \otimes T(|i\rangle \langle j|)^\dagger && (1.26)\\
&= \sum_{i,j} |j\rangle \langle i| \otimes T(|i\rangle \langle j|^\dagger) && (1.27)\\
&= \sum_{i,j} |j\rangle \langle i| \otimes T(|j\rangle \langle i|) && (1.28)\\
&= \tau && (1.29)
\end{aligned}$$

($\Leftarrow$) The other way, if $\tau = \tau^\dagger$, then $T(X)^\dagger = T(X^\dagger)$. We know that $\tau = \sum_{i,j} |i\rangle \langle j| \otimes T(|i\rangle \langle j|)$ so $\tau = \tau^\dagger$ allows us to write

$$\begin{aligned}
\left(\sum_{ij} |i\rangle \langle j| \otimes T(|i\rangle \langle j|) \right) &= \left(\sum_{ij} |i\rangle \langle j| \otimes T(|i\rangle \langle j|) \right)^\dagger && (1.30)\\
&= \sum_{ij} |j\rangle \langle i| \otimes T(|i\rangle \langle j|)^\dagger. && (1.31)
\end{aligned}$$

This is a matrix equality, so each element must be the same. Sandwiching $\langle l| \otimes \mathbb{1}_2 \cdot |k\rangle \otimes \mathbb{1}_2$ around both sides of the expression above yields

$$T(|l\rangle \langle k|) = T(|k\rangle \langle l|)^\dagger \quad \forall \quad l, k, \tag{1.32}$$

which implies

$$T(|l\rangle \langle k|)^\dagger = T(|k\rangle \langle l|) \quad \forall \quad l, k. \tag{1.33}$$

Finally, due to the linearity of T and the fact that all bounded operators can be expanded in the form $X = \sum_{l,k} x_{lk} |l\rangle \langle k|$, we have

$$T(X)^\dagger = T(X^\dagger) \quad \forall \quad X \in B(\mathcal{H}_1), \tag{1.34}$$

as desired.

2. ($\Rightarrow$) First we prove that if T is CP, then $\tau \geq 0$. By definition of the Choi operator, if $\gamma \geq 0$, then $\tau \geq 0$.

   ($\Leftarrow$) Going the other way, we want to prove that if $\tau \geq 0$, then T is CP. To do so, we need to show

$$(\mathrm{id}_n \otimes T)(\rho) \geq 0 \quad \forall \rho \in B(\mathbb{C}^n \otimes \mathcal{H}_2), \rho \geq 0. \tag{1.35}$$

   We can write $\rho = \sum_i \lambda_i |\psi_i\rangle \langle\psi_i|$ with $\lambda_i \geq 0$. The above claim then follows from the fact that each state in the decomposition satisfies

$$(\mathrm{id}_n \otimes T)(\psi_i) \geq 0 \quad \forall i, \tag{1.36}$$

   where $\psi_i = |\psi_i\rangle \langle\psi_i|$ is defined via the steering inequality. That is, there exists $K_i \in B(\mathcal{H}_1, \mathbb{C}^n)$ such that

$$|\psi_i\rangle = (K_i \otimes \mathbb{1}) |\gamma\rangle . \tag{1.37}$$

   Using this, we can write

$$\begin{aligned}
(\mathrm{id}_n \otimes T)(\psi_i) &= (\mathrm{id}_n \otimes T) \left((K_i \otimes \mathbb{1}) |\gamma\rangle \langle\gamma| (K_i \otimes \mathbb{1})^\dagger \right), && (1.38)\\
&= (K_i \otimes \mathbb{1}) (\mathrm{id}_n \otimes T(|\gamma\rangle \langle\gamma|)) (K_i \otimes \mathbb{1})^\dagger, && (1.39)\\
&= (K_i \otimes \mathbb{1}) \tau (K_i \otimes \mathbb{1})^\dagger, && (1.40)\\
&\geq 0, && (1.41)
\end{aligned}$$

   because $\tau \geq 0$ and because $X \geq 0$ implies $CXC^\dagger \geq 0$ for all C[7]. As this holds for all ψ_i, it also holds for all $\rho = \sum_i \lambda_i \psi_i$. Thus, we have

$$(\mathrm{id}_n \otimes T)(\rho) \geq 0 \quad \forall \rho \in B(\mathbb{C}^n \otimes \mathcal{H}_2), \tag{1.42}$$

   as desired.

[7] Remember that $C \geq 0$ means $\langle x| C |x\rangle \geq 0 \forall x \in \mathcal{H}$.

3. ($\Rightarrow$) If T is TP, then $\mathrm{tr}_2 \tau = \mathbb{1}_1$.

$$\begin{aligned}
\mathrm{tr}_2 \tau &= \mathrm{tr}_2 \left(\sum_{i,j} |i\rangle \langle j| \otimes T(|i\rangle \langle j|) \right) && (1.43)\\
&= \sum_{i,j} |i\rangle \langle j| \, \mathrm{tr}(T(|i\rangle \langle j|)) && (1.44)\\
&= \sum_{i,j} |i\rangle \langle j| \, T(\mathrm{tr}(|i\rangle \langle j|)) && (1.45)\\
&= \sum_{i,j} |i\rangle \langle j| \, \delta_{ij} && (1.46)\\
&= \sum_{i} |i\rangle \langle i|_1 && (1.47)\\
&= \mathbb{1}_1 && (1.48)
\end{aligned}$$

($\Leftarrow$) If $\operatorname{tr}_2 \tau = \mathbb{1}_1$, then $\operatorname{tr} T(X) = \operatorname{tr} X$:

$$\operatorname{tr} T(X) = \operatorname{tr}\left[\operatorname{tr}_1(\tau(X^T \otimes \mathbb{1}))\right] \tag{1.49}$$
$$= \operatorname{tr}(\tau(X^T \otimes \mathbb{1})) \tag{1.50}$$
$$= \operatorname{tr}\left(\operatorname{tr}_2(\tau) X^T\right) \tag{1.51}$$
$$= \operatorname{tr} X^T \tag{1.52}$$
$$= \operatorname{tr} X, \tag{1.53}$$

where the first equality holds by Prop. 3 and the second by the definition of partial trace: $\operatorname{tr}(\operatorname{tr}_i(\cdot)) = \operatorname{tr}(\cdot)$.

4. ($\Rightarrow$) T is unital if $\operatorname{tr}_1 \tau = \mathbb{1}_2$.

$$\operatorname{tr}_1 \tau = \sum_{i,j} \operatorname{tr}(|i\rangle \langle j|_1) T(|i\rangle \langle j|_1) \tag{1.54}$$
$$= \sum_i T(|i\rangle \langle i|_1) \tag{1.55}$$
$$= T(\mathbb{1}_1) \tag{1.56}$$
$$= \mathbb{1}_2 \tag{1.57}$$

($\Leftarrow$) If $\operatorname{tr}_1 \tau = \mathbb{1}_2$, then T is unital:

$$T(\mathbb{1}_1) = \operatorname{tr}_1\left(\tau(\mathbb{1}_1^T \otimes \mathbb{1}_2)\right) \tag{1.58}$$
$$= \operatorname{tr}_1 \tau \tag{1.59}$$
$$= \mathbb{1}_2, \tag{1.60}$$

which finishes the proof. □

Examples of CP maps

1. Unitary maps of the form $T(X) = UXU^\dagger$ for any unitary $U \in \mathcal{U}(\mathcal{H})$ are CP and also TP:
   - Clearly, $\operatorname{tr} T(X) = \operatorname{tr}(UXU^\dagger) = \operatorname{tr}(XU^\dagger U) = \operatorname{tr}(X)$, so T is TP.
   - We have $T \otimes \mathrm{id}_n = (U \otimes \mathbb{1}_n) \cdot (U \otimes \mathbb{1}_n)^\dagger$, where $\mathbb{1}_n$ denotes the identity operator acting on $\mathbb{C}^n$. This is clearly a positive map for all $n \in \mathbb{N}$, and hence T is CP.
2. More generally, isometries are also CPTP.[8]
3. The trace $\operatorname{tr}\colon \mathcal{H} \to \mathbb{C}$ is CPTP, since
$$(\mathrm{id}_1 \otimes \operatorname{tr})(\gamma) = \sum_{i,j} |i\rangle \langle j| \operatorname{tr}(|i\rangle \langle j|) = \mathbb{1}_{\mathcal{H}_1} \geq 0.$$
Note that this also implies that the partial trace, $\operatorname{tr}_2\colon B(\mathcal{H}_1 \otimes \mathcal{H}_2) \to B(\mathcal{H}_1)$ is CPTP. We will later show that all channels can be expressed as unitary evolution on system + environment followed by a partial trace over the environment.

[8] A linear map $V\colon \mathcal{H}_1 \to \mathcal{H}_2$ is an isometry if $V^\dagger V = \mathbb{1}_{\mathcal{H}_1}$, or equivalently $\langle\varphi| V^\dagger V |\psi\rangle = \langle\varphi|\psi\rangle$ for all $|\varphi\rangle, |\psi\rangle \in \mathcal{H}_1$. Note that we necessarily have $\dim \mathcal{H}_1 \leq \dim \mathcal{H}_2$ whenever there exists an isometry $V\colon \mathcal{H}_1 \to \mathcal{H}_2$.

4. $X \mapsto \sum_i K_i X K_i^\dagger$ are CP (more on this later).

We have seen examples of CP maps. What about a map that is not CP? The transposition map, $\vartheta\colon X \mapsto X^T$ with respect to a fixed basis, is positive but not *completely positive*.

Consider the following operator:

$$\begin{aligned}\mathbb{F} &= (\mathrm{id}\otimes\vartheta)(\gamma) && (1.61)\\ &= \sum_{i,j} |i\rangle\langle j| \otimes \vartheta(|i\rangle\langle j|) && (1.62)\\ &= \sum_{i,j} |i\rangle\langle j| \otimes |j\rangle\langle i|\,. && (1.63)\end{aligned}$$

We call this the swap operator because for $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}_1$ we have $\mathbb{F}(|\psi_1\rangle \otimes |\psi_2\rangle) = |\psi_2\rangle \otimes |\psi_1\rangle$. Consider, then, the swap operator's action on the state $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$:

$$\mathbb{F}\,|\Psi^-\rangle = -\,|\Psi^-\rangle\,, \tag{1.64}$$

thus, $\mathbb{F} \not\geq 0$ which implies ϑ is not CP.

1.4 *Kraus representation and the isometric picture*

We saw before that $X \mapsto \sum_i K_i X K_i^\dagger$ are CP. The converse is also true.

Proposition 6.

1. *A map $T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ is CP iff there exist $\{K_i\}_i$ with $K_i \in B(\mathcal{H}_1, \mathcal{H}_2)$ such that $T(X) = \sum_i K_i X K_i^\dagger$. We call the K_i's the Kraus operators of T.*

2. *The Kraus rank, $r(T)$, the minimal number of Kraus operators needed to represent T, is equal to the rank of the Choi operator, $\tau = (\mathrm{id}\otimes T)(\gamma)$. Furthermore,*

$$r(T) \leq \dim \mathcal{H}_1 \cdot \dim \mathcal{H}_2. \tag{1.65}$$

3. *There exists a Kraus representation with $r = \mathrm{rank}(\tau)$ operators such that $\langle K_i, K_j\rangle = \mathrm{tr}(K_i^\dagger K_j) = c_i \delta_{i,j}$ for some constant c_i that results from using an unnormalized γ state.*

4. *T is TP iff $\sum_i K_i^\dagger K_i = \mathbb{1}_1$, and unital iff $\sum_i K_i K_i^\dagger = \mathbb{1}_2$*

5. *Any two Kraus representations $\{K_i\}_i$ and $\{L_i\}_i$ of a channel T (i.e., $T(X) = \sum_i K_i X K_i^\dagger = \sum_j L_j X L_j^\dagger$) are related via a unitary U,*

$$K_i = \sum_j U_{ij} L_j. \tag{1.66}$$

Proof. 1. ($\Leftarrow$) is straightforward: Consider the Choi operator

$$\tau = (\mathrm{id} \otimes T)(\gamma) = \sum_i (\mathbb{1} \otimes K_i)\gamma(\mathbb{1} \otimes K_i^\dagger). \tag{1.67}$$

Clearly, $(\mathbb{1} \otimes A)\gamma(\mathbb{1} \otimes A)^\dagger \geq 0$ for any A. Thus, by Prop. 5 the channel is CP.

($\Rightarrow$) Going the other way, we know T is CP $\Leftrightarrow \tau \geq 0$, by Prop. 5.

Claim: $X \geq 0 \Leftrightarrow$ there are $\{|\psi_i\rangle\}_{i=1}^r$ with $r \geq \operatorname{rank} X$ such that $X = \sum_i |\psi_i\rangle \langle\psi_i|$.[9]

[9] Generally, $\langle\psi_i|\psi_j\rangle \neq \delta_{ij}$.

Proof of Claim. ($\Leftarrow$) $X = \sum_i |\psi_i\rangle \langle\psi_i|$ for some $\{|\psi_i\rangle\}_i$.

We know, $X \geq 0 :\Leftrightarrow \langle\varphi|X|\varphi\rangle \geq 0$ for all $|\varphi\rangle$. This allows us to write

$$\langle\varphi| \left(\sum_i |\psi_i\rangle \langle\psi_i| \right) |\varphi\rangle = \sum_i |\langle\varphi|\psi_i\rangle|^2 \geq 0. \tag{1.68}$$

($\Rightarrow$) $X \geq 0 : X = \sum_i \lambda_i |X_i\rangle \langle X_i|$ where $X|X_i\rangle = \lambda_i |X_i\rangle$ and $\langle X_i|X_j\rangle = \delta_{ij}$. Let $|\psi_i\rangle := \sqrt{\lambda_i}|X_i\rangle \implies X = \sum_i |\psi_i\rangle \langle\psi_i|$. ♦

Returning to the proof of the first item of the proposition, we have that $\tau \geq 0 \Rightarrow \exists\{|\psi_i\rangle\}$ such that $\tau = \sum_i |\psi_i\rangle \langle\psi_i|$. Further, for all i there exists K_i such that $|\psi_i\rangle = (\mathbb{1}_1 \otimes K_i)|\gamma\rangle$. Then we can write

$$\tau = \sum_i |\psi_i\rangle \langle\psi_i| = \sum_i (\mathbb{1}_1 \otimes K_i)|\gamma\rangle \langle\gamma| (\mathbb{1}_1 \otimes K_i)^\dagger \tag{1.69}$$

by the Choi-Jamiołkowski isomorphism. We conclude that $T(X) = \sum_i K_i X K_i^\dagger$.[10]

[10] Recall that the Choi-Jamiołkowski isomorphism is one-to-one and can be inverted. Thus, if we know that a Choi operator takes the form $\tau = (\mathbb{1} \otimes A)\gamma(\mathbb{1} \otimes A)^\dagger$ the corresponding channel must be $T(X) = AXA^\dagger$.

2. Clear from the proof of item 1.: $r = \operatorname{rank}(\tau) \implies$ there are at least r pure states in the decomposition of τ. This implies that there are at least r Kraus operators. On the other hand, τ acts on the space $\mathcal{H}_1 \otimes \mathcal{H}_2$ of dimension $\dim \mathcal{H}_1 \cdot \dim \mathcal{H}_2$, and hence we can always achieve $r \leq \dim \mathcal{H}_1 \cdot \dim \mathcal{H}_2$.

3. Let $\tau = \lambda_i |X_i\rangle \langle X_i|$ be a spectral decomposition of τ with $|\varphi_i\rangle = \sqrt{\lambda_i}|X_i\rangle$. We have

$$\tau = \sum_i |\varphi_i\rangle \langle\varphi_i| \tag{1.70}$$

$$= \sum_i (\mathbb{1} \otimes L_i)|\gamma\rangle \langle\gamma| (\mathbb{1} \otimes L_i)^\dagger \tag{1.71}$$

$$\Rightarrow \quad T = \sum_{i=1}^{r(\tau)} L_i(\cdot)L_i^\dagger \tag{1.72}$$

We also need to show $\langle L_i, L_j \rangle = c\delta_{ij}$. We know from the "steering equality" that

$$|\varphi_i\rangle = (\mathbb{1}_1 \otimes L_i) |\gamma\rangle . \tag{1.73}$$

We can thus write

$$\begin{aligned}
\lambda_i \delta_{ij} &= \langle \varphi_i | \varphi_j \rangle && (1.74)\\
&= \langle \gamma | \left(\mathbb{1}_1 \otimes L_i^\dagger \right) (\mathbb{1} \otimes L_i) |\gamma\rangle && (1.75)\\
&= \sum_{k,l} \langle k|l\rangle \langle k| L_i^\dagger L_j |l\rangle && (1.76)\\
&= \sum_{k} \langle k| L_i^\dagger L_j |k\rangle && (1.77)\\
&= \operatorname{tr}\left(L_i^\dagger L_j \right) && (1.78)\\
&= \langle L_i, L_j \rangle . && (1.79)
\end{aligned}$$

4. We have

$$\begin{aligned}
\operatorname{tr}(T(X)) &= \sum_i \operatorname{tr}\left(K_i X K_i^\dagger \right) && (1.80)\\
&= \sum_i \operatorname{tr}\left(K_i^\dagger K_i X \right) \quad \text{(by cyclicity of trace)} && (1.81)\\
&= \operatorname{tr}\left(\textstyle\sum_i K_i^\dagger K_i X \right) \quad \text{(by linearity of trace)} && (1.82)\\
&= \operatorname{tr} X, \quad \forall X && (1.83)\\
&\iff \sum_i K_i^\dagger K_i = \mathbb{1}_1, && (1.84)
\end{aligned}$$

and

$$T(\mathbb{1}_1) = \sum_i K_i \mathbb{1} K_i^\dagger = \sum_i K_i K_i^\dagger = \mathbb{1}_2. \tag{1.85}$$

5. **Claim**: $\sum_i |\psi_i\rangle \langle \psi_i| = \sum_j |\varphi_j\rangle \langle \varphi_j|$ iff there exists a unitary, U, with $|\psi_i\rangle = \sum_j U_{ij} |\varphi_j\rangle$.

Proof of Claim. Consider the purifications $|\Psi\rangle = \sum_i |\psi_i\rangle \otimes |i\rangle$ and $|\Phi\rangle = \sum_j |\varphi_j\rangle \otimes |j\rangle$, where $\{|i\rangle\}$ is a reference system's orthonormal basis. Then, there always exists an isometry[11] V such that $(\mathbb{1} \otimes V) |\Psi\rangle = |\Phi\rangle$.

The above result is useful for our proof because the isometry V

[11] Recall that an isometry is a map $V\colon \mathcal{H}_1 \to \mathcal{H}_2$ such that $\langle \psi|\varphi\rangle = \langle \psi|V^\dagger V|\varphi\rangle$ for all $|\psi\rangle, |\varphi\rangle \in \mathcal{H}_1$. equivalently, $V^\dagger V = \mathbb{1}_1$. In general $\dim \mathcal{H}_2 \geq \dim \mathcal{H}_1$. When the dimensions coincide, V is called a unitary.

above can always be extended to a unitary.

$$|\varphi_i\rangle = (\mathbb{1} \otimes \langle i|)\,|\Phi\rangle\,, \tag{1.86}$$
$$= (\mathbb{1} \otimes \langle i|)\,(\mathbb{1} \otimes V)\,|\Psi\rangle\,, \tag{1.87}$$
$$= (\mathbb{1} \otimes \langle i|)\,(\mathbb{1} \otimes V)\sum_j |\psi_j\rangle \otimes |j\rangle\,, \tag{1.88}$$
$$= \sum_j |\psi_j\rangle\,\langle i|V|j\rangle\,, \tag{1.89}$$
$$= \sum_j V_{ij}\,|\psi_j\rangle\,, \tag{1.90}$$

as desired. ♦

Applying this claim to the two different Kraus representations $T(X) = \sum_i K_i X K_i^\dagger$ and $T(X) = \sum_j L_j X L_j^\dagger$ proves the claim. □

Proposition 7. *Consider the map $T\colon B(\mathcal{H}_1 \to \mathcal{H}_2)$.*

1. *T is completely positive iff $\exists V\colon \mathcal{H}_1 \to \mathcal{H}_2 \otimes \mathbb{C}^n$ such that*
   - *$r \geq r(T)$,*
   - *$T(X) = \mathrm{tr}_E(VXV^\dagger)$,*

   where the trace is over the environment, in this case $\mathbb{C}^n$. And
2. *T is trace-preserving iff $V\colon \mathcal{H}_1 \to \mathcal{H}_2 \otimes \mathbb{C}^n$ is an isometry.*

Proof. 1. ($\Leftarrow$) ✓
($\Rightarrow$) We will show that T is completely positive by showing that there exists Kraus operators $\{K_i\}_{i=1}^n$ such that $T(X) = \sum_i K_i X K_i^\dagger$, $r \geq r(T)$. Choose an orthonormal basis $\{|i\rangle\}_{i=1}^n$ in $\mathbb{C}^n$, so we can write $V = \sum_{i=1}^r K_i \otimes |i\rangle$. Then we have

$$\mathrm{tr}_E\, VXV^\dagger = \sum_{i,j} \mathrm{tr}_1\left((K_i \otimes |i\rangle)X(K_j \otimes |j\rangle)^\dagger\right), \tag{1.91}$$
$$= \sum_{i,j} K_i X K_j\,\mathrm{tr}(|i\rangle\,\langle j|), \tag{1.92}$$
$$= \sum_i K_i X K_i^\dagger, \tag{1.93}$$
$$= T(X). \tag{1.94}$$

2. Now, we wish to show that T is trace-preserving iff $V^\dagger V = \mathbb{1}$. We

write

$$\begin{aligned} \operatorname{tr} X &= \operatorname{tr}(T(X)), && (1.95)\\ &= \operatorname{tr}(\operatorname{tr}_E(VXV^\dagger)), && (1.96)\\ &= \operatorname{tr}(VXV^\dagger), && (1.97)\\ &= \operatorname{tr}(V^\dagger VX), && (1.98)\\ &= \operatorname{tr}(X), \quad \forall X \Leftrightarrow V^\dagger V = \mathbb{1}, && (1.99) \end{aligned}$$

which finishes the proof. □

We conclude with a few remarks:

- The isometry V in Prop. 7 has a special name. It is called the *Stinespring isometry* or *Stinespring dilation* of the channel T.
- Given a Stinespring isometry V, a Kraus representation $\{K_i\}_i$ is obtained via[12]

$$K_i(\mathbb{1} \otimes \langle i|)V \leftrightarrow T(X) = \operatorname{tr}_E VXV^\dagger \tag{1.100}$$

[12] Remember that the partial trace is given as: $\operatorname{tr}_E Y = \sum_i (\mathbb{1} \otimes \langle i|)Y(\mathbb{1} \otimes |i\rangle)$.

1.5 Unitary Picture and Open System Dynamics

We now come to an important conclusion: *open system dynamics can always be viewed as a unitary on the system and environment, where we trace over the environment at the end.*

Proposition 7 shows that for a quantum channel $T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ there exists a Stinespring isometry $V\colon \mathcal{H}_1 \to \mathcal{H}_2 \otimes \mathbb{C}^r$ such that $T(X) = \operatorname{tr}_E VXV^\dagger$. Here, the dimension r of the auxiliary system is at least equal to the Kraus rank $r(T)$ of the channel T. Our goal is to reformulate this isometric representation of the channel using a unitary.

To this end, we set $d_i := \dim \mathcal{H}_i$ and choose $r = d_1 d_2$ for the auxiliary system in the definition of the channel isometry V. This allows us to complete V to a unitary U acting on $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \mathbb{C}^{d_2}$ such that

$$V = U(\mathbb{1}_{\mathcal{H}_1} \otimes |\varphi\rangle) \tag{1.101}$$

for some suitable vector $|\varphi\rangle \in \mathbb{C}^{d_2} \otimes \mathbb{C}^{d_2}$. We then have

$$T(X) = \operatorname{tr}_E VXV^\dagger = \operatorname{tr}_{E'} U(X \otimes |\varphi\rangle\langle\varphi|)U^\dagger, \tag{1.102}$$

where the 'new' environment E' is taken to be $E' \cong \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$.

1.6 Complementary Channels

Via Prop. 7 we know that any quantum channel can be represented through an isometry, $T(X_A) = \operatorname{tr}_E VX_AV^\dagger$. In a noisy channel the

input system interacts with the environment. We can then ask what the environment learns about the input X_A.

Given a channel $T\colon B(\mathcal{H}_A) \to B(\mathcal{H}_B)$ with Stinespring isometry $V\colon \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$, we define the *complementary channel* $T^c\colon B(\mathcal{H}_A) \to B(\mathcal{H}_E)$ as tracing out the receiver B instead of the environment E:

$$T^c(X_A) = \operatorname{tr}_B V X_A V^\dagger. \tag{1.103}$$

Note that, for a channel T with given isometry V, the isometry $\tilde{V} = (\mathbb{1} \otimes U)V$ with U unitary induces the same channel:

$$\begin{aligned} \operatorname{tr}_E \tilde{V} X_A \tilde{V}^\dagger &= \operatorname{tr}_E (\mathbb{1} \otimes U) V X_A V^\dagger (\mathbb{1} \otimes U^\dagger) && (1.104)\\ &= \operatorname{tr}_E V X_A V^\dagger && (1.105)\\ &= T(X_A), && (1.106) \end{aligned}$$

where the second equality holds via (partial) cyclicity of the partial trace. While this unitary shift leaves the channel invariant, it changes the complementary channel:

$$\begin{aligned} \operatorname{tr}_B \tilde{V} X_A \tilde{V}^\dagger &= \operatorname{tr}_B (\mathbb{1} \otimes U) V X_A V^\dagger (\mathbb{1} \otimes U^\dagger) && (1.107)\\ &= U \operatorname{tr}_B V X_A V^\dagger U^\dagger && (1.108)\\ &= U T^c(X_A) U^\dagger. && (1.109) \end{aligned}$$

This of course is related to the fact that we fix a basis in the construction of the isometry $V = \sum_i K_i \otimes |i\rangle$, where $\{K_i\}_i$ are the Kraus operators of the channel and $\{|i\rangle\}_i$ is the orthonormal basis of our choice.

While these different complementary channels are technically different from each other, they are all unitarily equivalent, as can be seen from (1.109). Unitarily equivalent channels have the same information-theoretic properties, so that the specific choice of complementary channel does not matter. This allows us to make a suitable choice of complementary channel for the specific problem of interest.

1.7 *Linear Representation*

Here, we quickly review some definitions from linear algebra that we will need. We denote a linear map acting from one bounded Hilbert space to another as $T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$. We then know, from linear algebra, that linear maps on a finite-dimensional Hilbert space correspond to matrices with respect to a fixed basis. Further, recall the following definitions and facts:

- $B(\mathcal{H})\ldots$ Hilbert space with inner product: $\langle X, Y\rangle = \operatorname{tr}(X^\dagger Y)$.
- We treat operators $X \in B(\mathcal{H})$ as vectors and maps $T\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ as matrices.

- Let $\{|i\rangle\}_{i=1}^{d}, d = \dim \mathcal{H}$, be an orthonormal basis for $\mathcal{H}$. Then $\{|i\rangle\langle j|\}_{i,j=1}^{d}$ is a basis for $B(\mathcal{H})$, which implies $\dim B(\mathcal{H}) = d^2$.

We now consider representing quantum channels through *transfer matrices*. To do so, we first introduce the following linear mapping

Definition 8 (Vectorization map)**.** The vectorization map vec is defined by its action on basis elements,

$$\text{vec}\colon B(\mathcal{H}) \to \mathcal{H} \otimes \mathcal{H} \tag{1.110}$$

$$\text{vec}(|i\rangle\langle j|) \to |i\rangle \otimes |j\rangle\,, \tag{1.111}$$

and extended linearly to all operators in $B(\mathcal{H})$.

Because $\{|i\rangle \otimes |j\rangle\}_{i,j=1}^{d}$ is an orthonormal basis for $\mathcal{H} \otimes \mathcal{H}$, the map vec is an isomorphism between $B(\mathcal{H})$ and $\mathcal{H} \otimes \mathcal{H}$. We note that vec has the following properties:[13]

[13] Note that $|\overline{x}\rangle$ represents the complex conjugate of x with respect to the basis $\{|i\rangle\}_i$.

1. $\text{vec}(|\psi\rangle\langle\varphi|) = |\psi\rangle \otimes |\overline{\varphi}\rangle$.
2. vec is an isometry and thus $\langle X, Y\rangle_{B(\mathcal{H})} = \langle \text{vec}(X), \text{vec}(Y)\rangle_{\mathcal{H}\otimes\mathcal{H}}$.
3. $\text{vec}(AXB) = (A \otimes B^T)\,\text{vec}(X)$.

The third property above lets us define the vectorization of a quantum channel using the operator-sum representation. Given a linear map $\mathcal{N}(X) = \sum_i A_i X B_i$, the vectorization map assigns to $\mathcal{N}\colon B(\mathcal{H}) \to B(\mathcal{H})$ the following operator $N \in B(\mathcal{H} \otimes \mathcal{H})$:

$$N = \textstyle\sum_i A_i \otimes B_i^T. \tag{1.112}$$

As we have seen before, any quantum channel admits a Kraus decomposition $\mathcal{N}(X) = \sum_i K_i X K_i^\dagger$. Thus, we make the following definition.

Definition 9 (Transfer matrix)**.** The *transfer matrix* of a quantum channel $\mathcal{N}\colon B(\mathcal{H}) \to B(\mathcal{H})$ is the matrix $N \in B(\mathcal{H} \otimes \mathcal{H})$ obtained through

$$\mathcal{N}(X) = \textstyle\sum_i K_i X K_i^\dagger \xleftrightarrow{\text{vec}} N = \sum_i K_i \otimes \overline{K}_i. \tag{1.113}$$

The transfer matrix has the following properties:

1. If $\mathcal{N}$ is unital, then $\sum_i K_i K_i^\dagger = \mathbb{1}$ and hence

$$N\,|\gamma\rangle = \left(\textstyle\sum_i K_i \otimes \overline{K}_i\right)|\gamma\rangle = |\gamma\rangle = \text{vec}(\mathbb{1}), \tag{1.114}$$

where the second to last equality holds via the transpose trick.[14]

[14] For the maximally entangled state $|\gamma\rangle = \sum_{i=1}^{\dim \mathcal{H}} |i\rangle|i\rangle$ and any $A \in B(\mathcal{H})$ the following "transpose trick" property holds: $A \otimes \mathbb{1}\,|\gamma\rangle = \mathbb{1} \otimes A^T\,|\gamma\rangle$

2. If $\mathcal{N}$ is trace preserving, then $\sum_i K_i^\dagger K_i = \mathbb{1}$, and hence

$$N^\dagger |\gamma\rangle = (\sum_i K_i^\dagger \otimes K_i^T) |\gamma\rangle = |\gamma\rangle . \tag{1.115}$$

The linear representation is useful, because it translates channel composition into matrix multiplication.

Proposition 10. *Let $\mathcal{M}\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ and $\mathcal{N}\colon B(\mathcal{H}_2) \to B(\mathcal{H}_3)$ be linear maps with transfer matrices M, N, respectively. Then NM is the transfer matrix of $\mathcal{N} \circ \mathcal{M}\colon B(\mathcal{H}_1) \to B(\mathcal{H}_3)$.*

Proof. We can express the linear maps as

$$\mathcal{M}(X) = \sum_i A_i X B_i^T, \quad A_i, B_i \in B(\mathcal{H}_1, \mathcal{H}_2), \tag{1.116}$$

$$\mathcal{N}(X) = \sum_i C_i X D_i^T, \quad C_i, D_i \in B(\mathcal{H}_2, \mathcal{H}_3), \tag{1.117}$$

and so their composition is given as

$$\mathcal{N} \circ \mathcal{M} = \sum_{i,j} C_i A_j X B_j^T D_i^T = \sum_{i,j} C_i A_j X (D_i B_j)^T, \tag{1.118}$$

$$\implies \operatorname{vec}(\mathcal{N} \circ \mathcal{M}) = \sum_{i,j} C_i A_j \otimes D_i B_j. \tag{1.119}$$

Then, from the definition of the transfer matrices,

$$M = \sum_i A_i \otimes B_i, N = \sum_i C_i \otimes D_i, \tag{1.120}$$

we see

$$NM = \sum_{i,j} C_i A_j \otimes D_i B_j, \tag{1.121}$$

as desired. □

Proposition 11. *Let $\mathcal{N}\colon B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ be a linear map with transfer matrix $N \in B(\mathcal{H}_1, \mathcal{H}_2)$ and Choi operator $\tilde{\tau} = (\mathcal{N} \otimes \mathbb{1}_2)(\gamma)$. Then $\tilde{\tau} = N^\Gamma$, where Γ is an involution ($\Gamma^2 = \mathbb{1}$) defined by $\langle i, j| X^\Gamma |k, l\rangle = \langle i, k| X |j, l\rangle$.*

Proof. We first consider the simple case where $\mathcal{N}(\cdot) = X(\cdot)Y^T$, where $X, Y \in B(\mathcal{H}_1, \mathcal{H}_2)$. Fix an orthonormal basis $\{|i\rangle\}_i$ for $\mathcal{H}_1$ and $\{|\alpha\rangle\}_\alpha$ for $\mathcal{H}_2$. Then X and Y can be expanded in these bases as

$$X = \sum_{i,\alpha} x_{\alpha,i} |\alpha\rangle \langle i| \qquad Y = \sum_{j,\beta} y_{\beta,j} |\beta\rangle \langle j| . \tag{1.122}$$

Thus, the transfer matrix of the channel is

$$N = X \otimes Y = \sum_{i,j,\alpha,\beta} x_{\alpha,i} y_{\beta,j} \left|\alpha\right\rangle \left\langle i\right| \otimes \left|\beta\right\rangle \left\langle j\right|, \tag{1.123}$$

and the Choi operator is

$$\begin{aligned}
\tilde{\tau} = (\mathcal{N} \otimes \mathbb{1}_2)(\gamma) &= \sum_{h,l} X \left|h\right\rangle \left\langle l\right| Y^T \otimes \left|h\right\rangle \left\langle l\right| && (1.124)\\
&= \sum_{\substack{h,l\\ i,j,\alpha,\beta}} x_{\alpha,i} y_{\beta,j} \left|\alpha\right\rangle \left\langle i|h\right\rangle \left\langle l|j\right\rangle \left\langle \beta\right| \otimes \left|h\right\rangle \left\langle l\right| && (1.125)\\
&= \sum_{i,j,\alpha,\beta} x_{\alpha,i} y_{\beta,j} \left|\alpha\right\rangle \left\langle \beta\right| \otimes \left|i\right\rangle \left\langle j\right| && (1.126)\\
&= N^{\Gamma}. && (1.127)
\end{aligned}$$

The general claim now holds by linearity. □

Note that $\langle \alpha, \beta | N | i, j \rangle = \langle \alpha, i | \tilde{\tau} | \beta, j \rangle$ and that $N \in B(\mathcal{H}_1 \otimes \mathcal{H}_1, \mathcal{H}_2 \otimes \mathcal{H}_2)$ but $\tilde{\tau} \in B(\mathcal{H}_2 \otimes \mathcal{H}_1, \mathcal{H}_2 \otimes \mathcal{H}_1)$.

2

Classes of quantum channels

2.1 *Qubit channels*

Bit-flip channel or X-dephasing channel

In classical information theory, the bit flip channel is a channel in which bits are flipped with probability p, and are transmitted unchanged with probability $1-p$. This is shown in Fig. 2.1.

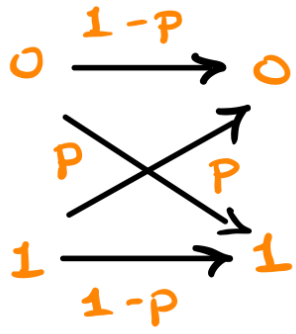

Figure 2.1: Classical bit-flip channel.

Before meeting the quantum version, recall that the Pauli-X operator is given as

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{2.1}$$

It follows that $X|0\rangle = |1\rangle$ and $X|1\rangle = |0\rangle$. Furthermore, the eigenstates of X are simply $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. The quantum version takes on the form

$$\mathscr{F}_p^X : \rho \mapsto (1-p)\rho + pX\rho X, \tag{2.2}$$

which is a channel with Kraus operators $\{\sqrt{1-p}\mathbb{1}, \sqrt{p}X\}$. Visually, the bit-flip channel squishes in the y- and z-direction in the Bloch sphere.

Phase-flip channel or Z-dephasing channel

The Z-analogue of the bit-flip channel is a phase-flip channel. For this channel we consider the Pauli-Z operator

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.3}$$

with $Z|0\rangle = |0\rangle$ and $Z|1\rangle = -|1\rangle$. Furthermore, $Z|+\rangle = |1\rangle$ and $Z|-\rangle = |+\rangle$. The Z-dephasing channel takes the form

$$\mathscr{F}_p^Z : \rho \mapsto (1-p)\rho + pZ\rho Z. \tag{2.4}$$

Bit-phase flip channel or Y-dephasing channel

Finally, we can look at a combined bit- and phase-flip channel, caused by the action of the Pauli-Y operator

$$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{2.5}$$

satisfying $Y|0\rangle = i|1\rangle$ and $Yi|1\rangle = |0\rangle$. The eigenbasis is given as $|\pm i\rangle = (|0\rangle \pm i|1\rangle)/\sqrt{2}$. The Y-dephasing channel takes the form

$$\mathscr{F}_p^Y : \rho \mapsto (1-p)\rho + pY\rho Y. \tag{2.6}$$

Because $Y = -iZX = iXZ$, we can also write this map as

$$\mathscr{F}_p^Y : \rho \mapsto (1-p)\rho + pXZ\rho ZX, \tag{2.7}$$

which is why the channel is sometimes called the "bit-phase flip channel."

We note that all of these Pauli channels are unital because each Pauli operator squares to the identity. For example,

$$\begin{aligned} \mathscr{F}_p^Y(\mathbb{1}) &= (1-p)\mathbb{1} + pY\mathbb{1}Y, && (2.8)\\ &= (1-p)\mathbb{1} + pY^2, && (2.9)\\ &= (1-p)\mathbb{1} + p\mathbb{1}, && (2.10)\\ &= \mathbb{1}, && (2.11) \end{aligned}$$

as claimed.

Depolarizing channel

We now introduce an important channel that represents each Pauli error (X, Y, Z) occurring with equal probability $p/3$. This is called the *depolarizing channel*, and is given as

$$\mathscr{D}_p : \rho \mapsto (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z). \tag{2.12}$$

The action of the depolarizing channel is to uniformly shrink the sphere. The Kraus operators for the depolarizing channel are given as

$$\{\sqrt{1-p}\mathbb{1}, \sqrt{\frac{p}{3}}X, \sqrt{\frac{p}{3}}Y, \sqrt{\frac{p}{3}}Z\}. \tag{2.13}$$

Alternatively, and perhaps more intuitively, this channel can be represented as

$$\rho \mapsto (1-q)\rho + q\frac{\mathbb{1}}{2}. \tag{2.14}$$

This form makes clear that the depolarizing channel is effectively replacing an input state with the completely mixed state with "probability" q.[1]

[1] Because q can exceed 1 (see (2.17)), it is not actually a probability.

We can work out the relationship between p and q as follows: An easy calculation shows that for all $\rho \in B(\mathbb{C}^2)$ we have

$$\frac{1}{4}(\rho + X\rho X + Y\rho Y + Z\rho Z) = \frac{\operatorname{tr}\rho}{2}\mathbb{1}. \tag{2.15}$$

With this in mind, we have

$$(1-q)\rho + q\frac{\mathbb{1}}{2} = (1-q)\rho + \frac{q}{4}(\rho + X\rho X + Y\rho Y + Z\rho Z), \tag{2.16}$$

$$\implies q = \frac{4}{3}p. \tag{2.17}$$

Hence, $p \in [0,1]$ means that $q \in [0, 4/3]$.

(Generalized) Pauli channels

Let $\mathbf{p} = (p, p_1, p_2, p_3)$ be a probability distribution. Then we consider the generalized Pauli channel

$$\mathcal{N}_{\mathbf{p}}(\rho) = p_0\rho + p_1 X\rho X + p_2 Y\rho Y + p_3 Z\rho Z. \tag{2.18}$$

For example, the depolarizing channel is obtained by setting $p_0 = 1-p$ and $p_i = \frac{p}{3}$. The channel $\mathcal{N}_{\mathbf{p}}$ is unital for all $\mathbf{p}$.

Pauli channels are interesting from an information theory point of view because

- classical information transmission is almost completely understood;
- quantum information transmission is really not understood at all (except in some special cases like the flip/dephasing channels).

Amplitude Damping Channel

Consider a physical 2-level system with ground state $|0\rangle$ and excited state $|1\rangle$. If the system is in the excited state it will decay down to the ground state with some probability γ, emitting a photon in the process. We model this process with the *amplitude damping* channel $\mathcal{A}_\gamma$. The channel isometry is given by

$$|0\rangle_A \to |0\rangle_B \otimes |0\rangle_E \tag{2.19}$$

$$|1\rangle_A \to \sqrt{1-\gamma}\,|1\rangle_B \otimes |0\rangle_E + \sqrt{\gamma}\,|0\rangle_B \otimes |1\rangle_E\,. \tag{2.20}$$

A choice of Kraus operators for this channel is the following:

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \qquad K_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}. \tag{2.21}$$

We make the following claims about amplitude damping channels:

1. $\mathcal{A}_\gamma(\mathbb{1}) = \begin{pmatrix} 1+\gamma & 0 \\ 0 & 1-\gamma \end{pmatrix}$ and thus this is not a unital channel.
2. $\mathcal{A}_\gamma$ is not a Pauli channel (more generally, it is not even a mixed unitary channel).
3. Sending quantum information through $\mathcal{A}_\gamma$ is completely understood. However, how much classical information can be sent through $\mathcal{A}_\gamma$ is not fully understood.

Erasure Channel

Classically, an erasure channel transmits a bit unchanged with probability $1-p$ and destroys the message with probability p. An erased bit is denoted by some flag e. The quantum version transmits a qubit state unchanged with probability $1-p$ and replaces it with some flag state $|e\rangle$ with probability p. Note that we take $|e\rangle \perp \mathcal{H}$. Mathematically, this is written as

$$\mathcal{E}_p \colon B(\mathcal{H}) \to B(\mathcal{H} \oplus \mathbb{C}), \tag{2.22}$$

$$\mathcal{E}_p(\rho) = (1-p)\rho) + p \operatorname{tr}(\rho) |e\rangle \langle e| . \tag{2.23}$$

This channel can be represented in Kraus form as follows:

$$\begin{aligned} K_0 &= \sqrt{1-p}\,\mathbb{1} = \sqrt{1-p} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \\ K_1 &= \sqrt{p}\,|e\rangle\langle 0| = \sqrt{p} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} \\ K_2 &= \sqrt{p}\,|e\rangle\langle 1| = \sqrt{p} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \tag{2.24}$$

Note that Bob can always tell if an erasure has occurred simply by measuring the output state. All major capacities of the erasure channel are understood because of this structure.

2.2 *Generalized Dephasing Channels*

We now extend the notion of a dephasing channel to larger systems. A generalized dephasing channel leaves a fixed orthonormal basis invariant and dephases the off-diagonal elements of a density matrix with respect to this basis. Consider some Hilbert space $\mathcal{H} = \mathbb{C}^d$ with orthonormal basis $\{|i\rangle\}_{i=1}^d$. Further, let us consider an environment

$\mathcal{H}_E$ of dimension at least 2. Let $\{|\varphi_i\rangle_E\}_{i=1}^d$ be some set of states (not necessarily orthogonal) on $\mathcal{H}_E$ and consider the isometry

$$|i\rangle_A \to |i\rangle_B \otimes |\psi_i\rangle_E \,. \tag{2.25}$$

This leads to the following channel:

$$\begin{aligned} \mathcal{N}(\rho) &= \operatorname{tr}_E V\rho V^\dagger && (2.26)\\ &= \sum_{i,j} \langle\varphi_i|\varphi_j\rangle \langle i|\rho|j\rangle |i\rangle \langle j|_B \,. && (2.27) \end{aligned}$$

Notice that diagonal elements are left unchanged ($[\mathcal{N}(\rho)]_{ii} = \rho_{ii}$). Off-diagonal elements, however, may be dampened as $\langle\varphi_i|\varphi_j\rangle$, which is generally not equal to 0.

We now introduce the generalized Pauli operators.

$$\text{"Shift Operator": } X|i\rangle = |i+1 \bmod d\rangle \,, \tag{2.28}$$

$$\text{"Clock Operator": } Z|j\rangle = \omega^j |j\rangle \,, \tag{2.29}$$

where $\omega = e^{\frac{2\pi i}{d}}$. Note that $X^d = Z^d = \mathbb{1}$. Lastly, $X \neq X^\dagger$ and $Z \neq Z^\dagger$ for $d > 2$. These are the generators of the Heisenberg-Weyl group:

$$\{\omega^j Z^k X^l \colon j,k,l \in [d]\}. \tag{2.30}$$

With these defined, we can construct generalized dephasing channels in a method similar to that of qubit channels. The following are examples of such:

$$\rho \mapsto (1-p)\rho + pX\rho X^\dagger, \tag{2.31}$$

$$\rho \mapsto (1-p)\rho + \frac{p}{3} Z\rho Z^\dagger + \frac{2p}{3} Z^2 \rho (Z^2)^\dagger. \tag{2.32}$$

The Holevo information of dephasing channels

A generalized dephasing channel $\mathcal{N}\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ as defined in (2.25) leaves the elements of an orthonormal basis $\{|i\rangle\}_{i=1}^d$ invariant. Intuitively, this means that a dephasing channel has maximal[2] classical capacity equal to $\log d$, where d is the dimension of the input and output spaces. To see this, note that we can encode a classical random variable taking values $i = 1, \cdots, d$ in the input space via the encoding $i \mapsto |i\rangle\langle i|$. Since $\mathcal{N}(|i\rangle\langle i|) = |i\rangle\langle i|$ for all i, these orthonormal basis states remain perfectly distinguishable, and hence Bob can fully recover the classical information encoded by Alice. In the following, we want to make this argument more formal using an entropic quantity called *Holevo information*.

[2] One can show that any capacity of a quantum channel $\mathcal{N}\colon \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ is bounded from above by $\min\{\log|A|, \log|B|\}$.

We start by recalling the *von Neumann entropy* of a quantum state ρ:

$$S(\rho) = -\operatorname{tr}\rho\log\rho = -\sum_i \lambda_i \log\lambda_i, \tag{2.33}$$

where the matrix logarithm of ρ is defined on the positive part of ρ as follows: Let $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$ be a spectral decomposition of ρ, then we define

$$\log \rho = \sum_{i:\, \lambda_i > 0} (\log \lambda_i) |e_i\rangle\langle e_i|. \tag{2.34}$$

The second equality in (2.33) follows in a straightforward manner using this definition, and shows that the von Neumann entropy is equal to the Shannon entropy of its eigenvalues $(\lambda_i)_i$. The most important properties of von Neumann entropy are summarized in Proposition 43 in Section 5.

Let now $E = (p_x, \rho_x)$ be an *ensemble of quantum states*, that is, $(p_x)_x$ is a probability distribution and $\rho_x \in \mathcal{B}(\mathcal{H})$ is a quantum state on a d-dimensional Hilbert space $\mathcal{H}$ for each x. Operationally, we can think of our system being prepared in state ρ_x with probability x. We define the *Holevo information* $\chi(E)$ of the quantum state ensemble E as

$$\chi(E) := S\left(\sum_x p_x \rho_x\right) - \sum_x p_x S(\rho_x). \tag{2.35}$$

Operationally, the Holevo information $\chi(E)$ quantifies the maximal amount of classical information that can be retrieved from the quantum state ensemble E by measuring the quantum systems.

In the setting where Alice and Bob are connected by a noisy quantum channel, sending classical information essentially corresponds to Alice encoding this information in a quantum system, sending it to Bob using the noisy channel, and Bob measuring the received system to retrieve the classical message. In analogy to (2.35), we thus define the *Holevo information* $\chi(\mathcal{N}, E)$ of a quantum channel $\mathcal{N}\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ and a given quantum state ensemble E as

$$\chi(\mathcal{N}, E) := S\left(\sum_x p_x \mathcal{N}(\rho_x)\right) - \sum_x p_x S(\mathcal{N}(\rho_x)). \tag{2.36}$$

Because Alice is free to choose the ensemble in the encoding, we optimize this quantity over all possible state ensembles, arriving at the *Holevo information* $\chi(\mathcal{N})$ *of a quantum channel*:

$$\chi(\mathcal{N}) := \max_E \chi(\mathcal{N}, E). \tag{2.37}$$

The Holevo-Schumacher-Westmoreland (HSW) theorem[3] states that the Holevo information of a quantum channel is an achievable rate for classical information transmission:

$$C(\mathcal{N}) \geq \chi(\mathcal{N}). \tag{2.38}$$

For any quantum channel $\mathcal{N}\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$, we also have the following upper bound on the classical capacity:

$$C(\mathcal{N}) \leq \min\{\log \dim \mathcal{H}, \log \dim \mathcal{H}'\}. \tag{2.39}$$

[3] Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998. DOI: 10.1109/18.651037; and Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56:131–138, 1997. DOI: 10.1103/PhysRevA.56.131

We now show that any generalized dephasing channel $\mathcal{N}\colon \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ has maximal classical capacity $C(\mathcal{N}) = \log d$. To see this, recall from (2.25) that any generalized dephasing channel is defined in terms of an ONB $\{|i\rangle\}_{i=1}^d$, and satisfies $\mathcal{N}(|i\rangle\langle i|) = |i\rangle\langle i|$. We can use this observation to achieve a value of $\chi(\mathcal{N}, E) = \log d$ by forming the quantum state ensemble $E = (1/d, |i\rangle\langle i|)_i$. Indeed,

$$\chi(N) \geq \chi(\mathcal{N}, E) = S\left(\frac{1}{d}\sum_i \mathcal{N}(|i\rangle\langle i|)\right) - \frac{1}{d}\sum_i S(\mathcal{N}(|i\rangle\langle i|)) \tag{2.40}$$

$$= S\left(\frac{1}{d}\mathbb{1}\right) \tag{2.41}$$

$$= \log d. \tag{2.42}$$

Here, we used $\mathcal{N}(|i\rangle\langle i|) = |i\rangle\langle i|$ in the second equality, together with the fact that the von Neumann entropy vanishes on pure state. On the other hand, since both the input and output spaces of $\mathcal{N}$ are d-dimensional, we get from(2.38) and (2.39) that

$$\chi(\mathcal{N}) \leq C(\mathcal{N}) \leq \log d, \tag{2.43}$$

which together with (2.42) shows that $C(\mathcal{N}) = \chi(\mathcal{N}) = \log d$ for generalized dephasing channels.

We will discuss the Holevo information further in Chapter 3.

2.3 *Entanglement-breaking channels*

Reminder: A bipartite state ρ_{AB} is called separable if

$$\rho_{AB} \in \operatorname{conv}\{\omega_A \otimes \sigma_B \colon \omega_A, \sigma_B \text{ states on } \mathcal{H}_{A/B}\}. \tag{2.44}$$

Explicitly,

$$\rho_{AB} = \sum_i p_i \omega_A^i \otimes \sigma_B^i, \tag{2.45}$$

where the only correlation in the state is the classical correlation contained in the index i.

Definition 12. A channel $\mathcal{N}\colon A \to B$ is called *entanglement-breaking* if $(\mathrm{id}_R \otimes \mathcal{N})(\rho_{RA})$ is separable for any input state ρ_{RA}.

Intuitively, whatever entanglement is present in the input state ρ_{RA} is broken by the channel.

We can characterize entanglement-breaking channels in different ways, as the next result shows.[4]

[4] For Prop. 13d), recall that a POVM $E = \{E_i\}_i$ is a collection of positive semidefinite operators E_i with $\sum_i E_i = \mathbb{1}$. This is the most general measurement considered in quantum mechanics. The different outcomes of the measurement correspond to the label i, being associated with the corresponding "effect operator" E_i. If we measure a system in the state ρ using the POVM $E = \{E_i\}_i$, we obtain the outcome i with probability $\mathrm{tr}(\rho E_i)$.

Proposition 13. *The following are equivalent:*

a) $\mathcal{N}: A \to B$ *is entanglement breaking.*

b) $\tau_{AB}^{\mathcal{N}} = (\mathrm{id}_A \otimes \mathcal{N})(\gamma_{AA'})$ *is separable.*

c) $\mathcal{N}$ *has a Kraus representation consisting of rank-1 Kraus operators.*

d) $\mathcal{N}$ *is a measure-and-prepare channel: there exists a POVM* $E = \{E_i\}_i$ *and states* $(\sigma_i)_i$ *such that*

$$\mathcal{N}(\rho) = \sum_i \mathrm{tr}[\rho E_i]\sigma_i \tag{2.46}$$

Proof. a) $\implies$ b) $(\mathbb{1}_R \otimes \mathcal{N}(\rho_{RA}))$ is separable for all ρ_{RA}. In particular, for γ_{RA}, where $(|\gamma\rangle_{RA} = \sum_i |i\rangle_R |i\rangle_A)$.

b) $\implies$ c) If $\tau_{AB}^{\mathcal{N}}$ 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 $\frac{1}{d}\tau_{AB}^{\mathcal{N}} = \sum_i p_i \psi_i \otimes \varphi_i$, where $d = |A|$. Then, simply set

$$K_i = \sqrt{d p_i}\, |\varphi_i\rangle_B \langle\overline{\psi}_i|_A \,. \tag{2.47}$$

We can check this construction. Observe

$$\frac{1}{d}\sum_i (\mathbb{1}_A \otimes K_i)(\gamma_{AA'})(\mathbb{1} \otimes K_i)^\dagger \tag{2.48}$$

$$= \frac{1}{d}\sum_{i,j,k} |j\rangle \langle k|_A \otimes K_i |j\rangle \langle k|_{A'} K_i^\dagger, \tag{2.49}$$

$$= \frac{1}{d}\sum_{i,j,k} d p_i\, |j\rangle \langle k|_A \otimes \langle\overline{\psi}_i|j\rangle \langle k|\overline{\psi}_i\rangle\, |\varphi_i\rangle \langle\varphi_i|_B \,, \tag{2.50}$$

$$= \sum_{i,j,k} p_i\, |j\rangle \langle k|_A \otimes \langle\overline{\psi}_i|j\rangle \langle k|\overline{\psi}_i\rangle\, |\varphi_i\rangle \langle\varphi_i|_B \,, \tag{2.51}$$

$$= \sum_i p_i \left(\sum_{j,k} \langle k|\, \overline{\psi}_i\, |j\rangle\, |j\rangle \langle k|\right) \otimes |\varphi_i\rangle \langle\varphi|_B \,, \tag{2.52}$$

$$= \sum_i p_i \psi_i \otimes \varphi_i, \tag{2.53}$$

$$= \frac{1}{d}\tau_{AB}^{\mathcal{N}}. \tag{2.54}$$

Further, we have

$$\sum_i K_i^\dagger K_i = \mathbb{1}, \tag{2.55}$$
$$= d \sum_i p_i \left|\overline{\psi}_i\right\rangle \left\langle\varphi_i|\varphi_i\right\rangle \left\langle\overline{\psi}_i\right|, \tag{2.56}$$
$$= d \sum_i p_i \overline{\psi}_i, \tag{2.57}$$
$$= d(\overline{\sum_i p_i \psi_i}), \tag{2.58}$$
$$= d \frac{1}{d} \overline{\mathbb{1}}_A, \tag{2.59}$$
$$= \mathbb{1}_A \tag{2.60}$$

c) $\Longrightarrow$ d) The rank of each Kraus operator is 1: $K_i = \left|\chi_i\right\rangle_B \left\langle\omega_i\right|_A$ for some vectors $\left|\chi_i\right\rangle_B, \left|\omega_i\right\rangle_A$, $\langle\chi_i|\chi_i\rangle = 1$.

$$\mathcal{N}(\rho) = \sum_i K_i \rho K_i^\dagger, \tag{2.61}$$
$$= \sum_i \langle\omega_i|\rho|\omega_i\rangle \left|\chi_i\right\rangle \left\langle\chi_i\right|_B. \tag{2.62}$$

Treating $\omega = \{\omega_i\}$ as a POVM and taking the states to be χ_i's we have

$$\mathbb{1} = \sum_i K_i^\dagger K_i, \tag{2.63}$$
$$= \sum_i \left|\omega_i\right\rangle \left\langle\chi_i|\chi_i\right\rangle \left\langle\omega_i\right|, \tag{2.64}$$
$$= \sum_i \left|\omega_i\right\rangle \left\langle\omega_i\right|. \tag{2.65}$$

d) $\Longrightarrow$ a) Let ρ_{RA} be arbitrary, $\mathcal{N}(\rho) = \sum_i \mathrm{tr}[E_i\rho]\sigma_i$. Then we have

$$(\mathbb{1}_R \otimes \mathcal{N})(\rho_{RA}) = \sum_i \mathrm{tr}_A \left[(\mathbb{1}_R \otimes E_i)\rho_{RA}\right] \otimes \sigma_i, \tag{2.66}$$
$$= \sum_i \mathrm{tr}_A \left[(\mathbb{1}_R \otimes \sqrt{E_i})\rho_{RA}(\mathbb{1}_R \otimes \sqrt{E_i})\right] \otimes \sigma_i, \tag{2.67}$$
$$= \sum_i p_i \omega_i \otimes \sigma_i, \tag{2.68}$$

where $p_i = \mathrm{tr}[E_i\rho_{RA}]$, and where

$$\omega_i = \frac{1}{p_i} \mathrm{tr}_A \left[(\mathbb{1}_R \otimes \sqrt{E_i})\rho_{RA}(\mathbb{1}_A \otimes \sqrt{E_i})\right] \geq 0. \tag{2.69}$$

□

Some channel capacities of entanglement-breaking channels are understood:

- Quantum capacity: Quantum information transmission is equivalent to entanglement generation, but entanglement-breaking channels cannot generate entanglement, thus their quantum capacity is zero.
- Classical capacity: For general channels we have $C(\mathcal{N}) \geq \chi(\mathcal{N})$, and a potential gap would arise from using entangled input states. But entanglement-breaking channels destroy any entanglement between different inputs, which can be used to show that $C(\mathcal{N}) = \chi(\mathcal{N})$.[5] Note however that $\chi(\mathcal{N})$ is NP-hard to compute![6]

2.4 *PPT-channels*

An important result in quantum information theory is that it is NP-hard to tell if a state is separable.[7] This begs the question: is there some easier criterion? One *necessary* criterion is the Peres-Horodecki criterion, which states that

$$\rho_{AB} \in \text{SEP} \implies \rho_{AB} \in \text{PPT}, \tag{2.70}$$

where PPT is the set of all states with *positive partial transpose*:

$$\rho_{AB}^{T_B} \geq 0. \tag{2.71}$$

To prove (2.70), we have for separable ρ_{AB}

$$\rho_{AB} = \sum_i p_i \omega_A^i \otimes \sigma_B^i \quad \implies \quad \rho_{AB}^{T_B} = \sum_i p_i \omega_A^i \otimes (\sigma_B^i)^T \geq 0. \tag{2.72}$$

The negation of (2.70) gives us the following sufficient condition for entanglement:

Proposition 14. *If $\rho_{AB} \notin$ PPT has negative partial transpose (NPT), then $\rho_{AB} \notin$ SEP.*

In small dimensions, the PPT criterion is actually necessary and sufficient:[8] If $|A||B| \leq 6$, then $\rho_{AB} \in \text{SEP} \leftrightarrow \rho_{AB} \in \text{PPT}$. However, already for $|A| = 3 = |B|$ there exist states that are PPT and entangled; these are so-called *bound entangled* states, since one cannot distill any entanglement from them.[9]

We now introduce a class of channels that can only produce PPT states.

Definition 15. A channel $\mathcal{N}\colon A \to B$ is called PPT if

$$(\mathrm{id}_R \otimes \mathcal{N})(\rho_{RA}) \tag{2.73}$$

is PPT for all ρ_{RA}.

[5] Peter W. Shor. Additivity of the classical capacity of entanglement-breaking quantum channels. *Journal of Mathematical Physics*, 43(9):4334–4340, 2002. DOI: 10.1063/1.1498000

[6] Salman Beigi and Peter W. Shor. On the complexity of computing zero-error and holevo capacity of quantum channels. *arXiv preprint*, 2007

[7] Leonid Gurvits. Classical deterministic complexity of edmonds' problem and quantum entanglement. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 10–19, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581136749. DOI: 10.1145/780542.780545. URL https://doi.org/10.1145/780542.780545

[8] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996. ISSN 0375-9601. DOI: 10.1016/S0375-9601(96)00706-2. URL http://www.sciencedirect.com/science/article/pii/S0375960196007062

[9] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Mixed-state entanglement and distillation: Is there a "bound" entanglement in nature? *Physical Review Letters*, 80:5239–5242, 6 1998. DOI: 10.1103/PhysRevLett.80.5239. URL https://link.aps.org/doi/10.1103/PhysRevLett.80.5239

For the following proposition, recall the transposition map $\vartheta\colon X \mapsto X^T$. Since $(\mathrm{id}_R \otimes \vartheta)(\gamma) = \mathbb{F} \not\geq 0$, this map is P but not CP.

Proposition 16. *The following are equivalent:*

a) $\mathcal{N}\colon A \to B$ *is PPT.*

b) $\tau_{AB}^{\mathcal{N}}$ *is PPT.*

c) $\vartheta \circ \mathcal{N}$ *is CP.*

Proof. a) $\implies$ b) $(\mathrm{id}_R \otimes \mathcal{N})(\rho_{RA})$ is PPT for all ρ_{RA} by definition, so in particular also for γ_{RA}.
b) $\implies$ c) If $\tau_{AB}^{\mathcal{N}}$ is PPT, then by definition

$$(\mathrm{id}_A \otimes \vartheta \circ \mathcal{N})(\gamma) = (\mathrm{id}_A \otimes \vartheta)(\tau_{AB}^{\mathcal{N}}) \tag{2.74}$$
$$= (\tau_{AB}^{\mathcal{N}})^{\Gamma_B} \geq 0. \tag{2.75}$$

c) $\implies$ a) If $\vartheta \circ \mathcal{N}$ is CP, then

$$(\mathrm{id}_R \otimes \vartheta \circ \mathcal{N})(\rho_{RA}) \geq 0 \tag{2.76}$$

for all ρ_{RA}, so $\mathcal{N}$ is PPT by definition. □

Note that $\vartheta \circ \mathcal{N}$ is not CP in general. However, $\vartheta \circ \mathcal{N} \circ \vartheta$ is CP for all CP maps $\mathcal{N}$.

What is the capacity of PPT channels? The Horodeckis showed that all PPT states are *undistillable*.[10] What do we mean by this?

[10] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Mixed-state entanglement and distillation: Is there a "bound" entanglement in nature? *Physical Review Letters*, 80:5239–5242, 6 1998. DOI: 10.1103/PhysRevLett.80.5239. URL https://link.aps.org/doi/10.1103/PhysRevLett.80.5239

Detour: Entanglement distillation

Given n independent and identically distributed (iid.) copies $\rho_{AB}^{\otimes n}$ of a bipartite state ρ_{AB}, the goal in entanglement distillation is to convert these to some (smaller) number m_n of maximally entangled states $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$. To achieve this task, Alice and Bob may use local operations and classical communication (LOCC). Local operations include any quantum operation and measurements that Alice and Bob can do in their respective labs. The (classical) measurement outcomes can be communicated to the other party, and there can be multiple rounds of communication in each direction.[11] If there is an LOCC protocol such that

1. the error[12] of the protocol goes to 0 as $n \to \infty$;
2. the protocol has a positive rate $c := \lim_{n\to\infty} \frac{m_n}{n} > 0$,

then ρ_{AB} is called *distillable*. The *distillable entanglement* $D(\rho_{AB})$ is defined as the supremum over all achievable rates in the sense defined above. Hence, a bipartite state is undistillable if $D(\rho_{AB}) = 0$, which

[11] Sometimes the communication is limited to only going from Alice to Bob. This set of operations is called *one-way LOCC*, or 1-LOCC. We call the more general class of LOCC operations with arbitrary communication *two-way LOCC*.

[12] This is typically a distance measure on quantum states such as the fidelity or the trace distance; in the asymptotic limit it doesn't matter too much which one is chosen, as they are all equivalent and lead to the same asymptotic quantity.

holds for all PPT states by the Horodecki result. One can show that this also implies that the quantum capacity vanishes for PPT channels, $Q(\mathcal{N}) = 0$. This holds even if we allow two-way classical communication.

The usual definition of quantum capacity $Q(\cdot)$ does not involve any classical communication between Alice and Bob. However, one can show that allowing for one-way classical communication from Alice to Bob does not increase the capacity.[13] If we further allow for two-way classical communication, then we get a different capacity called the *two-way quantum capacity* $Q_2(\cdot)$. Since one-way communication is a special case of this, we always have $Q(\mathcal{N}) \leq Q_2(\mathcal{N})$ for any channel $\mathcal{N}$. The Horodecki result shows that $Q(\mathcal{N}) = Q_2(\mathcal{N}) = 0$ for PPT channels. As a final remark, we note that the classical capacity of PPT channels is generally unknown.

[13] Howard Barnum, Emanuel Knill, and Michael A. Nielsen. On quantum fidelities and channel capacities. *IEEE Transactions on Information Theory*, 46(4):1317–1329, 2000. DOI: 10.1109/18.850671

2.5 *Antidegradable Channels*

Recall that a channel $\mathcal{N}\colon A \to B$ has an isometric representation $V\colon \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ such that $\mathcal{N}(\rho) = \mathrm{tr}_E\, V\rho V^\dagger$. This representation also yields a complementary channel $\mathcal{N}^c(\rho) = \mathrm{tr}_B\, V\rho V^\dagger$.

Definition 17. A channel $\mathcal{N}\colon A \to B$ is called antidegradable if there exists another channel $\mathcal{A}\colon E \to B$ such that

$$\mathcal{N} = \mathcal{A} \circ \mathcal{N}^c, \tag{2.77}$$

where $\mathcal{N}^c$ is the complementary channel.

Antidegradable channels allow Eve (the environment) to locally obtain Bob's output via the channel $\mathcal{A}$, which we call the *antidegrading map*. This implies that antidegradable channels cannot transmit quantum information, i.e., $Q(\mathcal{N}) = 0$: Assume otherwise, then Alice can faithfully send qubits to Bob at some positive rate. But there is some protocol based on the channel $\mathcal{A}$ that allows Eve to implement the same protocol that Alice and Bob use. This would clone Alice's state to the systems B and E, thus violating the no-cloning principle:[14]

[14] William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982

No-cloning theorem There is no unitary U such that for a fixed state $|0\rangle$ we have

$$U(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle \tag{2.78}$$

for all $|\psi\rangle$.

Examples of antidegradable channels:

1. Erasure channel $\mathcal{E}_p\colon \rho \mapsto (1-p)\rho + p\operatorname{tr}(\rho)\left|e\right\rangle\left\langle e\right|$ for $p \geq \frac{1}{2}$.
2. Amplitude dampening channel $\mathcal{A}_\gamma$ for $\gamma \geq \frac{1}{2}$.
3. Depolarizing channel $\mathcal{D}_p$ for $p \geq \frac{1}{4}$.

We prove the first example here:

Recall that $\mathcal{E}_p\colon B(\mathcal{H}_1) \to B(\mathcal{H}_1 \oplus \mathcal{H}_2)$, where $\mathcal{H}_1 = \mathbb{C}^2$ is the input space and $\mathcal{H}_2 = \mathbb{C}$ is the erasure flag. To highlight the structure of the output space, we write the erasure channel as

$$\mathcal{E}_p(\rho) = (1-p)\tilde{\rho} + p\operatorname{tr}(\rho)\left|e\right\rangle\left\langle e\right|, \tag{2.79}$$

where

$$\tilde{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} & 0 \\ \rho_{10} & \rho_{11} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \left|e\right\rangle\left\langle e\right| = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.80}$$

with $\rho = \left(\begin{smallmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{smallmatrix}\right)$ It is easy to verify that the complementary channel is given by the following formula:

$$\mathcal{E}_p^c(\rho) = p\tilde{\rho} + (1-p)\operatorname{tr}(\rho)\left|e\right\rangle\left\langle e\right|. \tag{2.81}$$

We now explicitly construct an antidegrading map $\mathcal{A}_p$ Roughly speaking, what we would like to do is apply an erasure channel with probability $q = \frac{2p-1}{p}$ to $\tilde{\rho}$ and do nothing if an erasure already occurred. Note that $q \geq 0$ precisely when $p \geq 1/2$.

$$\begin{aligned} \mathcal{E}_q(p\rho) &= p(1-q)\tilde{\rho} + pq\operatorname{tr}(\rho)\left|e\right\rangle\left\langle e\right| && (2.82)\\ &= (1-p)\tilde{\rho} + (2p-1)\operatorname{tr}(\rho)\left|e\right\rangle\left\langle e\right|. && (2.83) \end{aligned}$$

Combining this with the contribution from the identity channel acting on the erasure flag, $(1-p)\left|e\right\rangle\left\langle e\right|$, gives us

$$(1-p)\tilde{\rho} + p\operatorname{tr}(\rho)\left|e\right\rangle\left\langle e\right| = \mathcal{E}_p(\rho). \tag{2.84}$$

To make this work, we observe that the output of the complementary channel is in $B(\mathcal{H}_1 \oplus \mathcal{H}_2)$. We thus need to define $\mathcal{A}_p$ as acting on $B(\mathcal{H}_1 \oplus \mathcal{H}_2)$, which we achieve by suitably modifying the Kraus operators (2.24) of the erasure channel as follows:

$$K_1 = \sqrt{1-q}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad K_2 = \sqrt{q}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{2.85}$$

$$K_3 = \sqrt{q}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad K_4 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.86}$$

The operators K_1, K_2, K_3 enact erasure on $B(\mathcal{H}_1 \oplus 0)$ with probability q, while K_4 leaves operators in $B(0 \oplus \mathcal{H}_2)$ untouched. Defining $\mathcal{A}_p = \sum_{i=1}^4 K_i \cdot K_i^\dagger$, we obtain $\mathcal{E}_p = \mathcal{A} \circ \mathcal{E}_p^c$ for $p \geq \frac{1}{2}$.

Proposition 18. *A channel $\mathcal{N}\colon A \to B$ is antidegradable if and only if the Choi operator $\tau_{AB}^{\mathcal{N}}$ has a 2-extension, that is, there is a mixed state $\sigma_{ABB'}$ (with $B \cong B'$) satisfying*

$$\operatorname{tr}_B \sigma_{ABB'} = \operatorname{tr}_{B'} \sigma_{ABB'} = \tau_{AB}^{\mathcal{N}}. \tag{2.87}$$

Proof. "$\Longrightarrow$" We will construct a $\sigma_{ABB'}$ that satisfies the desired properties. Let $V\colon \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ be such that $\mathcal{N}(\cdot) = \operatorname{tr}_E V \cdot V^\dagger$. Consider a purification of the Choi operator, $|\chi\rangle_{ABE} = (\mathbb{1}_A \otimes V)\,|\gamma\rangle_{AA'}$, and define $\sigma_{ABB'} = (\mathrm{id}_{AB} \otimes \mathcal{A})(\chi_{ABE})$, where $\mathcal{A}\colon E \to B'$ is the antidegrading map. Then we have $\operatorname{tr}_B \sigma_{ABB'} = \operatorname{tr}_{B'} \sigma_{ABB'} = \tau_{AB}^{\mathcal{N}}$.

"$\Longleftarrow$" Let $\tau_{ABB'}$ be an extension of $\tau_{AB}^{\mathcal{N}}$ satisfying $\operatorname{tr}_B \sigma_{ABB'} = \operatorname{tr}_{B'} \sigma_{ABB'} = \tau_{AB}^{\mathcal{N}}$, and let $|\psi\rangle_{ABB'R}$ be a purification of $\tau_{ABB'}$. Then (1.103) shows that $\psi_{AB'R} = \operatorname{tr}_B \psi_{ABB'R}$ is the Choi operator of $\mathcal{N}^c$. Define the antidegrading map simply as $\mathcal{A} = \operatorname{tr}_R$. Thus, $\mathcal{N} = \mathcal{A} \circ \mathcal{N}^c$ because $\operatorname{tr}_R \psi_{AB'R} = \psi_{AB'} = \psi_{AB} = \tau_{AB}^{\mathcal{N}}$. □

Note that without loss of generality the extension $\sigma_{ABB'}$ can be taken to be symmetric under $B \leftrightarrow B'$,

$$\mathbb{F}_{BB'} \sigma_{ABB'} \mathbb{F}_{BB'} = \sigma_{ABB'}. \tag{2.88}$$

Corollary 19. *Let $\mathcal{N}_1\colon A \to B_1$ be antidegradable and $\mathcal{N}_2\colon B_1 \to B_2$ be an arbitrary channel. Then $\mathcal{M} = \mathcal{N}_2 \circ \mathcal{N}_1$ is also antidegradable.*

Proof. Let $\sigma_{AB_1B_1'}$ be an extension of the Choi operator $\tau_{AB_1}^{\mathcal{N}_1}$ of $\mathcal{N}_1$. Then

$$\omega_{AB_2B_2'} = (\mathrm{id}_A \otimes \mathcal{N}_2 \otimes \mathcal{N}_2)(\sigma_{AB_1B_1'}) \tag{2.89}$$

is an extension of $\tau_{AB_2}^{\mathcal{M}} = (\mathrm{id} \otimes \mathcal{N}_2)(\tau_{AB_1}^{\mathcal{N}_1})$. By Prop. 18, $\mathcal{M}$ is antidegradable. □

We will now use this to prove that the depolarizing channel $\mathcal{D}_p$ is antidegradable for $p \geq \frac{1}{4}$. First, recall that the depolarizing channel takes the form

$$\mathcal{D}_p\colon \rho \mapsto (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z), \tag{2.90}$$

and can equivalently be written as

$$\tilde{\mathcal{D}}_q\colon \rho \mapsto (1-q)\rho + q \operatorname{tr}(\rho)\frac{1}{2}\mathbb{1}, \tag{2.91}$$

where $q = \frac{4p}{3}$. The proof continues by considering the following cases:

1. $p = \frac{1}{4}$, $q = \frac{1}{3}$: consider the two states

$$|\psi_1\rangle = \frac{1}{\sqrt{6}}(2|000\rangle + |101\rangle + |110\rangle), \tag{2.92}$$

$$|\psi_2\rangle = \frac{1}{\sqrt{6}}(|001\rangle + |010\rangle + 2|111\rangle). \tag{2.93}$$

It can be verified that $\sigma_{ABB'} = \psi_1 + \psi_2$ is a symmetric extension of $\tau_{AB}^{\mathcal{D}_{1/4}}$, and thus $\tilde{\mathcal{D}}_{\frac{1}{3}}$ is antidegradable.

2. $q_1 \leq q_2 \leq 1$: We have $\tilde{\mathcal{D}}_{q_2} = \tilde{\mathcal{D}}_w \circ \tilde{\mathcal{D}}_{q_1}$, where $w = \frac{q_2 - q_1}{1 - q_1}$. Corollary 19 now shows that $\tilde{\mathcal{D}}_q$ is antidegradable for $\frac{1}{3} \leq q \leq 1$.

3. $1 \leq q \leq \frac{4}{3}$, $\frac{3}{4} \leq p \leq 1$: We will prove a stronger statement here. We claim that $\mathcal{D}_p$ is entanglement-breaking for all $p \geq \frac{1}{2}$. Recall that $\mathcal{D}_p$ is entanglement-breaking if and only if the Choi operator $\tau^{\mathcal{D}_p}$ is separable. Because $|A| = |B| = 2$, separability is equivalent to PPT.[15] We work with the Choi operator of $\tilde{\mathcal{D}}_q$,

$$\tau_{AB}^{\tilde{\mathcal{D}}_q} = (1 - q)|\gamma\rangle\langle\gamma|_{AB} + q\frac{1}{2}\mathbb{1} \otimes \mathbb{1}. \tag{2.94}$$

Recall that we have

$$\mathbb{F}_{AB} = \sum_{i,j} |i\rangle\langle j| \otimes |j\rangle \otimes i = \sum_{i,j} |i\rangle\langle j| \otimes |i\rangle\langle j|^T = |\gamma\rangle\langle\gamma|^{\Gamma_B}, \tag{2.95}$$

where Γ_B denotes partial transposition on the B system. Applying this to the Choi operator yields

$$(\tau^{\tilde{\mathcal{D}}_q})^{\Gamma_B} = (1 - q)\mathbb{F}_{AB} + \frac{q}{2}\mathbb{1} \otimes \mathbb{1}. \tag{2.96}$$

The two operators $\mathbb{F}_{AB}$ and $\mathbb{1} \otimes \mathbb{1}$ commute and so they are simultaneously diagonalizable. Hence, in order to check positive semidefiniteness of $(\tau^{\tilde{\mathcal{D}}_q})^{\Gamma_B}$ it suffices to consider the sums of the eigenvalues of each. Recall that $\mathbb{F}_{AB}$ has a single negative eigenvalue -1 (corresponding to the eigenvector $|01\rangle - |10\rangle$).

If $q \leq 1$, then the smallest eigenvalue of $(\tau^{\tilde{\mathcal{D}}_p})^{\Gamma_B}$ is equal to $-(1 - q) + \frac{q}{2}$, which is nonnegative iff $q \geq \frac{2}{3}$.

If $q \geq 1$, this smallest eigenvalue is equal to $(1 - q) + \frac{q}{2}$, which is nonnegative iff $q \leq \frac{4}{3}$.

In summary, $\tilde{\mathcal{D}}_q$ is PPT (and entanglement breaking) for $\frac{2}{3} \leq q \leq \frac{4}{3}$. To finish our proof, we need one more lemma:

Lemma 20. *Entanglement-breaking channels are antidegradable.*

Proof. An entanglement-breaking channel $\mathcal{N}$ has a separable Choi operator $\tau_{AB}^{\mathcal{N}} = d\sum_i p_i \omega_A^i \otimes \sigma_B^i$. Then $\varphi_{ABB'} = \sum_i p_i \omega_A^i \otimes \sigma_B^i \otimes \sigma_{B'}^i$ is a (symmetric) extension of $\frac{1}{d}\tau_{AB}^{\mathcal{N}}$. By Prop. 18, $\mathcal{N}$ is antidegradable. □

[15] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996. ISSN 0375-9601. DOI: 10.1016/S0375-9601(96)00706-2. URL http://www.sciencedirect.com/science/article/pii/S0375960196007062

More generally, one can define k-extensions of a bipartite state as follows. A state ρ_{AB} is (symmetrically) k-extendible if there exists $\sigma_{AB_1\ldots B_k}$ such that

1. $\mathrm{tr}_{B_i^c}\, \sigma_{AB_1\ldots B_k} = \rho_{AB}$ for all $i = 1, \ldots, k$, where $\mathrm{tr}_{B_i^c}$ denotes the partial trace over all B systems except B_i.

2. $P_\pi \sigma_{AB_1\ldots B_k} P_\pi^\dagger = \sigma_{AB_1\ldots B_k}$, where $\pi \in S_k$ is a permutation and P_π is the representation of S_k on $\mathcal{H}_B^{\otimes k}$ acting by permuting tensor factors.

One can show that a state is ∞-extendible if and only if it is separable. This creates a type of hierarchy: if a state is k-extendible it is also l-extendible for $l \leq k$, by simply tracing out the extra parties from the k-extension.

2.6 *Degradable channels*

Degradable channels are the dual concept to anti-degradable channels.

Definition 21. A channel $\mathcal{N} : A \to B$ with complementary channel $\mathcal{N}^c : A \to E$ is called degradable, if there exists a channel $\mathcal{D} : B \to E$ (called the degrading map) such that $\mathcal{N}^c = \mathcal{D} \circ \mathcal{N}$.

It follows that $\mathcal{N}$ is degradable iff $\mathcal{N}^c$ is antidegradable. The intuition for degradability is that Bob can locally "simulate" the environment. As a result, we understand the quantum capacity of degradable channels and can efficiently compute it. Further, we know that there *are* degradable channels $\mathcal{N}$ such that $Q(\mathcal{N}) > 0$.
Examples:

- Erasure channel[16] $\mathcal{E}_p$ for $p \leq \frac{1}{2}$.
- Amplitude damping channel A_γ for $\gamma \leq \frac{1}{2}$.
- generalized dephasing channels
- the complementary channel of any entanglement-breaking channel, so-called "Hadamard channels'.'

[16] We have already proved this! The quantum capacity is given as $Q(\mathcal{E}_p) = 1 - 2p$

We now list some fundamental differences between the sets DEG and ADG of degradable and antidegradable channels, respectively.

- $\mathcal{N} \in \mathrm{DEG} \implies Q(\mathcal{N}) \geq 0, \mathcal{N} \in \mathrm{ADG} \implies Q(\mathcal{N}) = 0.$
- DEG is not convex, whereas ADG is convex.[17]

[17] That is, if $\mathcal{N}_1, \mathcal{N}_2 \in \mathrm{ADG}$, then $\lambda\mathcal{N}_1 + (1-\lambda)\mathcal{N}_2 \in \mathrm{ADG}$ as well.

Now, why isn't the set of degradable channels convex? Suppose $\mathcal{N}_1 \in$ DEG and has a Choi operator ρ_{AB}. Similarly, $\mathcal{N}_2 \in$ DEG and has a Choi operator σ_{AB}. And let

$$\mathcal{N} = \lambda\mathcal{N}_1 + (1-\lambda)\mathcal{N}_2 \tag{2.97}$$

be a channel with Choi operator ω_{AB}. Then, suppose that $|\psi^\rho\rangle_{ABE}$ purifies ρ_{AB} and $|\varphi^\sigma\rangle_{ABE}$ purifies σ_{AB}. Then

$$|\chi^\omega\rangle_{ABEX} = \sqrt{\lambda}\,|\psi^\rho\rangle_{ABE}\,|0\rangle_X + \sqrt{1-\lambda}\,|\varphi^\sigma\rangle_{ABE}\,|1\rangle_X \tag{2.98}$$

purifies ω_{AB} and $\mathrm{tr}_{EX}\,\chi^\omega = \omega_{AB}$. The choi operator of $\mathcal{N}^c$ is given as

$$\omega_{AEX} = \begin{pmatrix} \lambda\rho_{AE} & \sqrt{\lambda(1-\lambda)}\,\mathrm{tr}_B\,|\psi^\rho\rangle\langle\varphi^\sigma|_{ABE} \\ \text{h.c.} & (1-\lambda)\sigma_{AE} \end{pmatrix} \tag{2.99}$$

Proposition 22. *A qubit-qubit channel with qubit environment (i.e., $|A| = |B| = |E| = 2$) is degradable or antidegradable.*

The proof idea is as follows.[18]. Suppose $\mathcal{N}$ is degradable. Then, there exists $\mathcal{D} : B \to E$ such that $\mathcal{N}^c = \mathcal{D} \circ \mathcal{N}$. Using transfer matrices, we can write

$$N^c = DN \Leftrightarrow D = N^c N^{-1}, \tag{2.100}$$

Then, we note that degradability is equivalent to complete positivity of $\mathcal{D}$ defined via $D = N^c N^{-1}$.

Similarly for $\mathcal{N}$ ADG, there exists $A : E \to B$ such that $\mathcal{A} \circ \mathcal{N}^c$. Using transfer matrices, we have

$$N = AN^c \Leftrightarrow A = N(N^c)^{-1}. \tag{2.101}$$

And antidegradability is equivalent to complete positivity of $\mathcal{A}$ defined via $A = N(N^c)^{-1}$. For $|\mathrm{A}| = |\mathrm{B}| = |\mathrm{E}| = 2$, we always have that D or D^{-1} defines a completely positive map.

[18] Michael M Wolf and David Perez-Garcia. Quantum capacities of channels with small environment. *Physical Review A*, 75(1):012303, 2007

3
Covariant channels and minimum output entropy

3.1 Definition and Basics of Representation Theory

Motivation

Consider a channel $\mathcal{N} : A \to B$. For any unitaries U, V we can construct another channel unitarily equivalent to $\mathcal{N}$:

$$\mathcal{M}(X) = U\mathcal{N}(VXV^\dagger)U^\dagger. \tag{3.1}$$

$\mathcal{M}$ and $\mathcal{N}$ have the same capacities and thus, from the perspective of information theory, they are effectively the same channel. Any protocol for $\mathcal{N}$, such as generating entanglement, can be turned into a protocol for $\mathcal{M}$ achieving the same task at the same rate by absorbing U, V into the protocol.

For certain channels and unitaries, we may have that $\mathcal{M} = \mathcal{N}$. Thus, (U, V) is a symmetry of $\mathcal{N}$. Channels with such symmetries are called *covariant*.

Basics of Representation Theory

Here we will review some of the fundamental tenets of representation theory. For our purposes, we are only considering groups G that are either finite or compact (such as the unitary group).

1. A representation of G on a vector space R is a group homomorphism

$$\varphi : G \to GL(R) \tag{3.2}$$

   Recall that homomorphisms preserve group structure, i.e. $\varphi(g \cdot h) = \varphi(g) \cdot \varphi(h)\ \forall g, h \in G$. We say that the representation (φ, R) is finite if $\dim(R) < \infty$. (Note that we will only consider representations over complex vector spaces.) We say that (φ, R) is a unitary representation if R is a Hilbert space and $\varphi(g) \in U(R)\ \forall g \in G$.

2. A subspace $S \subseteq R$ is called G-invariant if $\varphi(g)s \in S \; \forall s \in S, \; \forall g \in G$. The subspaces $\{0\}$ and R are G-invariant for any (φ, R) and are sometimes referred to as the trivial invariant subspaces.

3. A representation (φ, R) is called *irreducible* if the only G-invariant subspaces are the trivial ones ($\{0\}$ and R). We denote such a representation with the shorthand, "irrep".

4. Finite-dimensional unitary representations of any group are completely reducible $\varphi = \bigoplus_i \varphi_i$, where φ_i is irreducible.

5. Let (φ_V, V) and (φ_W, W) be representations of a group G. A linear map $f : V \to W$ is called G-linear if $f \circ \varphi_V(g) = \varphi_W \circ f$ for all $g \in G$ (in other words, f commutes with the representations).

The following Lemma is extremely useful (offered without proof, which can be found in introductory texts on representation theory).

Schur's Lemma: Let φ_V and φ_W be irreps of a group G and let f be a G-linear map. Then either

1. $V \ncong W$ and f is the trivial map $f = 0$.
2. $V \cong W$ and $f = \lambda \mathrm{id}_{V \to W}$, for some $\lambda \in \mathbb{C}$.

Covariant Channels

With these concepts in our toolbox, we will now explicitly define covariant channels:

Definition 23. Let $\mathcal{N} : A \to B$ be a quantum channel and G 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

$$V_g \mathcal{N}(\cdot) V_g^\dagger = \mathcal{N}(U_g \cdot U_g^\dagger) \quad \text{for all } g \in G. \tag{3.3}$$

The following are examples of covariant channels.

1. Pauli channels $\rho \mapsto p_0\rho + p_1 X\rho X + p_2 Y\rho Y + p_3 Z\rho Z$ are covariant with respect to the Pauli group

$$P = \{\pm 1, \pm i\} \cup \{\mathbb{1}, X, Y, Z\}. \tag{3.4}$$

This is easy to see: let $o_1, o_2 \in P$ and note that $o_1 o_2 \cdot o_2^\dagger o_1^\dagger = o_2 o_1 \cdot o_1^\dagger o_2^\dagger$.[1]

[1] This is a simple consequence of $\sigma_i \sigma_j = \pm i \sigma_k$ for distinct $\sigma_i, \sigma_j, \sigma_k \in P$.

2. For special cases of Pauli groups there are additional symmetries. Consider the depolarizing channel $\rho \mapsto (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y +$

$Z\rho Z$). Here the covariance group is the entire unitary group $U(2)$. This is easy to see in the "q" representation: $\rho \mapsto (1-q)\rho + q\,\mathrm{tr}(\rho)\frac{1}{2}\mathbb{1}$,[2]

[2] Recall that $q = \frac{4p}{3}$.

$$\mathcal{D}_q(U\rho U^\dagger) = (1-q)U\rho U^\dagger + q\,\mathrm{tr}(\rho)\frac{1}{2}\mathbb{1} \tag{3.5}$$

$$= UD_q(\rho)U^\dagger. \tag{3.6}$$

Thus far we have only talked about qubit depolarizing channels. But it is straightforward to generalize this channel to $d \geq 2$:

$$\rho \mapsto (1-q)\rho + q\,\mathrm{tr}(\rho)\frac{1}{d}\mathbb{1}. \tag{3.7}$$

This has the covariance group $\mathcal{U}_d$. We will shortly see that any channel covariant under the full unitary must be a depolarizing channel.

3. Amplitude damping channel A_γ with Kraus operators

$$K_0 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} \quad K_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}. \tag{3.8}$$

This has the small covariance group $\{\mathbb{1}, Z\} \cong \mathbb{Z}_2$.

4. Erasure channel $\rho \mapsto (1-p)\rho + p\,\mathrm{tr}(\rho)\,|e\rangle\langle e|$. This has the full unitary group $U(2)$ as a covariance group. Here, however, we have to specify the input/output representations:

$$\begin{aligned} &\text{rep. on input space: } U \mapsto U \\ &\text{rep. on output space: } U \mapsto \begin{pmatrix} U & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \tag{3.9}$$

Note that the representation on the output space is not irreducible as $|e\rangle\langle e|$ is a non-trivial invariant subspace. We can generalize the erasure channel in a natural way to d dimensions, and it is clear that the resulting channel will be covariant with respect to the unitary group $\mathcal{U}_d$ and its representations given in (3.9).

Proposition 24. *A channel $\mathcal{N}\colon A \to B$ is (G, U_g, V_g)-covariant iff $\tau^{\mathcal{N}}_{AB} = (\overline{U}_g \otimes V_g)\tau^{\mathcal{N}}_{AB}(\overline{U}_g \otimes V_g)^\dagger$ for all $g \in G$.*

Proof. "$\Longrightarrow$" By assumption, $V_g\mathcal{N}(\cdot)V_g^\dagger = \mathcal{N}(U_g \cdot U_g^\dagger)$ for all $g \in G$, which is equivalent to $\mathcal{N}(\cdot) = V_g^\dagger\mathcal{N}(U_g \cdot U_g^\dagger)V_g^\dagger$. We then have

$$\tau^{\mathcal{N}}_{AB} = (\mathbb{1} \otimes \mathcal{N})(\gamma) \tag{3.10}$$

$$= (\mathbb{1} \otimes V_g^\dagger)(\mathbb{1} \otimes \mathcal{N})((\mathbb{1} \otimes U_g)\gamma(\mathbb{1} \otimes U_g)^\dagger)(\mathbb{1} \otimes V_g) \tag{3.11}$$

$$= (\mathbb{1} \otimes V_g^\dagger)(\mathbb{1} \otimes \mathcal{N})((U_g^T \otimes \mathbb{1})\gamma(U_g^T \otimes \mathbb{1})^\dagger)(\mathbb{1} \otimes V_g) \tag{3.12}$$

$$= (U_g^T \otimes V_g^\dagger)(\mathbb{1} \otimes \mathcal{N})(\gamma)(\overline{U}_g \otimes V_g) \tag{3.13}$$

$$= (U_g^T \otimes V_g^\dagger)\tau^{\mathcal{N}}_{AB}(\overline{U}_g \otimes V_g) \tag{3.14}$$

for all $g \in G$, where we used the "transpose trick"

$$(X \otimes \mathbb{1})\gamma = (\mathbb{1} \otimes X^T)\gamma, \tag{3.15}$$

and $\overline{U}_g$ denotes the entry-wise complex conjugate of U_g. To get the form in the proposition statement we consider g^{-1} and note that $U_{g^{-1}} = (U_g)^{-1} = U_g^\dagger$, since the considered representation is unitary. Thus,

$$\tau_{AB}^{\mathcal{N}} = (\overline{U}_g \otimes V_g)\tau_{AB}^{\mathcal{N}}(\overline{U}_g \otimes V_g)^\dagger \quad \text{for all } g \in G. \tag{3.16}$$

"$\Longleftarrow$" We have $\overline{U}_g U_g^T = \mathbb{1}$, and thus $\tau_{AB}^{\mathcal{N}} = (\overline{U}_g \otimes V_g)\tau_{AB}^{\mathcal{N}}(\overline{U}_g \otimes V_g)^\dagger$ for all $g \in G$ is equivalent to

$$(U_g^T \otimes \mathbb{1})\tau_{AB}^{\mathcal{N}}(U_g^T \otimes \mathbb{1})^\dagger = (\mathbb{1} \otimes V_g)\tau_{AB}^{\mathcal{N}}(\mathbb{1} \otimes V_g)^\dagger. \tag{3.17}$$

Recall from the Choi-isomorphism that $\mathcal{N}(X) = \operatorname{tr}_1(\tau_{AB}^{\mathcal{N}}(X^T \otimes \mathbb{1}))$. Then for all $g \in G$ we have

$$\begin{aligned}
\mathcal{N}(U_g X U_g^\dagger) &= \operatorname{tr}_1\left(\tau_{AB}^{\mathcal{N}}(\overline{U}_g X^T U_g^T \otimes \mathbb{1})\right) && (3.18)\\
&= \operatorname{tr}_1\left((U_g^T \otimes \mathbb{1})\tau_{AB}^{\mathcal{N}}(U_g^T \otimes \mathbb{1})^\dagger (X^T \otimes \mathbb{1})\right) && (3.19)\\
&= \operatorname{tr}_1\left((\mathbb{1} \otimes V_g)\tau_{AB}^{\mathcal{N}}(\mathbb{1} \otimes V_g)^\dagger (X^T \otimes \mathbb{1})\right) && (3.20)\\
&= V_g \operatorname{tr}_1\left(\tau_{AB}^{\mathcal{N}}(X^T \otimes \mathbb{1})\right) V_g^\dagger && (3.21)\\
&= V_g \mathcal{N}(X) V_g^\dagger, && (3.22)
\end{aligned}$$

where we used cyclicity of the trace in (3.19) and (3.17) in (3.20). □

A slight problem with Prop 24 is that the Choi operator and its covariance properties are basis-dependent. This can be fixed by instead working with the Jamiołkowski operator:

$$J_{AB}^{\mathcal{N}} = (\mathrm{id} \otimes \mathcal{N})(\mathbb{F}_{AA'}). \tag{3.23}$$

Since $\mathbb{F} = |\gamma\rangle\langle\gamma|^{T_A}$, this is related to the Choi operator by the partial transpose: $(J_{AB}^{\mathcal{N}})^{T_A} = \tau_{AB}^{\mathcal{N}}$. In terms of the Jamiołkowski operator, the (G, U_g, V_g)-covariance of $\mathcal{N}$ is equivalent to

$$J_{AB}^{\mathcal{N}} = (U_g \otimes V_g)J_{AB}^{\mathcal{N}}(U_g \otimes V_g)^\dagger \quad \text{for all } g \in G. \tag{3.24}$$

Now recall that the d-dimensional depolarizing channel $\rho \mapsto (1-q)\rho + q \operatorname{tr}\rho \frac{1}{d}\mathbb{1}_d$ has $\mathcal{U}_d$ as a covariance group.

Proposition 25. *Let $\mathcal{N} : A \to B$ be a channel with $|A| = |B| = d$. If $U\mathcal{N}(\cdot)U^\dagger = \mathcal{N}(U \cdot U^\dagger)$ for all $U \in \mathcal{U}_d$, then*

$$\mathcal{N}(X) = (1-q)X + q \operatorname{tr}(X)\frac{1}{d}\mathbb{1}_d, \tag{3.25}$$

where $q = \frac{f-d^2}{1-d^2}$ *with* $f = \langle\gamma|\tau_{AB}^{\mathcal{N}}|\gamma\rangle$.

Proof. By our remark after Prop. 24 we know that the $(\mathcal{U}_d, U, U)$ covariance of $\mathcal{N}$ implies that the Jamiołkowski operator $J_{AB}^{\mathcal{N}} = (\mathrm{id} \otimes \mathcal{N})(\mathbb{F}_{AA'})$ satisfies $J_{AB}^{\mathcal{N}} = (U \otimes U) J_{AB}^{\mathcal{N}} (U \otimes U)^\dagger$ for all $U \in \mathcal{U}_d$. We will prove in Lemma 26 below that this implies $J_{AB}^{\mathcal{N}} = x\mathbb{1}_A \otimes \mathbb{1}_B + y\mathbb{F}_{AB}$ for some $x, y \in \mathbb{R}$.[3] Then,

$$\tau_{AB}^{\mathcal{N}} = (J_{AB}^{\mathcal{N}})^{T_A} \tag{3.26}$$
$$= x\mathbb{1}_A^T \otimes \mathbb{1}_B + y\mathbb{F}_{AB}^{T_A} \tag{3.27}$$
$$= x\mathbb{1}_A \otimes \mathbb{1}_B + y\,|\gamma\rangle\langle\gamma|\,. \tag{3.28}$$

As the action of a depolarizing channel is $X \mapsto (1-q)X + q\,\mathrm{tr}(X)\frac{1}{d}\mathbb{1}_d$, it is clear that this is the Choi operator for a depolarizing channel. In general we require that the Choi operator for a CPTP channel satisfy $\mathrm{tr}(\tau_{AB}^{\mathcal{N}}) = |A| = d$. Applied to the form of the Choi operator given above and using $\mathrm{tr}\,\gamma = \langle\gamma|\gamma\rangle = d$, this yields $1 = xd + y$.

Next, we define

$$f = \langle\gamma|\tau_{AB}^{\mathcal{N}}|\gamma\rangle = xd + yd^2. \tag{3.29}$$

Combining, we have that $f = xd + yd^2 = 1 - y + yd^2$. This implies that $y = \frac{1-f}{1-d^2}$ and $1 - y = \frac{f-d^2}{1-d^2}$. Comparing this with the Choi operator of the depolarizing channel reveals after a short calculation that $q = 1 - y$, so that $q = \frac{f-d^2}{1-d^2}$. □

[3] Note that $\mathbb{F}$ is invariant under $U \otimes U$ since for all $|\psi\rangle, |\phi\rangle$ we have

$$(U \otimes U)\mathbb{F}(|\psi\rangle \otimes |\phi\rangle) = U|\phi\rangle \otimes U|\psi\rangle$$
$$= \mathbb{F}(U|\psi\rangle \otimes U|\phi\rangle),$$

so that $[U \otimes U, \mathbb{F}] = 0$.

Above we noted that $J_{AB}^{\mathcal{N}}$ is invariant when acted upon by the representation $U \otimes U$. These are a special class of states called **Werner states** with a particularly nice structure.

Lemma 26. *Let* $R \in B(\mathbb{C}^d \otimes \mathbb{C}^d)$ *be such that*

$$(U \otimes U)R(U \otimes U)^\dagger = R \quad \textit{for all } U \in \mathcal{U}_d. \tag{3.30}$$

Then there are $a, b \in \mathbb{C}$ *such that* $R = a\mathbb{1}_A \otimes \mathbb{1}_B + b\mathbb{F}_{AB}$.

Proof. There are two ways to prove this:

1. Elegant way via Schur-Weyl duality:

   On $\mathcal{H} = (\mathbb{C}^d)^{\otimes 2}$ there are natural representations of the unitary group U_d and symmetric group S_2:

$$\varphi_{U_d} : U \mapsto U^{\otimes 2}, \tag{3.31}$$
$$\varphi_{S_2} : \pi \mapsto (|\psi_1\rangle \otimes |\psi_2\rangle \mapsto |\psi_{\pi^{-1}(1)}\rangle \otimes |\psi_{\pi^{-1}(2)}\rangle). \tag{3.32}$$

These representations clearly commute as $[U \otimes U, \mathbb{F}] = 0$. We denote the commutant of a subalgebra $A \subseteq \mathcal{C}$ by

$$A' = \{b \in \mathcal{C} : ab = ba \; \forall a \in A\}. \tag{3.33}$$

Let $A = \langle \varphi_{U_d}(U) \rangle$ and $B = \langle \varphi_{S_2}(\pi) \rangle = \langle \mathbb{1}, \mathbb{F} \rangle$ be the algebras generated by the above representations of the unitary and symmetric group. **Schur-Weyl duality** states that $A' = B$ and $B' = A$. That is, the representation (3.31) of the unitary group spans the commutant of the representation (3.32) of the symmetric group, and vice versa.[4] Returning to the lemma: $(U \otimes U)R(U \otimes U)^\dagger = R$ for all $U \in U_d$ implies that $R \in A' = B$. Thus, $R = x\mathbb{1} \otimes \mathbb{1} + y\mathbb{F}$, for some $x, y \in \mathbb{C}$.

[4] Schur-Weyl duality holds more generally for representations on $(\mathbb{C}^d)^{\otimes n}$. That is, the representation of the symmetric group S_n acting by permuting tensor factors on $(\mathbb{C}^d)^{\otimes n}$ spans the commutant of the representation of $\mathcal{U}_d$ acting via $U^{\otimes n}$, and vice versa.

2. We will now work through a direct proof for the case of $d = 2$. Consider the Bell basis for $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \tag{3.34}$$

$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \tag{3.35}$$

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \tag{3.36}$$

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \tag{3.37}$$

We define the symmetric subspace $\mathcal{H}_s$ and antisymmetric subspace $\mathcal{H}_a$ of $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$ as follows:

$$\mathcal{H}_s = \{|\psi\rangle \in \mathcal{H} : \mathbb{F}\,|\psi\rangle = |\psi\rangle\} \tag{3.38}$$

$$\mathcal{H}_a = \{|\psi\rangle \in \mathcal{H} : \mathbb{F}\,|\psi\rangle = -|\psi\rangle\}. \tag{3.39}$$

It's easy to see that $|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle \in \mathcal{H}_s$ and $|\Psi^-\rangle \in \mathcal{H}_a$. Thus, $\mathcal{H} = \mathcal{H}_s \oplus \mathcal{H}_a$. We define projectors $P_s = \frac{1}{2}(\mathbb{1} + \mathbb{F})$ and $P_a = \frac{1}{2}(\mathbb{1} - \mathbb{F})$ such that $P_{s/a}\mathcal{H} = \mathcal{H}_{s/a}$. We can express R in this basis:

$$R = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \begin{matrix} \Phi^+ \\ \Phi^- \\ \Psi^+ \\ \Psi^- \end{matrix} \tag{3.40}$$

$$\Phi^+ \; \Phi^- \; \Psi^+ \; \Psi^- \tag{3.41}$$

We now continue by choosing specific U and showing how it restricts R:

(a) $X \otimes X$:

$$|\Phi^+\rangle \mapsto |\Phi^+\rangle, \tag{3.42}$$
$$|\Phi^-\rangle \mapsto -|\Phi^-\rangle, \tag{3.43}$$
$$|\Psi^+\rangle \mapsto |\Psi^+\rangle, \tag{3.44}$$
$$|\Psi^-\rangle \mapsto -|\Psi^-\rangle. \tag{3.45}$$

(b) $Z \otimes Z$:

$$|\Phi^+\rangle \mapsto |\Phi^+\rangle, \tag{3.46}$$
$$|\Phi^-\rangle \mapsto |\Phi^-\rangle, \tag{3.47}$$
$$|\Psi^+\rangle \mapsto -|\Psi^+\rangle, \tag{3.48}$$
$$|\Psi^-\rangle \mapsto -|\Psi^-\rangle. \tag{3.49}$$

(c) $H \otimes H$:[5]

$$|\Phi^+\rangle \mapsto |\Phi^+\rangle, \tag{3.50}$$
$$|\Phi^-\rangle \mapsto |\Psi^+\rangle, \tag{3.51}$$
$$|\Psi^+\rangle \mapsto |\Phi^-\rangle, \tag{3.52}$$
$$|\Psi^-\rangle \mapsto -|\Psi^-\rangle. \tag{3.53}$$

[5] H is the Hadamard matrix $H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

(d) $S \otimes S$:[6]

$$|\Phi^+\rangle \mapsto |\Phi^-\rangle, \tag{3.54}$$
$$|\Phi^-\rangle \mapsto |\Phi^+\rangle, \tag{3.55}$$
$$|\Psi^+\rangle \mapsto i|\Psi^+\rangle, \tag{3.56}$$
$$|\Psi^-\rangle \mapsto i|\Psi^-\rangle. \tag{3.57}$$

[6] S is the phase gate $S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$.

Thus, $R = aP_s + bP_a$. Expanding these projectors we have that $R = \frac{a+b}{2}\mathbb{1} \otimes \mathbb{1} + \frac{a-b}{2}\mathbb{F}$.

□

3.2 *Channel Twirling*

Say we have a channel $\mathcal{N}\colon A \to B$ and a group G with unitary representations U_g on $\mathcal{H}_A$ and V_g on $\mathcal{H}_B$. Can we impose (G, U_g, V_g)-covariance on $\mathcal{N}$?

The answer is yes if the group belongs to a special class of topological groups called *compact* groups.[7] These are groups with a unique measure called the Haar measure which allows us to integrate over the group and average operators with respect to this measure.

Let us first look at the simpler case of a finite group,[8] where the measure we seek is just the *counting measure* $\frac{1}{|G|}\sum_{g\in G}$. Let (φ, V) be

[7] A topological group is a topological space endowed with a group structure (i.e., a binary operation on elements of the space satisfying the well-known group axioms) such that group multiplication and inversion are continuous with respect to the chosen topology.

[8] A finite group can always be regarded as a topological group when choosing the so-called discrete topology, which declares every subset to be open. The finiteness then implies that this topological space is also compact.

a representation of the finite group and $X \in \mathcal{B}(V)$ be an arbitrary operator. Then the averaged operator

$$X_G = \frac{1}{|G|} \sum_{g \in G} \varphi(g) X \varphi(g)^\dagger \tag{3.58}$$

is G-invariant in the sense that $\varphi(g) X_G = X_G \varphi(g)$ for all $g \in G$.[9] Hence, averaging an operator over the group forces it to be G-invariant. We can use the same trick also for channels:

[9] The argument for this uses the fact that any representation satisfies $\varphi(gh) = \varphi(g)\varphi(h)$ for $g, h \in G$, and is essentially given in the proof of Prop. 27 below.

Proposition 27. *Let $\mathcal{N} \colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a channel, and for a finite group let $g \mapsto U_g$, $g \mapsto V_g$ be representations of G on $\mathcal{H}$ and $\mathcal{K}$, respectively. Then the channel $\mathcal{N}_G$ defined by*

$$\mathcal{N}_G = \frac{1}{|G|} \sum_{g \in G} V_g^\dagger \mathcal{N}(U_g \cdot U_g^\dagger) V_g^\dagger. \tag{3.59}$$

is (G, U_g, V_g)-covariant.

Proof. Recall the proof of Prop. 24 where we used the transpose trick $(\mathbb{1}_A \otimes X) |\gamma\rangle = (X^T \otimes \mathbb{1}_B) |\gamma\rangle$ for any operator X. Applying this identity with U_h for $h \in G$, we obtain

$$\tilde{\tau} \equiv \tau^{\mathcal{N}_G} = \frac{1}{|G|} \sum_{g \in G} (U_g^T \otimes V_g^\dagger) \tau^{\mathcal{N}} (U_g^T \otimes V_g^\dagger)^\dagger \tag{3.60}$$

$$= \frac{1}{|G|} \sum_{h \in G} (\overline{U}_h \otimes V_h) \tau^{\mathcal{N}} (\overline{U}_h \otimes V_h)^\dagger, \tag{3.61}$$

where we made the variable change $g = h^{-1}$ in the second identity and used $(U^T)^{-1} = \overline{U}$ for a unitary U. From Prop. 24 we know that $\mathcal{N}_g$ is (G, U_g, V_g)-covariant if and only if $\tilde{\tau} = (\overline{U}_g \otimes V_g) \tilde{\tau} (\overline{U}_g \otimes V_g)^\dagger$ for all $g \in G$. Using (3.61), we have for any $g \in G$ that

$$(\overline{U}_g \otimes V_g) \tilde{\tau} (\overline{U}_g \otimes V_g)^\dagger$$

$$= \frac{1}{|G|} \sum_{h \in G} (\overline{U}_g \otimes V_g)(\overline{U}_h \otimes V_h) \tau^{\mathcal{N}} (\overline{U}_h \otimes V_h)^\dagger (\overline{U}_g \otimes V_g)^\dagger \tag{3.62}$$

$$= \frac{1}{|G|} \sum_{h \in G} (\overline{U}_{gh} \otimes V_{gh}) \tau^{\mathcal{N}} (\overline{U}_{gh} \otimes V_{gh})^\dagger \tag{3.63}$$

$$= \frac{1}{|G|} \sum_{g \in G} (\overline{U}_g \otimes V_g) \tau^{\mathcal{N}} (\overline{U}_g \otimes V_g)^\dagger, \tag{3.64}$$

where we used the fact that $h \mapsto gh$ is a bijection on G in the last equality. Thus, $\tilde{\tau} = (\overline{U}_g \otimes V_g) \tilde{\tau} (\overline{U}_g \otimes V_g)^\dagger$ for all $g \in G$, so that $\mathcal{N}_G$ is (G, U_g, V_g)-covariant by Prop. 24. □

We now turn to the more general case of arbitrary compact groups that are not necessarily finite. A compact group G admits a unique

measure called Haar measure, denoted $\mu(g)$, satisfying the following two properties:

1. Finiteness: $\int_G d\mu(g) < \infty$.

2. Left- and right-invariance: For any Haar-integrable function f on G and any $h \in G$,

$$\int_G d\mu(g) f(hg) = \int_G d\mu(g) f(g) = \int_G d\mu(g) f(gh). \tag{3.65}$$

Compact groups arise frequently, with some classic examples being the Lie groups $\mathcal{U}_d$, $\mathcal{SU}(d)$, $O(d)$, $SO(d)$, and $Sp(n)$. For a finite group G with the discrete topology, the Haar integral coincides with the counting measure $\frac{1}{|G|}\sum_{g\in G}$.

Property 1 allows us to normalize the Haar measure, so that we may always assume that $\int_G d\mu(g) = 1$. Property 2 lets us mimick the step (3.64) in the proof of Proposition 27, so that we arrive at the following version for compact groups:

Proposition 28. *Let $\mathcal{N}\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a channel, and for a compact group G with Haar measure $\mu(g)$ let $g \mapsto U_g$, $g \mapsto V_g$ be representations of G on $\mathcal{H}$ and $\mathcal{K}$, respectively. Then the channel $\mathcal{N}_G$ defined by*

$$\mathcal{N}_G = \int_G d\mu(g) V_g^\dagger \mathcal{N}(U_g \cdot U_g^\dagger) V_g^\dagger. \tag{3.66}$$

is (G, U_g, V_g)-covariant.

We now characterize channels acting on a d-dimensional space with full $\mathcal{U}_d$-covariance.

Proposition 29. *Let $\mathcal{N}\colon A \to B$ be a channel where the input and output spaces satisfy $|A| = |B| = d$. Then $N_{\mathcal{U}_d}$ is $(\mathcal{U}_d, U, U)$-covariant and hence a depolarizing channel:*

$$\mathcal{N}_{\mathcal{U}_d}(X) = (1-q)X + q\,\mathrm{tr}(X)\frac{1}{d}\mathbb{1}_d, \tag{3.67}$$

where $q = \frac{f-d^2}{1-d^2}$ with $f = \langle\gamma|\tau^{\mathcal{N}}|\gamma\rangle$.

Proof. The unitary group $\mathcal{U}_d$ is compact, and hence Prop. 28 shows that $\mathcal{N}_{\mathcal{U}_d}$ is $(\mathcal{U}_d, U, U)$-covariant.

By Prop. 25 we have $\mathcal{N}_{\mathcal{U}_d} = (1-q)X + q\,\mathrm{tr}(X)\frac{1}{d}\mathbb{1}_d$, where $q = \frac{f-d^2}{1-d^2}$ and $f = \langle\gamma|\tau^{\mathcal{N}_{\mathcal{U}_d}}|\gamma\rangle$. We conclude the proof by showing that f coin-

cides for $\mathcal{N}$ and $\mathcal{N}_{\mathcal{U}_d}$:

$$\langle\gamma|\mathbb{1}\otimes\mathcal{N}_{\mathcal{U}_d}(\gamma)|\gamma\rangle \tag{3.68}$$

$$= \langle\gamma|\int_{\mathcal{U}_d} d\mu(U)(\mathbb{1}\otimes U^\dagger)(\mathbb{1}\otimes\mathcal{N})((\mathbb{1}\otimes U)|\gamma\rangle\langle\gamma|(\mathbb{1}\otimes U^\dagger))(\mathbb{1}\otimes U)|\gamma\rangle \tag{3.69}$$

$$= \int_{\mathcal{U}_d} d\mu(U)\,\langle\gamma|(U^T\otimes U^\dagger)(\mathbb{1}\otimes\mathcal{N})(\gamma)(\overline{U}\otimes U)|\gamma\rangle \tag{3.70}$$

$$= \int_{\mathcal{U}_d} d\mu(U)\,\langle\gamma|\tau^{\mathcal{N}}|\gamma\rangle \tag{3.71}$$

$$= f_{\mathcal{N}}, \tag{3.72}$$

where we have made repeated usage of the transpose trick. □

3.3 *Irreducibly Covariant Channels and Minimum Output Entropy*

Recall that a representation (φ, V) of a group G is said to be irreducible if $\{0\}$ and V are the only G-invariant subspaces of V.[10] Schur's Lemma states that, given two irreducible representations (φ_1, V_1) and (φ_2, V_2) of a group G, the only G-linear maps are 0 (if $V_1 \ncong V_2$) or $\lambda\mathbb{1}$ for some $\lambda \neq 0$ (if $V_1 = V_2$).

[10] A subspace $S \subseteq V$ is said to be G-invariant if $\varphi(g)s \in S$ for all $g \in G$ and $s \in S$.

Lemma 30. *An irreducible representation (φ, V) of a compact group G forms a* 1*-design:*

$$\int_G d\mu(g)\varphi(g)X\varphi(g)^\dagger = \frac{\operatorname{tr}(X)}{d}\mathbb{1}_V, \tag{3.73}$$

for all $X \in \mathcal{B}(V)$, where $d = \dim V$.

Proof. Let $\mu(g)$ denote either the counting measure (if G is finite) or the Haar measure (if G is compact).

Let $Y = \int_G d\mu(g)\varphi(g)X\varphi(g)^\dagger$, which satisfies $\varphi(g)Y\varphi(g)^\dagger = Y$ for all $g \in G$. Rearranging, we have that $\varphi(g)Y = Y\varphi(g)$, that is, Y is a G-linear map. Schur's Lemma then states that $Y = c\mathbb{1}_V$ for some $c \in \mathbb{C}$. To determine c, note that

$$\operatorname{tr}(Y) = c\operatorname{tr}(\mathbb{1}_V) = cd = \int_G d\mu(g)\operatorname{tr}\left(\varphi(g)X\varphi(g)^\dagger\right) = \operatorname{tr}(X), \tag{3.74}$$

using cyclicity of the trace and normalization of the measure on the group in the last step. Thus, $c = \frac{\operatorname{tr}(X)}{d}$ as claimed. □

Proposition 31. *Let $\mathcal{N}\colon B(\mathcal{H}) \to B(\mathcal{K})$ be a (G, U_g, V_g)-covariant channel, where G is a compact group, and where $g \mapsto V_g$ on $\mathcal{K}$ is*

an irreducible representation. Then $\mathcal{N}$ satisfies $\mathcal{N}(\frac{1}{|\mathcal{H}|}\mathbb{1}_{\mathcal{H}}) = \frac{1}{|\mathcal{K}|}\mathbb{1}_{\mathcal{K}}$. Furthermore, if $|\mathcal{H}| = |\mathcal{K}|$ (i.e. $\mathcal{H} \cong \mathcal{K}$), then $\mathcal{N}$ is unital.

Proof. We have $\mathcal{N}(\mathbb{1}_{\mathcal{H}}) = \mathcal{N}(U_g \mathbb{1}_{\mathcal{H}} U_g^\dagger) = V_g \mathcal{N}(\mathbb{1}_{\mathcal{H}}) V_g^\dagger$ for all $g \in G$ because of covariance. Integrating both sides of this equation with respect to the Haar measure $\mu(g)$ and using the irreducibility of $g \mapsto V_g$ with Lemma 30, we obtain

$$\begin{aligned}
\mathcal{N}(\mathbb{1}_{\mathcal{H}}) &= \int_G d\mu(g) \mathcal{N}(\mathbb{1}_{\mathcal{H}}) && (3.75)\\
&= \int_G d\mu(g) V_g \mathcal{N}(\mathbb{1}_{\mathcal{H}}) V_g^\dagger && (3.76)\\
&= \frac{\operatorname{tr}(\mathcal{N}(\mathbb{1}_{\mathcal{H}}))}{|\mathcal{K}|} \mathbb{1}_{\mathcal{K}} && (3.77)\\
&= \frac{|\mathcal{H}|}{|\mathcal{K}|} \mathbb{1}_{\mathcal{K}}, && (3.78)
\end{aligned}$$

from which all claims follow. □

We now turn to the information-theoretic properties of covariant channels. In particular, the Holevo information assumes a simple form for irreducibly covariant channels. Recall from Section 2.2 that the Holevo information is defined as

$$\chi(\mathcal{N}) = \max_{\{p_x, \rho_x\}} \{S(\mathcal{N}(\sum_x p_x \rho_x)) - \sum_x p_x S(\mathcal{N}(\rho_x))\}. \tag{3.79}$$

As a first step, we show that this optimization problem can be simplified.

Lemma 32. *The maximization in the definition of $\chi(\mathcal{N})$ can be restricted to pure states.*

Proof. Let $E = (p_x, \rho_x)_x$ be an ensemble achieving $\chi(\mathcal{N}) = \chi(\mathcal{N}, E)$ and define a classical-quantum state

$$\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_A^x \equiv \bigoplus_x p_x \rho_A^x. \tag{3.80}$$

Here, we have changed our notation for convenience, writing $\rho_x = \rho^x$. Letting the channel $\mathcal{N}\colon A \to B$ act on the A-part of ρ_{XA} gives a state

$$\sigma_{XB} = (\mathrm{id}_X \otimes \mathcal{N})(\rho_{XA}) = \sum_x p_x |x\rangle\langle x| \otimes \mathcal{N}(\rho_A^x). \tag{3.81}$$

We have $S(\mathcal{N}(\sum_x p_x \rho_A^x)) = S(\sigma_B)$. Furthermore, using the relation

$$S\left(\sum_x q_x |x\rangle\langle x| \otimes \omega_x\right) = \sum_x q_x S(\omega_x) + H(\{q_x\}) \tag{3.82}$$

with $H(\{q_x\}) = -\sum_x q_x \log q_x$ the *Shannon entropy* of the probability distribution $\{q_x\}$, we also obtain

$$\sum_x p_x S(\mathcal{N}(\rho_A^x)) = S(\sigma_{XB}) - S(\sigma_X) = S(\sigma_{XB}) - H(\{p_x\}). \tag{3.83}$$

Plugging these into the definition of Holevo information, we obtain

$$\begin{aligned} \chi(\mathcal{N}) &= S(\sigma_B) - (S(\sigma_{XB}) - S(\sigma_X)) && (3.84)\\ &= S(\sigma_B) + S(\sigma_X) - S(\sigma_{XB}) && (3.85)\\ &= I(X;B)_\sigma, && (3.86) \end{aligned}$$

which is called the *mutual information* between X and B.

Let us now consider spectral decompositions of the states ρ_A^x, i.e., $\rho_A^x = \sum_y p_{y|x} |\varphi_{x,y}\rangle \langle\varphi_{x,y}|$. We define a classical-classical-quantum state

$$\sigma_{XYB} = \sum_{x,y} p_x p_{y|x} |x\rangle \langle x| \otimes |y\rangle \langle y| \otimes \mathcal{N}(|\varphi_{x,y}\rangle \langle\varphi_{x,y}|). \tag{3.87}$$

It is straightforward to check that $\sigma_{XB} = \mathrm{tr}_Y \sigma_{XYB}$, and hence σ_{XYB} is an extension of this state. We finish this proof by making use of the data-processing inequality (see Proposition 46), which states that tracing out a subsystem does not increase mutual information: $I(AB;C) \geq I(A;C)$ and $I(A;BC) \geq I(A;B)$. Applying this to the state σ_{XYB} from above, we get

$$I(XY;B)_\sigma \geq I(X;B)_\sigma = \chi(\mathcal{N}). \tag{3.88}$$

But the left-hand side of this inequality is the Holevo information of a state ensemble consisting of *pure* states. Hence, maximizing over arbitrary ensembles we can always find such a pure-state ensemble that maximizes the Holevo information. □

Proposition 33. *Let G be a finite or compact group with representations $g \mapsto U_g$ and $g \mapsto V_g$ on $\mathcal{H}$ and $\mathcal{K}$, respectively, and set $d_1 = \dim \mathcal{H}$ and $d_2 = \dim \mathcal{K}$. Let $\mathcal{N}\colon B(\mathcal{H}) \to B(\mathcal{K})$ be a (G, U_g, V_g)- covariant channel.*

1. *If $g \mapsto U_g$ is an irreducible representation, then*

$$\chi(\mathcal{N}) = S\left(\mathcal{N}\left(\frac{1}{d_1}\mathbb{1}_{\mathcal{H}}\right)\right) - \min_{|\psi\rangle} S(\mathcal{N}(|\psi\rangle \langle\psi|). \tag{3.89}$$

2. *If in addition $g \mapsto V_g$ is also an irreducible representation, then $\mathcal{N}(\frac{1}{d_1}\mathbb{1}_{\mathcal{H}}) = \frac{1}{d_2}\mathbb{1}_{\mathcal{K}}$ and*

$$\chi(\mathcal{N}) = \log d_2 - \min_{|\psi\rangle} S(\mathcal{N}(|\psi\rangle \langle\psi|). \tag{3.90}$$

Proof. 1. By Lem. 32 we can assume that there is a pure state ensemble $\{p_x, |\psi_x\rangle\langle\psi_x|\}$ achieving $\chi(\mathcal{N})$:

$$\chi(\mathcal{N}) = S\left(\sum_x p_x \mathcal{N}(|\psi_x\rangle\langle\psi_x|)\right) - \sum_x p_x S(\mathcal{N}(|\psi_x\rangle\langle\psi_x|)). \quad (3.91)$$

We first derive an upper bound on $\chi(\mathcal{N})$ via an upper bound on the first term and a lower bound on the second one. Later, we will see that this upper bound can be achieved.

The second term in (3.91) can easily be bounded by taking a minimization over all pure states:

$$\sum_x p_x S(\mathcal{N}(|\psi_x\rangle\langle\psi_x|)) \geq \min_{|\psi\rangle} S(\mathcal{N}(|\psi\rangle\langle\psi|)). \quad (3.92)$$

For the first term in (3.91), we use the unitary invariance of von Neumann entropy to show that for all $g \in G$ we have

$$S\left(\textstyle\sum_x p_x \mathcal{N}(|\psi_x\rangle\langle\psi_x|)\right) = S\left(\textstyle\sum_x p_x V_g \mathcal{N}(|\psi_x\rangle\langle\psi_x|) V_g^\dagger\right) \quad (3.93)$$

$$= S\left(\textstyle\sum_x p_x \mathcal{N}(U_g|\psi_x\rangle\langle\psi_x|U_g^\dagger)\right), \quad (3.94)$$

where the second equality follows from covariance of $\mathcal{N}$. Averaging both sides of this equation over the group using the Haar measure $\mu(g)$, we get

$$S\left(\textstyle\sum_x p_x \mathcal{N}(|\psi_x\rangle\langle\psi_x|)\right) \quad (3.95)$$

$$= \int_G d\mu(g) S\left(\textstyle\sum_x p_x \mathcal{N}(U_g|\psi_x\rangle\langle\psi_x|U_g^\dagger)\right) \quad (3.96)$$

$$\leq S\left(\textstyle\sum_x p_x \mathcal{N}\left(\int_G d\mu(g) U_g|\psi_x\rangle\langle\psi_x|U_g^\dagger\right)\right) \quad (3.97)$$

$$= S\left(\mathcal{N}\left(\frac{1}{d_1}\mathbb{1}_{\mathcal{H}}\right)\right), \quad (3.98)$$

where we have used concavity of von Neumann entropy (Prop. 43) in the third line, and Lem. 30 in the fourth line.

Combining both bounds, we have

$$\chi(\mathcal{N}) = S(\sum_s p_x \mathcal{N}(|\psi_x\rangle\langle\psi_x|)) - \sum_x p_x S(\mathcal{N}(|\psi_x\rangle\langle\psi_x|)) \quad (3.99)$$

$$\leq S\left(\mathcal{N}\left(\frac{1}{d_1}\mathbb{1}_{\mathcal{H}}\right)\right) - \min_{|\psi\rangle} S(\mathcal{N}(|\psi\rangle\langle\psi|)). \quad (3.100)$$

This value is achieved by the following pure-state ensemble. Let $|\varphi\rangle \in \mathcal{H}$ be such that $\min_{|\psi\rangle} S(\mathcal{N}(|\psi\rangle\langle\psi|)) = S(\mathcal{N}(\phi))$, and consider the ensemble E consisting of the vectors $|\varphi_g\rangle = U_g|\varphi\rangle$ distributed according to the Haar measure $\mu(g)$.[11] Then,

[11] For an infinite compact group this is a continuous ensemble.

$$\chi(\mathcal{N}, E) = S\left(\int_G d\mu(g)\mathcal{N}(U_g |\varphi\rangle \langle\varphi| U_g^\dagger)\right) - \int_G d\mu(g) S(\mathcal{N}(U_g |\varphi\rangle \langle\varphi| U_g^\dagger) \tag{3.101}$$

$$= S\left(\mathcal{N}\left(\frac{1}{d_1}\mathbb{1}_{\mathcal{H}}\right)\right) - \min_{|\psi\rangle} S(\mathcal{N}(|\psi\rangle \langle\psi|)), \tag{3.102}$$

where we used unitary invariance of the von Neumann entropy and normalization of the Haar measure.

2. If $g \mapsto V_g$ is also irreducible, then Prop. 31 implies that $\mathcal{N}(\frac{1}{d_1}\mathbb{1}_{\mathcal{H}}) = \frac{1}{d_2}\mathbb{1}_{\mathcal{K}}$, so that $S(\mathcal{N}(\frac{1}{d_1}\mathbb{1}_{\mathcal{H}})) = S(\frac{1}{d_2}\mathbb{1}_{\mathcal{K}}) = \log d_2$.

□

3.4 *Minimal Output Entropy and Classical Capacity*

The HSW theorem states that $\chi(\mathcal{N})$ is exactly the classical capacity of $\mathcal{N}\colon A \to B$ when we restrict to only product encodings. That is, if we use the channel k times in parallel, $\mathcal{N}^{\otimes k}(\rho_k)$, then the input ρ_k on A^k must be separable.

Allowing entangled inputs, the full classical capacity is equal to

$$C(\mathcal{N}) = \sup_{n\in\mathbb{N}} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}). \tag{3.103}$$

It is an easy consequence of the first equality above that $C(\mathcal{N}) \geq \chi(\mathcal{N})$. For a long time, a major open problem in quantum information theory was to determine if there exists a channel such that $C(\mathcal{N}) > \chi(\mathcal{N})$ (non-additivity of Holevo information).

In 2004 Peter Shor showed that Holevo information is additive if and only if the following two quantities are additive.[12]

1. Minimum output entropy

$$S_{\min}(\mathcal{N}) = \min_{|\psi\rangle} S(\mathcal{N}(|\psi\rangle \langle\psi|)).$$

2. Entanglement of formation

$$E_F(\rho_{AB}) = \min_{(p_x, |\psi_x\rangle)} \sum_x p_x S(\mathrm{tr}_B |\psi_x\rangle \langle\psi_x|),$$

where the minimum is over (not necessarily orthogonal) ensembles $(p_x, |\psi_x\rangle)$ satisfying $\rho_{AB} = \sum_x p_x |\psi_x\rangle \langle\psi_x|$.

Note that these claims do not necessarily hold for a specific channel. That is, the existence of channels with non-additive minimum output entropy implies the existence of channels with non-additive Holevo information. However, these may not be the same channels.

[12] Peter W. Shor. Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics*, 246(3):473–473, 4 2004. ISSN 1432-0916. DOI: 10.1007/s00220-004-1071-1. URL https://doi.org/10.1007/s00220-004-1071-1

While these are all equivalent statements, the minimum output entropy is much easier to work with. We will not prove the above claim in full, but we will give a construction of channels such that non-additivity of minimum output entropy implies non-additivity of Holevo information.

Let $\mathcal{N}_1, \mathcal{N}_2 \colon A_i \to B_i$ be channels such that

$$S_{\min}(\mathcal{N}_1 \otimes \mathcal{N}_2) < S_{\min}(\mathcal{N}_1) + S_{\min}(\mathcal{N}_2). \tag{3.104}$$

That is, minimum output entropy is non-additive for these channels. We now define new channels $\mathcal{N}_i' : A_i C_i \to B_i$, where $|C_i| = |B_i|^2$, via

$$\mathcal{N}_i'(\rho_{A_i} \otimes \sigma_{C_i}) = \sum_{k,l} X^k Z^l \mathcal{N}_i(\rho_{A_i})(X^k Z^l)^\dagger \langle k,l|\sigma_{C_i}|k,l\rangle , \tag{3.105}$$

where X, Z are the generalized Pauli operators on B_i. We can bound the Holevo information of this channel as follows:

$$\chi(\mathcal{N}_i') = \max_{\{p_x,\rho_x\}} S\left(\sum_x p_x \mathcal{N}_i'(\rho_x)\right) - \sum_x p_x S(\mathcal{N}_i'(\rho_x)) \tag{3.106}$$

$$\leq \log|B_i| - S_{\min}(\mathcal{N}_i'). \tag{3.107}$$

Consider the distribution $p_{k,l} = \frac{1}{|B_i|^2}$ and states $|\psi_{k,l}\rangle = |\varphi\rangle \otimes |k,l\rangle$, where $|\phi\rangle$ achieves the minimum output entropy.

$$\sum_{k,l} p_{k,l} \mathcal{N}_i'(|\varphi\rangle\langle\varphi| \otimes |k,l\rangle\langle k,l|) = \frac{1}{|B_i|^2} \sum_{k,l} X^k Z^l \mathcal{N}_i(|\varphi\rangle\langle\varphi|)(X^k Z^l)^\dagger \tag{3.108}$$

$$= \frac{1}{|B_i|} \mathbb{1}_{B_i}, \tag{3.109}$$

where the second equality follows from the generalized Pauli operators being an irreducible representation of the Heisenberg-Weyl group. By Lem. 30 twirling with this representation outputs the maximally mixed state. Thus, $S(\sum_{k,l} p_{k,l} \mathcal{N}_i'(|\varphi\rangle\langle\varphi| \otimes |k,l\rangle\langle k,l|) = \log|B_i|$.

Next, it is easy to see that

$$S(\mathcal{N}_i'(|\varphi\rangle\langle\varphi| \otimes |k,l\rangle\langle k,l|)) = S(X^k Z^l \mathcal{N}_i(|\varphi\rangle\langle\varphi|)(X^k Z^l)^\dagger) \tag{3.110}$$

$$= S(\mathcal{N}_i(|\varphi\rangle\langle\varphi|) \tag{3.111}$$

$$= S_{\min}(\mathcal{N}_i), \tag{3.112}$$

where we are using the fact that Von Neumann entropy is invariant under conjugation by unitaries. Together, these results show that

$$\chi(\mathcal{N}_i') = \log|B_i| - S_{\min}(\mathcal{N}_i). \tag{3.113}$$

Now consider the joint channel:

$$\chi(\mathcal{N}_1' \otimes \mathcal{N}_2') = \log|B_1| + \log|B_2| - S_{\min}(\mathcal{N}_1 \otimes \mathcal{N}_2) \tag{3.114}$$

$$> \log|B_1| + \log|B_2| - S_{\min}(\mathcal{N}_1) - S_{\min}(\mathcal{N}_2) \tag{3.115}$$

$$= \chi(\mathcal{N}_1') + \chi(\mathcal{N}_2'). \tag{3.116}$$

where we are using the assumed non-additivity of minimum output entropy.

We know that minimum output entropy is strongly additive for certain classes of channels, that is, $S_{\min}(\mathcal{N} \otimes \mathcal{M}) = S_{\min}(\mathcal{N}) + S_{\min}(\mathcal{M})$ for any channel $\mathcal{M}$. Some classes of such channels are the following:

1. Entanglement-breaking channels (Shor '02).
2. d-dimensional depolarizing channels (King '03).
3. Unital qubit channels (King '02).
4. Hadamard channels[13] (King et al. '04).

[13] These are complementary channels to entanglement-breaking channels.

Whether minimum output entropy was additive was an open question until Hastings gave a counter-example in 2008. However, the proof is based on random matrix theory and no explicit examples are known. We will give a high-level overview of the proof here.

Consider two quantum systems A and B. Define a channel:

$$\mathcal{E}(\rho_A) = \mathrm{tr}_A(U_{AB}(\rho_A \otimes |0\rangle\langle 0|_B)U_{AB}^\dagger), \tag{3.117}$$

with conjugate channel

$$\overline{\mathcal{E}}(\rho_A) = \mathrm{tr}_A(\overline{U}_{AB}(\rho_A \otimes |0\rangle\langle 0|_B)\overline{U}_{AB}^\dagger). \tag{3.118}$$

The idea behind the proof is to take $|A| \gg |B|$ and draw U_{AB} from the Haar measure on $\mathcal{U}(|A||B|)$. Hastings proved the following:

1. For all $U \in \mathcal{U}(|A||B|)$,
$$S_{\min}(\mathcal{E} \otimes \overline{\mathcal{E}}) \leq S(\mathcal{E} \otimes \overline{\mathcal{E}}(\Phi^+)) \leq 2\log|B| - \frac{\log|B|}{|B|}. \tag{3.119}$$
2. Let U be drawn from the Haar measure on $\mathcal{U}(|A||B|)$. Then there exists a constant c_0 such that for all constants $c \geq c_0$ and for $|A|$ sufficiently large (i.e., $|B|/|A| \in o(1)$) the following holds:
$$S_{\min}(\mathcal{E}) = S_{\min}(\overline{\mathcal{E}}) \geq \log|B| - \frac{c}{|B|}, \tag{3.120}$$
with probability $1 - o(1)$.
3. If (2) is satisfied and $\log|B| > 2c$, then
$$S_{\min}(\mathcal{E} \otimes \overline{\mathcal{E}}) \leq 2\log|B| - \frac{\log|B|}{|B|} \tag{3.121}$$
$$< 2\log|B| - \frac{2c}{|B|} \tag{3.122}$$
$$= S_{\min}(\mathcal{E}) + S_{\min}(\overline{\mathcal{E}}). \tag{3.123}$$

Details of this calculation can be found in the articles by [Fukuda et al. '09] and [Brandao and Horodecki '09]. See also the original paper by [Hastings '08].

4
Relative Entropy and Data-Processing

4.1 Motivation: quantum state discrimination

We begin by considering the task of quantum state discrimination. Assume you are given one of two quantum states ρ_0, ρ_1 with equal probability, and your goal is to decide which state you got. The strategy will be to perform a measurement[1] on the unknown state, the outcome of which will determine our guess for which state we were given.

To this end, we use a POVM $\Lambda = \{\Lambda_0, \Lambda_1\}$ corresponding to a measurement with two outcomes and measurement operators $\Lambda_1 = \mathbb{1} - \Lambda_0$ and $\Lambda_0 \geq 0$. The outcomes "0" and "1" correspond to detecting ρ_0 and ρ_1, respectively, with probabilities $p_i = \mathrm{tr}(\Lambda_i \sigma)$ for $\sigma \in \{\rho_0, \rho_1\}$.

With this data, the success probability $p_{\mathrm{succ}}(\Lambda)$ of correctly identifying the state is given by

$$\begin{aligned} p_{\mathrm{succ}}(\Lambda) &= \frac{1}{2}\mathrm{Pr}(\rho_0|\rho_0) + \frac{1}{2}\mathrm{Pr}(\rho_1|\rho_1) && (4.1)\\ &= \frac{1}{2}(\mathrm{tr}\,\Lambda_0\rho_0 + \mathrm{tr}\,\Lambda_1\rho_1) && (4.2)\\ &= \frac{1}{2}(\mathrm{tr}\,\Lambda_0\rho_0 + \mathrm{tr}((\mathbb{1} - \Lambda_0)\rho_1) && (4.3)\\ &= \frac{1}{2}(\mathrm{tr}\,\Lambda_0\rho_0 + 1 - \mathrm{tr}\,\Lambda_0\rho_1) && (4.4)\\ &= \frac{1}{2}\left(1 + \mathrm{tr}\left[\Lambda_0(\rho_0 - \rho_1)\right]\right). && (4.5) \end{aligned}$$

We aim to maximize the success probability with respect to the POVM $\Lambda = \{\Lambda_0, \Lambda_1\}$, that is, we want to find

$$p^*_{\mathrm{succ}} = \max_{\Lambda} p_{\mathrm{succ}}(\Lambda) = \frac{1}{2}\left(1 + \max_{0\leq\Lambda_0\leq\mathbb{1}} \mathrm{tr}\left[\Lambda_0(\rho_0 - \rho_1)\right]\right). \qquad (4.6)$$

Recall the definition of the trace norm

$$\|X\|_1 = \mathrm{tr}\sqrt{X^\dagger X} = \textstyle\sum_i s_i(X), \qquad (4.7)$$

where the s_i are the singular values of X.[2] This norm gives rise to the

[1] Recall that a measurement is described by a positive operator-valued measure (POVM): $\{\Lambda_i\}_{i=1}^k, \Lambda_i \geq 0, \sum_i \Lambda_i = \mathbb{1}_{\mathcal{H}}$. For a given state ρ, a POVM gives outcome i with probability $p_i = \mathrm{tr}(\Lambda_i\rho)$.

[2] That is, they are the eigenvalues of $|X| = \sqrt{X^\dagger X}$.

trace distance between two quantum states,

$$T(\rho_0, \rho_1) = \frac{1}{2}\|\rho_0 - \rho_1\|_1. \tag{4.8}$$

Our first result is a variational characterization of this trace distance that connects it to state discrimination via (4.6) above.

Lemma 34. *Let ρ_0, ρ_1 be quantum states, then*

$$\frac{1}{2}\|\rho_0 - \rho_1\|_1 = \max_{0\leq\Lambda\leq\mathbb{1}} \operatorname{tr}\left[\Lambda(\rho_0 - \rho_1)\right] \tag{4.9}$$

Proof. Since $\rho_0 - \rho_1$ is Hermitian,we can write its spectral decomposition as $\sum_i \lambda_i |i\rangle\langle i|$ where $\lambda_i \in \mathbb{R}$ and $\langle i|j\rangle = \delta_{ij}$. Defining two positive semidefinite operators

$$P = \sum_{i:\ \lambda_i \geq 0} \lambda_i |i\rangle\langle i| \geq 0, \tag{4.10}$$

$$Q = \sum_{i:\ \lambda_i < 0} (-\lambda_i) |i\rangle\langle i| \geq 0, \tag{4.11}$$

we have $P - Q = \rho_0 - \rho_1$ and

$$\|\rho_0 - \rho_1\|_1 = \operatorname{tr}|\rho_0 - \rho_1| = \operatorname{tr}|P - Q| = \operatorname{tr} P + \operatorname{tr} Q = 2\operatorname{tr} P, \tag{4.12}$$

where $\operatorname{tr}|P - Q| = \operatorname{tr} P + \operatorname{tr} Q$ follows from the fact that P, Q are each positive semidefinite and $PQ = 0 = QP$. The last equality follows from $\operatorname{tr}(P - Q) = \operatorname{tr}(\rho_0 - \rho_1) = \operatorname{tr}\rho_0 - \operatorname{tr}\rho_1 = 0$. In summary, we have $\frac{1}{2}\|\rho_0 - \rho_1\|_1 = \operatorname{tr} P$.

We now want to relate this result to $\operatorname{tr}\left[\Lambda(\rho_0 - \rho_1)\right]$ for arbitrary $0 \leq \Lambda \leq \mathbb{1}$. We write

$$\begin{aligned}
\operatorname{tr}\Lambda(\rho_0 - \rho_1) &= \operatorname{tr}\Lambda(P - Q) && (4.13)\\
&\leq \operatorname{tr}\Lambda P && (4.14)\\
&\leq \operatorname{tr} P && (4.15)\\
&= \frac{1}{2}\|\rho_0 - \rho_1\|_1 && (4.16)
\end{aligned}$$

Hence, $\max_{0\leq\Lambda_0\leq\mathbb{1}} \operatorname{tr}\Lambda(\rho_0 - \rho_1) \leq \frac{1}{2}\|\rho_0 - \rho_1\|_1$, and it remains to be shown that there exists a Λ that achieves this maximum. Setting $\Lambda = \Pi_P = \sum_{i:\ \lambda_i \geq 0} |i\rangle\langle i|$ (the projector onto the support of P), we see that

$$\begin{aligned}
\operatorname{tr}\Pi_P(\rho_0 - \rho_1) &= \operatorname{tr}\Pi_P(P - Q) && (4.17)\\
&= \operatorname{tr}\Pi_P P - \operatorname{tr}\Pi_P Q && (4.18)\\
&= \operatorname{tr} P && (4.19)\\
&= \frac{1}{2}\|\rho_0 - \rho_1\|_1, && (4.20)
\end{aligned}$$

which concludes the proof. □

For our state discrimination problem above, Lemma 34 shows that the success probability of correctly identifying the state is

$$p^*_{\text{succ}} = \frac{1}{2}(1 + \frac{1}{2}\|\rho_0 - \rho_1\|_1). \tag{4.21}$$

It follows from this formula that the success probability is equal to one if the trace distance is one, equal to $1/2$ if the trace distance is 0, and interpolates between these extreme cases. *Thus, the trace distance between two states is a measure of distinguishability. The further apart two states are with respect to this distance measure, the more distinguishable they are.* Now we ask what happens if the quantum system supporting the states ρ_i is affected by noise. Intuitively, the noise should not make the two states more distinguishable, which is the content of the next result.

Proposition 35. *Let $T(\rho_0, \rho_1) = \frac{1}{2}\|\rho_0 - \rho_1\|_1$. For any two states ρ_0, ρ_1 and a quantum channel $\mathcal{N}$, we have*

$$T(\rho_0, \rho_1) \geq T(\mathcal{N}(\rho_0), \mathcal{N}(\rho_1)). \tag{4.22}$$

Proof. We know from Lemma (34) that

$$\frac{1}{2}\|\rho_0 - \rho_1\|_1 = \max_{0 \leq \Lambda \leq 1} \operatorname{tr} \Lambda(\rho_0 - \rho_1). \tag{4.23}$$

Let $\Lambda \geq 0$ with $\Lambda \leq \mathbb{1}$ be optimal for $\frac{1}{2}\|\mathcal{N}(\rho_0) - \mathcal{N}(\rho_1)\|_1$:

$$\begin{aligned} \frac{1}{2}\|\mathcal{N}(\rho_0) - \mathcal{N}(\rho_1)\|_1 &= \operatorname{tr}\left[\Lambda(\mathcal{N}(\rho_0) - \mathcal{N}(\rho_1))\right] && (4.24)\\ &= \operatorname{tr}\left[\Lambda\mathcal{N}(\rho_0 - \rho_1)\right] && (4.25)\\ &= \operatorname{tr}\left[\mathcal{N}^\dagger(\Lambda)(\rho_0 - \rho_1)\right], && (4.26) \end{aligned}$$

where the last equality uses the definition of adjoint map. Now, we want to show 1) $\mathcal{N}^\dagger(\Lambda) \geq 0$ and 2) $\mathcal{N}^\dagger(\Lambda) \leq \mathbb{1}$.

1. Recall that the adjoint map of a completely positive map is completely positive again. In particular, $\mathcal{N}^\dagger$ is positive, which means that $\mathcal{N}^\dagger(\Lambda) \geq 0$ for any $\Lambda \geq 0$.

2. The above argument also shows that $\mathcal{N}^\dagger(\mathbb{1} - \Lambda) \geq 0$, since by assumption $\Lambda \leq \mathbb{1} :\Leftrightarrow \mathbb{1} - \Lambda \geq 0$. Furthermore, if $\mathcal{N}$ is trace-preserving then $\mathcal{N}^\dagger$ is unital, $\mathcal{N}^\dagger(\mathbb{1}) = \mathbb{1}$. We thus get

$$0 \leq \mathcal{N}^\dagger(\mathbb{1} - \Lambda) = \mathcal{N}^\dagger(\mathbb{1}) - \mathcal{N}^\dagger(\Lambda) = \mathbb{1} - \mathcal{N}^\dagger(\Lambda), \tag{4.27}$$

and hence $\mathcal{N}^\dagger(\Lambda) \leq \mathbb{1}$.

From 1) and 2) we conclude that $\mathcal{N}^\dagger(\Lambda)$ is feasible[3] in

$$\max_{0\leq K\leq \mathbb{1}} \operatorname{tr} K(\rho_0-\rho_1), \tag{4.28}$$

and thus we can bound the trace distance as

$$\begin{aligned}\frac{1}{2}\|\rho_0-\rho_1\|_1 &= \max_{0\leq K\leq \mathbb{1}} \operatorname{tr} K(\rho_0-\rho_1), &(4.29)\\ &\geq \operatorname{tr}\mathcal{N}^\dagger(\Lambda)(\rho_0-\rho_1), &(4.30)\\ &= \frac{1}{2}\|\mathcal{N}(\rho_0)-\mathcal{N}(\rho_1)\|_1, &(4.31)\end{aligned}$$

concluding the proof. □

We have thus shown that the trace distance cannot increase when subject to the same noisy process. States will never become more distinguishable after processing.

[3] A feasible region is the set of all possible points of an optimization problem that satisfy the problem's constraints.

4.2 *Error analysis and hypothesis testing*

The probability of success can be expressed

$$p_{\text{succ}} = \frac{1}{2}\Pr(\rho_0|\rho_0)+\frac{1}{2}\Pr(\rho_1|\rho_1). \tag{4.32}$$

Then, the error probability will be given as

$$p_{\text{error}} = 1-p_{\text{succ}} = \frac{1}{2}\left(\Pr(\rho_1|\rho_0)+\Pr(\rho_0|\rho_1)\right). \tag{4.33}$$

In hypothesis testing, there is a null hypothesis $H_0(\rho_0)$ and an alternative hypothesis $H_1(\rho_1)$. Then, we say a type-1 error[4] is committed when we infer ρ_1 when you actually have ρ_0 (false rejection or false negative). A type-2 error[5] is when you infer ρ_0 when you really have ρ_1 (false acceptance or false positive). There is a trade-off between these two errors, usually handled in one of the following ways:

- Symmetric hypothesis testing ("Bayesian"): try and minimize the sum of the two errors. This leads to the trace distance as discussed above.
- Asymmetric hypothesis testing: assume that type-1 error is constant and small. The question is then, how small can I make the type-2 error under this constraint.[6]

[4] Often denoted by the Greek letter α.

[5] Often denoted by the Greek letter β.

[6] In realistic settings we want to avoid false negatives at all costs!

Definition 36. Let ρ,σ be two quantum states. Let ρ be the null hypothesis and let σ be the alternative hypothesis. Further, let Λ with $0\leq\Lambda\leq\mathbb{1}$ be a test operator defining a two-element POVM $\{\Lambda,\mathbb{1}-\Lambda\}$, with the operator Λ corresponding to the

state ρ and $\mathbb{1} - \Lambda$ corresponding to the state σ. Then we have the following errors:

$$\alpha(\Lambda) = \operatorname{tr}[\rho(\mathbb{1} - \Lambda)] \qquad \text{(type-1 error)} \tag{4.34}$$

$$\beta(\Lambda) = \operatorname{tr}\sigma\Lambda \qquad \text{(type-2 error)} \tag{4.35}$$

In an information-theoretic setting we usually assume that we have access to identical copies of the unknown state in the discrimination problem, and thus we aim to distinguish the states $\rho^{\otimes n}$ and $\sigma^{\otimes n}$ for large n. In this case we have the errors[7]

[7] Note that $\Lambda_n \in B(\mathcal{H}^{\otimes n}), \Lambda_n \geq 0$, and $\Lambda_n \leq \mathbb{1}_{\mathcal{H}}^{\otimes n}$

$$\alpha_n(\Lambda_n) = \operatorname{tr}\left[\rho^{\otimes n}(\mathbb{1} - \Lambda_n)\right] \tag{4.36}$$

$$\beta_n(\Lambda_n) = \operatorname{tr}\left[\sigma^{\otimes}\Lambda_n\right]. \tag{4.37}$$

Then, for some $\epsilon > 0$, we define the following quantity describing the minimal type-II error when the type-I error is bounded:

$$\beta_n^*(\epsilon) = \min\{\beta_n(\Lambda_n) \colon 0 \leq \Lambda_n \leq \mathbb{1}, \alpha_n(\Lambda_n) \leq \epsilon\}. \tag{4.38}$$

The question is now: How does $\beta_n^*(\epsilon)$ behave as $n \to \infty$? The answer will be determined in terms of a fundamental quantity in (quantum) information theory called *relative entropy* $D(\rho\|\sigma)$.

To define it, we first define the *support* $\operatorname{supp} X$ of a linear operator $X \in B(\mathcal{H})$ as

$$\operatorname{supp} X = (\ker X)^{\perp}. \tag{4.39}$$

If X is Hermitian with spectral decomposition $X = \sum_i \lambda_i |i\rangle\langle i|$, then $\operatorname{supp} X = \operatorname{span}\{|i\rangle \colon \lambda_i \neq 0\}$. The orthogonal projection Π_X onto $\operatorname{supp} X$ is given by[8]

[8] Here, $X^{\alpha} := \sum_i \lambda_i^{\alpha} |i\rangle\langle i|$.

$$\Pi_X = \sum_{i \colon \lambda_i \neq 0} |i\rangle\langle i| = \lim_{\alpha \to 0} X^{\alpha} =: X^0. \tag{4.40}$$

Definition 37. Let ρ be a quantum state and $\sigma \geq 0$. The *relative entropy* is defined as

$$D(\rho\|\sigma) = \begin{cases} \operatorname{tr}(\rho\log\rho - \rho\log\sigma) & \text{if } \operatorname{supp}\rho \subseteq \operatorname{supp}\sigma, \\ \infty & \text{else.} \end{cases} \tag{4.41}$$

With this definition, we can give the answer to the above question, which is known as the *quantum Stein's lemma*.[9] The lemma says that for all $\epsilon > 0$,

[9] This result was shown by Hiai and Petz and Ogawa and Nagaoka.

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n^*(\epsilon) = -D(\rho\|\sigma). \tag{4.42}$$

Intuitively, this means that $\beta_n^* \approx \exp(-nD(\rho\|\sigma))$. Thus, we have another measure of distinguishability, this time in the asymmetric setting. The larger $D(\rho\|\sigma)$, the better we can distinguish between ρ and σ, in the sense that the optimal type-II error β_n^* decays exponentially if the type-I error is bounded by a constant. Note that because $D(\rho\|\sigma) \neq D(\sigma\|\rho)$, the relative entropy is not a metric in the mathematical sense. However, if ρ and σ are quantum states (that is, σ is also normalized), then $D(\rho\|\sigma) \geq 0$ with equality if and only if $\rho = \sigma$ (see below).

4.3 *Properties of relative entropy*

Measures of distinguishability between quantum states should be monotonic under quantum operations to match the intuition that noise cannot make quantum states more distinguishable. This is summarized in the following fundamental theorem.

Theorem 38 (Data-processing inequality)**.** *Let ρ be a quantum state, $\sigma \geq 0$, and $\mathcal{N}$ a quantum channel. Then,*

$$D(\rho\|\sigma) \geq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{4.43}$$

We will defer the proof to later and first study properties of the relative entropy that follow from this inequality.

Proposition 39. *The quantum relative entropy satisfies the following properties:*

1. *Let ρ be a classical state, and $\sigma \geq 0$ be classical (diagonal with respect to the same basis), that is, $\rho = \sum_x p_x |x\rangle\langle x|$ with $p = (p_x)_x$ a probability distribution, and $\sigma = \sum_x q_x |x\rangle\langle x|$ with $q_x \geq 0$ for all x. Then the quantum relative entropy equals the classical* Kullback-Leibler divergence*:*

$$D(\rho\|\sigma) = \sum_x p_x \log \frac{p_x}{q_x} = D(p\|q). \tag{4.44}$$

2. *If both ρ, σ are quantum states, then $D(\rho\|\sigma) \geq 0$, and $D(\rho\|\sigma) = 0$ iff $\rho = \sigma$.*

3. *Isometric invariance: $D(\rho\|\sigma) = D(V\rho V^\dagger \| V\sigma V^\dagger)$ for an isometry V.*

4. *For a classical-quantum state $\rho_{XA} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_A^x$ and a classical-quantum operator $\sigma_{XA} = \sum_x q_x |x\rangle\langle x|_X \otimes \sigma_A^x$ (i.e., $q_x \geq$*

0 *for all* x *and the* σ_A^x *are PSD but not necessarily normalized), we have*

$$D(\rho_{XA}\|\sigma_{XA}) = D(p\|q) + \sum_x p_x D(\rho_A^x\|\sigma_A^x). \tag{4.45}$$

5. *Joint convexity: Let* $\{\rho_x\}$ *be quantum states,* $\{\sigma_x\}$ *be positive semi-definite operators and* $\{\lambda_x\}$ *be a probability distribution. Then,*

$$D\left(\sum_x \lambda_x\rho_x\|\sum_x \lambda_x\sigma_x\right) \leq \sum_x \lambda_x D(\rho_x\|\sigma_x). \tag{4.46}$$

6. *For a state* ρ *and* $\sigma, \sigma' \geq 0$ *with* $\sigma \leq \sigma'$*, we have*

$$D(\rho\|\sigma) \geq D(\rho\sigma'). \tag{4.47}$$

Proof. 1. This follows directly from the definition of the relative entropy.

2. If ρ, σ are states, we can use the data-processing inequality in Thm. 38 with respect to the quantum channel $\mathcal{N} = \mathrm{tr}$ to obtain

$$\begin{aligned} D(\rho\|\sigma) &\geq D(\mathrm{tr}\,\rho\|\,\mathrm{tr}\,\sigma), && (4.48)\\ &= D(1\|1), && (4.49)\\ &= 0. && (4.50) \end{aligned}$$

It is clear that $D(\rho\|\sigma) = 0$ if $\rho = \sigma$. The "only if" part will be proved later.

3. Let $V\colon \mathcal{H} \to \mathcal{K}$ be an isometry. Since the map $X \mapsto VXV^\dagger$ is a quantum channel, Thm. 38 implies that

$$D(\rho\|\sigma) \geq D(V\rho V^\dagger\|V\sigma V^\dagger). \tag{4.51}$$

Now, let $\Pi = VV^\dagger$ denote the projection onto the image of V, and define a quantum channel $W\colon B(\mathcal{K}) \to B(\mathcal{H})$ as

$$W(X) = V^\dagger XV + \mathrm{tr}\left((\mathbb{1} - \Pi)X\right)|0\rangle\langle 0|, \tag{4.52}$$

where $|0\rangle \in \mathcal{H}$ is some fixed normalized vector. The map W is manifestly completely positive, and furthermore we have

$$\begin{aligned} \mathrm{tr}\,W(X) &= \mathrm{tr}\,V^\dagger XV + \mathrm{tr}((\mathbb{1} - \Pi)X) && (4.53)\\ &= \mathrm{tr}(\Pi X) + \mathrm{tr}((\mathbb{1} - \Pi)X) && (4.54)\\ &= \mathrm{tr}\,X, && (4.55) \end{aligned}$$

so that W is also trace-preserving, i.e., a quantum channel. By construction, W inverts the isometric channel $X \mapsto VXV^\dagger$ on its image,

$$W(VXV^\dagger) = V^\dagger VXV^\dagger V + \operatorname{tr}((\mathbb{1} - \Pi)VXV^\dagger) \tag{4.56}$$
$$= X + \operatorname{tr}((\mathbb{1} - V^\dagger \Pi V)X) \tag{4.57}$$
$$= X, \tag{4.58}$$

where we used $V^\dagger V = \mathbb{1}$ and $V^\dagger \Pi V = V^\dagger V V^\dagger V = \mathbb{1}$. Hence, with another application of Thm. 38, we have

$$D(V\rho V^\dagger \| V\sigma V^\dagger) \geq D(W(V\rho V^\dagger) \| W(V\sigma V^\dagger)) = D(\rho\|\sigma), \tag{4.59}$$

which together with (4.51) completes the proof.

4. Recall that we can rewrite classical-quantum states as

$$\rho_{XA} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_A^x = \bigoplus_x p_x \rho_A^x, \tag{4.60}$$

and hence

$$\log \rho_{XA} = \bigoplus_x \left[(\log p_x)\mathbb{1} + \log \rho_A^x\right], \tag{4.61}$$

and similarly for $\log \sigma_{XA}$. Thus,

$$\log \rho_{XA} - \log \sigma_{XA} = \bigoplus_x \left[\log \frac{p_x}{q_x}\mathbb{1} + \log \rho_A^x - \log \sigma_A^x\right] \tag{4.62}$$

Using the definition of $D(\cdot\|\cdot)$,

$$D(\rho_{XA}\|\sigma_{XA}) \tag{4.63}$$
$$= \operatorname{tr}\left[\rho_{XA}(\log \rho_{XA} - \log \sigma_{XA})\right] \tag{4.64}$$
$$= \operatorname{tr}\left[\left(\bigoplus_x p_x \rho_A^x\right)\left(\bigoplus_x \left(\log \frac{p_x}{q_x}\mathbb{1} + \log \rho_A^x - \log \sigma_A^x\right)\right)\right] \tag{4.65}$$
$$= \sum_x \left[p_x \log \frac{p_x}{q_x} \operatorname{tr} \rho_A^x + p_x \operatorname{tr}\left(\rho_A^x(\log \rho_A^x - \log \sigma_A^x)\right)\right] \tag{4.66}$$
$$= D(p\|q) + \sum_x p_x D(\rho_A^x \| \sigma_A^x), \tag{4.67}$$

which completes the proof.

5. To show joint convexity, define classical-quantum operators $\rho_{XA} = \sum_x \lambda_x |x\rangle\langle x| \otimes \rho_x$ and $\sigma_{XA} = \sum_x \lambda_x |x\rangle\langle x| \otimes \sigma_x$. Then from (4.45) and using $D(\lambda\|\lambda) = 0$, we get

$$D(\rho_{XA}\|\sigma_{XA}) = \sum_x \lambda_x D(\rho_x\|\sigma_x). \tag{4.68}$$

Since $\mathrm{tr}_X \rho_{XA} = \sum_x \lambda_x \rho_x$ and similarly for $\mathrm{tr}_X \sigma_{XA}$, the data-processing inequality in Thm. 38 with respect to tr_X shows that

$$\sum_x \lambda_x D(\rho_x \| \sigma_x) = D(\rho_{XA} \| \sigma_{XA}) \tag{4.69}$$

$$\geq D(\rho_A \| \sigma_A) \tag{4.70}$$

$$= D\left(\textstyle\sum_x \lambda_x \rho_x \| \sum_x \lambda_x \sigma_x\right), \tag{4.71}$$

which proves the claim.

6. Finally, we wish to show that if $\sigma \leq \sigma'$, then $D(\rho\|\sigma) \geq D(\rho\|\sigma')$. Using (4.45), we can rewrite

$$D(\rho\|\sigma) = D(\rho \otimes |0\rangle\langle 0| \,\|\sigma \otimes |0\rangle\langle 0| + (\sigma' - \sigma) \otimes |1\rangle\langle 1|). \tag{4.72}$$

Then by data-processing (Thm. 38), we have

$$D(\rho\|\sigma) \geq D(\rho\|\sigma + \sigma' - \sigma) = D(\rho\|\sigma'), \tag{4.73}$$

concluding the proof.

□

Note that the isometric invariance of the relative entropy can also be proved directly without making use of data-processing. To see this, let V be an isometry. Using the fact that $\log VXV^\dagger = V(\log X)V^\dagger$, we then have

$$D(V\rho V^\dagger \| V\sigma V^\dagger) = \mathrm{tr}\left[V\rho V^\dagger(\log V\rho V^\dagger - \log V\sigma V^\dagger)\right] \tag{4.74}$$

$$= \mathrm{tr}\left[V\rho V^\dagger V(\log\rho - \log\sigma)V^\dagger\right] \tag{4.75}$$

$$= \mathrm{tr}\left[\rho(\log\rho - \log\sigma)\right] \tag{4.76}$$

$$= D(\rho\|\sigma). \tag{4.77}$$

However, the argument using data-processing applies to any so-called *divergence*, i.e., a real-valued functional $\mathbf{D}$ on pairs of positive semidefinite operators satisfying data-processing. We have also seen that any divergence satisfying data-processing and the classical-quantum structure relation (4.45) is also jointly convex.

These results show that data-processing is a fundamental property of the relative entropy, and the next section is dedicated to proving Theorem 38.

4.4 *Proving the data-processing inequality*

Functions on operators and operator convexity

Let $A \in B(\mathcal{H})$ be Hermitian with spectral decomposition

$$A = \sum_i \lambda_i |i\rangle\langle i|, \tag{4.78}$$

that is, $\lambda_i \in \mathbb{R}$ and $\langle i|j\rangle = \delta_{ij}$. For a function $f\colon I \to \mathbb{R}$ such that $\operatorname{spec}(A) \subseteq I \subseteq \mathbb{R}$, we can define an operator

$$f(A) = \sum_i f(\lambda_i) \, |i\rangle \langle i| \, . \tag{4.79}$$

That is, $f(A)$ is a Hermitian operator with the same eigenbasis as A and spectrum $\{f(\lambda_i)\}_i$.[10] If $V\colon \mathcal{H} \to \mathcal{K}$ is an isometry, $V^\dagger V = \mathbb{1}$, the definition (4.79) of $f(A)$ implies that

$$f(VAV^\dagger) = Vf(A)V^\dagger. \tag{4.80}$$

We have already seen an example of this: the matrix logarithm. Given a state with spectral decomposition $\rho = \sum_i \lambda_i \, |i\rangle \langle i|$, the matrix logarithm is the operator $\log \rho = \sum_i \log \lambda_i \, |i\rangle \langle i|$. Another important examples are the power function $t \mapsto t^\alpha$ for $\alpha \in \mathbb{R}$ and the entropy function $\eta(t) = t \log t$.[11]

We define a partial order on Hermitian operators (called the *Löwner* order) via $A \leq B :\Leftrightarrow B - A \geq 0$. A function $f\colon I \to \mathbb{R}$ is called *operator-convex*, if

$$f(\lambda A + (1-\lambda)B) \leq \lambda f(A) + (1-\lambda) f(B), \tag{4.81}$$

holds for all A, B with $\operatorname{spec} A, \operatorname{spec} B \subseteq I$ and $\lambda \in [0,1]$. An immediate consequence of (4.81) is that every operator convex function is convex as a real function[12]. However, the converse of this statement is *not* true; for example, $t \mapsto t^3$ is not operator convex. Some examples of operator convex functions are the following:

1. $t \mapsto t^p$ for $-1 \leq p \leq 0$ and $1 \leq p \leq 2$,
2. $t \mapsto -t^p$ for $0 \leq p \leq 1$,
3. $t \mapsto -\log t$,
4. $t \mapsto \eta(t) = t \log t$.

Finally, we state (without proof) an important result called the *operator Jensen*[13] *inequality*: Let $f\colon I \to \mathbb{R}$ be operator convex, $V\colon \mathcal{H} \to \mathcal{K}$ an isometry, and A Hermitian with $\operatorname{spec} A \subseteq I$. Then,

$$f(V^\dagger A V) \leq V^\dagger f(A) V. \tag{4.82}$$

Relative Modular Operator

We now define relative modular operators, which will be a useful concept for proving the data-processing inequality for the relative entropy. For given $A, B \in B(\mathcal{H})$ we define maps $L_A, R_B\colon B(\mathcal{H}) \to B(\mathcal{H})$ via

$$L_A(X) = AX \qquad\qquad R_B(X) = XB. \tag{4.83}$$

We collect some useful properties in the following proposition.

[10] Note that this implies the often used property that $[A, f(A)] = 0$.

[11] Since $\lim_{t\to 0} \eta(t) = 0$, we define $0 \log 0 := 0$.

[12] Just take $\dim \mathcal{H} = 1$

[13] Named after Johan Jensen (1859-1925).

Lemma 40. 1. $[L_A, R_B] = 0$.

2. *If A is invertible, then L_A, R_A are invertible, and*

$$L_A^{-1} = L_{A^{-1}} \qquad R_A^{-1} = R_{A^{-1}}. \tag{4.84}$$

3. *We have $L_A^\dagger = L_{A^\dagger}$ and $R_A^\dagger = R_{A^\dagger}$ with respect to the Hilbert-Schmidt inner product $\langle X, Y\rangle = \operatorname{tr}(X^\dagger Y)$. Hence, if A is Hermitian, then so are L_A, R_A.*

4. *If A is Hermitian and $f\colon I \to \mathbb{R}$ with $\operatorname{spec} A \subseteq I$, then $f(L_A), f(R_A)$ are well-defined and*

$$f(L_A) = L_{f(A)} \qquad f(R_A) = R_{f(A)}. \tag{4.85}$$

Proof. 1. We have for all X:

$$\begin{aligned}
[L_A, R_B](X) &= L_A(R_B(X)) - R_B(L_A(X)), && (4.86)\\
&= L_A(XB) - R_B(AX), && (4.87)\\
&= AXB - AXB, && (4.88)\\
&= 0, && (4.89)
\end{aligned}$$

and hence L_A and R_B commute for all A, B.

2. This is immediate from the definition.

3. The adjoint $L_A^\dagger$ is defined via $\langle X, L_A(Y)\rangle = \langle L_A^\dagger(X), Y\rangle$ for all X, Y. Hence,

$$\begin{aligned}
\langle X, L_A(Y)\rangle &= \operatorname{tr}(X^\dagger AY), && (4.90)\\
&= \operatorname{tr}((A^\dagger X)^\dagger Y), && (4.91)\\
&= \langle A^\dagger X, Y\rangle, && (4.92)
\end{aligned}$$

which implies $L_A^\dagger = L_{A^\dagger}$. If $A^\dagger = A$, then also $L_A^\dagger = L_A$. The claims for R_A are proved analogously.

4. Let $A = \sum_{i=1}^d a_i |i\rangle\langle i|$ where $d = \dim \mathcal{H}$. Since

$$L_A(|i\rangle\langle j|) = A|i\rangle\langle j| = a_i |i\rangle\langle j|, \tag{4.93}$$

we see that for each fixed i there are d eigenoperators $|i\rangle\langle j|$ with eigenvalues a_i (for $i, j = 1, \ldots, d$). In total there are d^2 eigenvalues with orthogonal eigenoperators $|i\rangle\langle j|$ (with respect to the Hilbert-Schmidt inner product). These eigenoperators form a basis since $\dim B(\mathcal{H}) = d^2$, and thus we can define $f(L_A)$ through

$$f(L_A)(|i\rangle\langle j|) \coloneqq f(a_i)|i\rangle\langle j| \tag{4.94}$$

But $f(a_i)|i\rangle\langle j| = L_{f(A)}(|i\rangle\langle j|)$, and hence $f(L_A) = L_{f(A)}$ since they agree on a basis. The same argument applies for R_A. □

Definition 41. Let $X, Y \in B(\mathcal{H})$ be Hermitian and Y invertible. The relative modular operator $\Delta \equiv \Delta^{X,Y}$ is defined as

$$\Delta^{X,Y} = L_X R_{Y^{-1}} \colon Z \mapsto XZY^{-1}. \tag{4.95}$$

Lemma 42. *Let $X, Y \geq 0$ with Y invertible, and set $\eta(t) = t \log t$. Then $\eta(\Delta^{X,Y}) = \Delta^{X,Y}(L_{\log X} - R_{\log Y})$.*

Proof. Let X, Y have spectral decompositions[14]

$$X = \sum_i x_i \left|e_i\right\rangle \left\langle e_i\right|, \tag{4.96}$$

$$Y = \sum_i y_i \left|f_i\right\rangle \left\langle f_i\right|. \tag{4.97}$$

[14] Note that $Y \geq 0$ being invertible implies $Y > 0$, i.e., $y_i > 0$ for all i.

Then, we have

$$\begin{aligned} \Delta^{X,Y}(\left|e_i\right\rangle \left\langle f_j\right|) &= L_X R_{Y^{-1}}(\left|e_i\right\rangle \left\langle f_j\right|), && (4.98)\\ &= X \left|e_i\right\rangle \left\langle f_j\right| Y^{-1}, && (4.99)\\ &= x_i \left|e_i\right\rangle \left\langle f_i\right| y_j^{-1}, && (4.100)\\ &= x_i y_j^{-1} \left|e_i\right\rangle \left\langle f_j\right|, && (4.101) \end{aligned}$$

which implies that $\left|e_i\right\rangle \left\langle f_j\right|$ is an eigenoperator of $\Delta^{X,Y}$ with eigenvalue $x_i y_j^{-1}$, so that

$$\eta(\Delta)(\left|e_i\right\rangle \left\langle f_j\right|) = \eta(x_i y_j^{-1}) \left|e_i\right\rangle \left\langle f_j\right| = x_i y_j^{-1}(\log x_i - \log y_j) \left|e_i\right\rangle \left\langle f_j\right|. \tag{4.102}$$

Since we also have

$$\Delta^{X,Y}(L_{\log X} - R_{\log Y})(\left|e_i\right\rangle \left\langle f_j\right|) = x_i y_j^{-1}(\log x_i - \log y_j) \left|e_i\right\rangle \left\langle f_j\right| \tag{4.103}$$

and the $\{\left|e_i\right\rangle \left\langle f_j\right|\}_{i,j=1}^d$ form a basis for $B(\mathcal{H})$, we can conclude that

$$\eta(\Delta^{X,Y}) = \Delta^{X,Y}(L_{\log X} - R_{\log Y}), \tag{4.104}$$

as desired. □

We are now ready to prove Theorem 38, the data processing inequality for the quantum relative entropy.

Proof. First, recall that $D(\rho\|\sigma) \equiv \infty$ if the support condition $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$ is *not* satsified. In this case the data-processing inequality is trivially satisfied, and so we assume $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$ in the sequel. Without loss of generality, we restrict the Hilbert space to the support of σ, which allows us to assume that σ is invertible.[15]

[15] In general, $\mathcal{H} = \ker\sigma \oplus \operatorname{supp}\sigma$, and we can restrict $\mathcal{H}$ to $\operatorname{supp}\sigma$ by projecting.

Recall that for any quantum channel $\mathcal{N}\colon A \to B$ there exists an isometry $V\colon \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ with an auxiliary Hilbert space $\mathcal{H}_E$ such that

$$\mathcal{N}(X_A) = \operatorname{tr}_E(VX_AV^\dagger) \tag{4.105}$$

We further have $f(VAV^\dagger) = Vf(A)V^\dagger$ by definition of the operator function f. Applying this to the matrix logarithm yields the previously noted isometric variance of the relative entropy, $D(V\rho V^\dagger \| V\sigma V^\dagger) = D(\rho\|\sigma)$. Hence, the claim of Thm. 38 follows from proving the data-processing inequality for the partial trace $\mathcal{N} = \operatorname{tr}_B\colon AB \to A$,

$$D(\rho_{AB}\|\sigma_{AB}) \geq D(\rho_A\|\sigma_A). \tag{4.106}$$

We have $\operatorname{supp}\sigma_{AB} \subseteq \operatorname{supp}\sigma_A \otimes \operatorname{supp}\sigma_B$,[16] and so we can assume without loss of generality that both σ_{AB} and σ_A are invertible.

[16] Proved in Lemma B.4.1 of Renato Renner's PhD thesis.

Recall the Hilbert-Schmidt inner product $\langle X, Y\rangle := \operatorname{tr}(X^\dagger Y)$ on $B(\mathcal{H})$. Defining relative modular operators

$$\Delta_{AB} = L_{\rho_{AB}}R_{\sigma_{AB}^{-1}} \qquad \Delta_A = L_{\rho_A}R_{\sigma_A^{-1}}, \tag{4.107}$$

we can rewrite the relative entropy $D(\cdot\|\cdot)$ as follows:

$$\begin{aligned} D(\rho_{AB}\|\sigma_{AB}) &= \operatorname{tr}\rho_{AB}(\log\rho_{AB} - \log\sigma_{AB}) && (4.108)\\ &= \langle\sigma_{AB}^{1/2}, \rho_{AB}\log\rho_{AB}\sigma_{AB}^{1/2}\sigma_{AB}^{-1} - \rho_{AB}\sigma_{AB}^{1/2}\log\sigma_{AB}\sigma_{AB}^{-1}\rangle, && (4.109)\\ &= \langle\sigma_{AB}^{1/2}, \Delta_{AB}(L_{\log\rho_{AB}} - R_{\log\sigma_{AB}})(\sigma_{AB}^{1/2})\rangle, && (4.110)\\ &= \langle\sigma_{AB}^{1/2}, \eta(\Delta_{AB})(\sigma_{AB}^{1/2})\rangle, && (4.111) \end{aligned}$$

where the last equality follows from Lem. 42. Similarly,

$$D(\rho_A\|\sigma_A) = \langle\sigma_A^{1/2}, \eta(\Delta_A)(\sigma_A^{1/2})\rangle. \tag{4.112}$$

Our goal is to use the operator Jensen's inequality, which states that for any operator-convex function f and isometry V we have

$$f(V^\dagger XV) \leq V^\dagger f(X)V. \tag{4.113}$$

To this end, we note that $\eta(t) = t\log t$ is operator-convex, and we wish to find a map $V\colon B(\mathcal{H}_A) \to B(\mathcal{H}_{AB})$ satisfying the following properties:

1. V is an isometry, $V^\dagger V = \mathrm{id}_A$;
2. $V^\dagger\Delta_{AB}V = \Delta_A$;
3. $V(\sigma_A^{1/2}) = \sigma_{AB}^{1/2}$.

Once we have found such an isometry, we can proceed with the proof of the data-processing inequality as follows:

$$\begin{aligned}
D(\rho_A\|\sigma_A) &= \langle \sigma_A^{1/2}, \eta(\Delta_A)(\sigma_A^{1/2})\rangle && (4.114)\\
&= \langle \sigma_A^{1/2}, \eta(V^\dagger \Delta_{AB} V)(\sigma_A^{1/2})\rangle, \quad \text{by property 2} && (4.115)\\
&\leq \langle \sigma_A^{1/2}, V^\dagger \eta(\Delta_{AB}) V(\sigma_A^{1/2})\rangle, \quad \text{by (4.113)} && (4.116)\\
&= \langle V(\sigma_A^{1/2}), \eta(\Delta_{AB}) V(\sigma_A^{1/2})\rangle && (4.117)\\
&= \langle \sigma_{AB}^{1/2}, \eta(\Delta_{AB})(\sigma_{AB}^{1/2})\rangle, \quad \text{by property 3} && (4.118)\\
&= D(\rho_{AB}\|\sigma_{AB}), && (4.119)
\end{aligned}$$

as desired.

It remains to choose the isometry V. We set

$$V\colon X_A \mapsto (X_A \sigma_A^{-1/2} \otimes \mathbb{1}_B)\sigma_{AB}^{1/2}, \tag{4.120}$$

and check the three properties above:
Property 1: The adjoint of V is given by

$$V^\dagger(Y_{AB}) = \mathrm{tr}_B(Y_{AB}\sigma_{AB}^{1/2}(\sigma_A^{-1/2}\otimes \mathbb{1}_B)), \tag{4.121}$$

and hence we have

$$\begin{aligned}
V^\dagger V(X_A) &= V^\dagger(X_A \sigma_A^{-1/2}\sigma_{AB}^{1/2}), && (4.122)\\
&= \mathrm{tr}_B(X_A \sigma_A^{-1/2}\sigma_{AB}^{1/2}\sigma_{AB}^{1/2}\sigma_A^{-1/2}), && (4.123)\\
&= X_A \sigma_A^{-1/2}\sigma_A\sigma_A^{-1/2}, && (4.124)\\
&= X_A. && (4.125)
\end{aligned}$$

Since this holds for all X_A, we can conclude that $V^\dagger V = \mathrm{id}_A$ as desired.
Property 2: We have

$$\begin{aligned}
V^\dagger \Delta_{AB} V(X_A) &= V^\dagger \Delta_{AB}(X_A \sigma_A^{-1/2}\sigma_{AB}^{1/2}), && (4.126)\\
&= V^\dagger(\rho_{AB} X_A \sigma_A^{-1/2}\sigma_{AB}^{1/2}\sigma_{AB}^{-1}), && (4.127)\\
&= V^\dagger(\rho_{AB} X_A \sigma_A^{-1/2}\sigma_{AB}^{-1/2}) && (4.128)\\
&= \mathrm{tr}_B(\rho_{AB} X_A \sigma_A^{-1}), && (4.129)\\
&= \rho_A X_A \sigma_A^{-1}, && (4.130)\\
&= \Delta_A(X_A), && (4.131)
\end{aligned}$$

which holds for all X_A, and thus $V^\dagger \Delta_{AB} V = \Delta_A$ as desired.
Property 3: We have $V(\sigma_A^{1/2}) = \sigma_{AB}^{1/2}$ by definition of V.

This concludes the proof of Theorem 38. □

A natural question is now: how does the data-processing inequality generalize? We have proved it above for quantum channels, but Denes Petz extended these methods to prove the DPI for trace-preserving 2-positive maps.[17] Moreover, a very recent result by Mueller-Hermes and Reeb shows that DPI holds for all trace-preserving, positive maps.[18]

[17] Φ is 2-positive if $\Phi \otimes \mathrm{id}_2$ is positive.

[18] Note that they use different proof methods based on complex interpolation.

5
Entropies and equality in data-processing

5.1 Entropic quantities

Von Neumann entropy

Entropies are fundamental quantities in information theory. Probably the most important one is the von Neumann entropy,[1] defined as

$$S(\rho) = -\operatorname{tr}\rho\log\rho = -D(\rho\|\mathbb{1}). \tag{5.1}$$

[1] We often denote the entropy as $S(A)_\rho = S(\rho_A)$ where $\rho_A \in B(\mathcal{H}_A)$ is a quantum state.

The most useful entropic quantities also have operational interpretation. For the von Neumann entropy, there are two operational interpretations: one related to source/data compression and one to entanglement conversion of pure states; more on this later. First, let us review some of the basic properties of the von Neumann entropy.

Proposition 43. *The von Neumann entropy $S(\cdot)$ satisfies the following properties:*

1. $0 \leq S(A)_\rho \leq \log|A|$ *for any state ρ_A, with equality in the first inequality iff ρ_A is pure, and equality in the second inequality iff $\rho_A = \frac{1}{|A|}\mathbb{1}$ is completely mixed.*
2. *Concavity: for any collection of states ρ_i and a probability distribution $(\lambda_i)_i$,*

$$S(\textstyle\sum_i \lambda_i\rho_i) \geq \sum_i \lambda_i S(\rho_i). \tag{5.2}$$

3. *Strong subadditivity: for any tripartite state $\rho_{ABC} \in B(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$,*

$$S(ABC) + S(C) \leq S(AC) + S(BC). \tag{5.3}$$

Taking $|C| = 1$ gives the weaker subadditivity relation $S(AB) \leq S(A) + S(B)$.

4. *Let $\mathcal{N}$ be a unital quantum channel, $\mathcal{N}(\mathbb{1}) = \mathbb{1}$. Then, $S(\rho) \leq S(\mathcal{N}(\rho))$ for all $\rho \in B(\mathcal{H})$. In particular, let $\{\Pi_i\}_{i=1}^k$ be a projective measurement satisfying $\Pi_i \geq 0$ for all i, $\Pi_i\Pi_j = \delta_{ij}\Pi_i$, and $\sum_i \Pi_i = \mathbb{1}$. Then $\mathcal{N}(X) = \sum_i \Pi_i X \Pi_i$ is a unital channel and hence $S(\rho) \leq S(\mathcal{N}(\rho))$. In other words, measuring a system always increases entropy. Choosing $\Pi_i = |i\rangle\langle i|$ for an ONB $\{|i\rangle\}$ gives the special case $S(\rho) \leq S(\mathrm{diag}(\rho))$, where $\mathrm{diag}(\rho)$ denotes the diagonal part of ρ.*

Proof. 1. Let $\rho = \sum_i \lambda_i |i\rangle\langle i|$ be the spectral decomposition of ρ. Then the entropy can be expressed as $S(\rho) = -\sum_i \lambda_i \log \lambda_i$.[2] Then clearly $S(\rho) \geq 0$ because $\lambda_i \in [0, 1]$ for all i. The only way to have $S(\rho) = 0$ is if there is exactly one non-zero eigenvalue equal to one; that is, $S(\rho) = 0$ iff ρ is pure.

[2] That is, the von Neumann entropy of a quantum state is equal to the Shannon entropy of its eigenvalue distribution.

For the upper bound on the entropy, observe that

$$D\left(\rho_A \middle\| \frac{1}{|A|}\mathbb{1}_A\right) = \mathrm{tr}\,\rho_A \left(\log \rho_A - \log \frac{1}{|A|}\mathbb{1}_A\right), \tag{5.4}$$

$$= -S(\rho_A) + \log |A| \tag{5.5}$$

$$\geq 0, \tag{5.6}$$

and we have equality iff $\rho_A = \frac{1}{|A|}\mathbb{1}_A$ is maximally mixed.[3]

[3] We will prove this later!

2. To show concavity, observe that

$$S(\textstyle\sum_i \lambda_i \rho_i) = -D(\sum_i \lambda_i \rho_i \| \mathbb{1}), \tag{5.7}$$

$$= -D(\textstyle\sum_i \lambda_i \rho_i \| \sum_i \lambda_i \mathbb{1}), \tag{5.8}$$

$$\geq \textstyle\sum_i \lambda_i (-D(\rho_i \| \mathbb{1})), \tag{5.9}$$

$$= \textstyle\sum_i \lambda_i S(\rho_i) \tag{5.10}$$

where the inequality follows from the joint convexity of relative entropy proved in Proposition 39.

3. To show the strong subadditivity relation $S(ABC) + S(C) \leq S(AC) + S(BC)$, we first note that

$$D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) = \mathrm{tr}\,\rho_{ABC}(\log \rho_{ABC} - \log \rho_A \otimes \rho_{BC}). \tag{5.11}$$

We have[4]

[4] Here, we use the fact that $\log XY = \log X + \log Y$ if $[X, Y] = 0$. We apply this to $\rho_A \otimes \rho_{BC} = (\rho_A \otimes \mathbb{1}_{BC})(\mathbb{1}_A \otimes \rho_{BC})$.

$$\log \rho_A \otimes \rho_{BC} = \log (\rho_A \otimes \mathbb{1}_{BC}) + \log (\mathbb{1}_A \otimes \rho_{BC}), \tag{5.12}$$

$$= \log \rho_A \otimes \mathbb{1}_{BC} + \mathbb{1}_A \otimes \log \rho_{BC}, \tag{5.13}$$

which follows from the definition of the matrix logarithm.[5] Then,

[5] See, e.g., this post.

substituting back into the expression above we have

$$\begin{aligned} D(\rho_{ABC}\|\rho_A \otimes \rho_{BC}) &= \operatorname{tr} \rho_{ABC} \log \rho_{ABC} && (5.14)\\ &\quad - \operatorname{tr} \rho_{ABC}(\log \rho_A \otimes \mathbb{1}_{BC}) && (5.15)\\ &\quad - \operatorname{tr} \rho_{ABC}(\mathbb{1}_A \otimes \rho_{BC}) && (5.16)\\ &= -S(ABC) + S(A) + S(BC). && (5.17) \end{aligned}$$

An analogous calculation shows that

$$D(\rho_{AC}\|\rho_A \otimes \rho_C) = -S(AC) + S(A) + S(C). \tag{5.18}$$

By the data-processing inequality with respect to $\mathcal{N} = \operatorname{tr}_B$, we have

$$\begin{aligned} -S(ABC) + S(A) + S(BC) &= D(\rho_{ABC}\|\rho_A \otimes \rho_{BC}) && (5.19)\\ &\geq D(\rho_{AC}\|\rho_A \otimes \rho_C) && (5.20)\\ &= -S(AC) + S(A) + S(C), && (5.21) \end{aligned}$$

which rearranged gives strong subadditivity. Taking $|C| = 1$, we obtain $S(AB) \leq S(A) + S(B)$.

4. Finally we need to show that if $\mathcal{N}$ is unital then $S(\rho) \leq S(\mathcal{N}(\rho))$. This is a straightforward application of data-processing:

$$S(\rho) = -D(\rho\|\mathbb{1}) \leq -D(\mathcal{N}(\rho)\|\mathcal{N}(\mathbb{1})) = -D(\mathcal{N}(\rho)\|\mathbb{1}) = S(\mathcal{N}(\rho)). \tag{5.22}$$

□

Conditional entropy

Another ubiquitous quantity in quantum information theory is the conditional entropy

$$S(A|B)_\rho = S(AB)_\rho - S(B)_\rho = -D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B). \tag{5.23}$$

While the *classical* conditional entropy $H(X|Y) = H(XY) - H(Y)$ is always non-negative, this fails for the quantum conditional entropy. To see this, note that a maximally entangled state $|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB})$ satisfies $S(AB)_{\Phi^+} = 0$ (since it is a pure state) and $S(B)_{\Phi^+} = \log 2 = 1$ (since the marginal on B is completely mixed). Hence,

$$S(A|B)_{\Phi^+} = -1. \tag{5.24}$$

The negativity of the conditional entropy is due to the entanglement; remarkably, there is an operational interpretation in terms of a task called state merging,[6] which gives this negative conditional entropy a precise information-theoretic meaning.

[6] Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269:107–136, 2007

Proposition 44. *The quantum conditional entropy satisfies the following properties:*

1. *Conditioning reduces entropy:*

$$S(A|B) \leq S(A). \tag{5.25}$$

2. *Duality relation: let ρ_{AB} have a purification $|\rho\rangle_{ABC}$, then*

$$S(A|B)_\rho = -S(A|C)_\rho. \tag{5.26}$$

3. *Bounds on conditional entropy:*

$$-\log|A| \leq S(A|B) \leq \log|A|, \tag{5.27}$$

   with equality in the first inequality for the maximally entangled state between systems A and B (if $|A| = |B|$), and equality in the second inequality if ρ_{AB} is product and completely mixed on A.

4. *Data-processing (strong subadditivity):*

$$S(A|BC) \leq S(A|B). \tag{5.28}$$

5. *Weak monotonicity:*

$$S(A|B) + S(A|C) \geq 0. \tag{5.29}$$

6. *Classical conditioning: Let $\rho_{AX} = \sum_x p_x |x\rangle\langle x|_A \otimes \rho_A^x$ be a quantum-classical state, then*

$$S(A|X) = \sum_x p_x S(A)_{\rho^x}. \tag{5.30}$$

   In particular, $S(A|X) \geq 0$.

7. *Concavity: For a collection of states ρ_{AB}^i and a probability distribution (λ_i) denote by $\bar{\rho} = \sum_i \lambda_i \rho_{AB}^i$ the average state. Then,*

$$S(A|B)_{\bar{\rho}} \geq \sum_i \lambda_i S(A|B)_{\rho^i}. \tag{5.31}$$

Proof. 1. We have $S(A|B) = S(AB) - S(B) \leq S(A)$, which is equivalent to subadditivity of the von Neumann entropy.

2. We have $S(AB) = S(C)$ for any pure state $|\rho\rangle_{ABC}$ by Schmidt decomposition, and similarly $S(B) = S(AC)$. Hence,

$$S(A|B) = S(AB) - S(B) = S(C) - S(AC) = -S(A|C). \tag{5.32}$$

3. The lower bound on the range of the conditional entropy follows from the proof of 1) above and from Proposition 43: For a state $|\rho\rangle_{ABC}$ purifying ρ_{AB},

$$S(A|B) = -S(A|C) \geq -S(A) \geq -\log|A|. \tag{5.33}$$

Equality holds for a maximally entangled state $|\Phi^+\rangle_{AB}$ between A and B, as observed above.

For the upper bound, note that

$$\begin{aligned} 0 &\leq D\left(\rho_{AB}\middle\|\frac{1}{|A|}\mathbb{1}_A \otimes \rho_B\right) && (5.34)\\ &= \operatorname{tr}\rho_{AB}\log\rho_{AB} - \operatorname{tr}\rho_{AB}\log\left[\frac{1}{|A|}\mathbb{1}_A \otimes \rho_B\right] && (5.35)\\ &= -S(AB) + \log|A| + S(B), && (5.36)\end{aligned}$$

which gives the desired upper bound. Furthermore, we have equality iff $\rho_{AB} = \frac{1}{|A|}\mathbb{1}_A \otimes \rho_B$.

4. We have $S(ABC) + S(B) \leq S(AB) + S(BC)$ by strong subadditivity, and hence

$$S(A|BC) = S(ABC) - S(BC) \leq S(AB) - S(B) = S(A|B). \tag{5.37}$$

5. Let $|\rho\rangle_{ABCD}$ be a purification of the tripartite state ρ_{ABC}. Then by Schmidt decomposition we have $S(AC) = S(BD)$ and $S(C) = S(ABD)$, so that

$$\begin{aligned} S(A|B) + S(A|C) &= S(AB) - S(B) + S(AC) - S(C) && (5.38)\\ &= S(AB) - S(B) + S(BD) - S(ABD) && (5.39)\\ &\geq 0 && (5.40)\end{aligned}$$

by strong subadditivity.

6. For $\rho_{AX} = \sum_x p_x \rho_A^x \otimes |x\rangle\langle x|_X$ we have $\rho_X = \sum_x p_x |x\rangle\langle x|$. Then,

$$\begin{aligned} S(A|X) &= -D(\rho_{AX}\|\mathbb{1}_A \otimes \rho_x), && (5.41)\\ &= -D(\textstyle\sum_x p_x \rho_A^x \otimes |x\rangle\langle x|_X \,\|\, \mathbb{1}_A \otimes \sum_x p_x |x\rangle\langle x|), && (5.42)\\ &= -\sum_x p_x D(\rho_A^x\|\mathbb{1}_A), && (5.43)\\ &= \sum_x p_x S(A)_{\rho^x}, && (5.44)\end{aligned}$$

where the third equality holds by part 4) of Proposition 39.

7. Note that $\bar{\rho}_{AB} = \sum_i \lambda_i \rho^i_{AB}$ implies $\bar{\rho}_B = \sum_i \lambda_i \rho^i_B$. Then,

$$\begin{aligned} S(A|B)_{\bar{\rho}} &= -D(\bar{\rho}_{AB} \| \mathbb{1}_A \otimes \bar{\rho}_B) && (5.45)\\ &= -D(\textstyle\sum_i \lambda_i \rho^i_{AB} \| \mathbb{1}_A \otimes \sum_i \rho^i_B), && (5.46)\\ &\geq \sum_i \lambda_i (-D(\rho^i_{AB} \| \mathbb{1}_A \otimes \rho^i_B)), && (5.47)\\ &= \sum_i \lambda_i S(A|B)_{\rho^i}, && (5.48) \end{aligned}$$

where the inequality follows from the joint convexity of relative entropy.

□

Corollary 45.

1. *The conditional entropy of product states is equal to a von Neumann entropy and thus non-negative:* $S(A|B)_{\omega\otimes\tau} = S(A)_\omega \geq 0$ *for any state* $\omega_A \otimes \tau_B$.
2. $S(A|B)_\rho \geq 0$ *for any separable state* ρ_{AB}.

Proof. 1. The von Neumann entropy is additive on product states: $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$. It then immediately follows that

$$\begin{aligned} S(A|B)_{\omega\otimes\tau} &= S(AB)_{\omega\otimes\tau} - S(B)_\tau && (5.49)\\ &= S(A)_\omega + S(B)_\tau - S(B)_\tau && (5.50)\\ &= S(A)_\omega. && (5.51) \end{aligned}$$

2. A separable state ρ_{AB} can be written as $\rho_{AB} = \sum_i \lambda_i \omega^i_A \otimes \tau^i_B$. The concavity of conditional entropy and item 1) above then imply that

$$S(A|B)_\rho \geq \sum_i \lambda_i S(A|B)_{\omega^i\otimes\tau^i} = \sum_i \lambda_i S(A)_{\omega^i} \geq 0. \quad (5.52)$$

□

Note that the converse of part 2) of Cor. 45 does *not* hold, since there are bound entangled states with non-negative conditional entropy.

Coherent information

Another well-studied entropic quantity is the *coherent information* which is defined as

$$\begin{aligned} I_c(A\rangle B)_\rho &= -S(A|B)_\rho, && (5.53)\\ &= S(B) - S(AB), && (5.54)\\ &= D(\rho_{AB} \| \mathbb{1}_A \otimes \rho_B) && (5.55) \end{aligned}$$

The coherent information has multiple operational interpretations in entanglement distillation, quantum information transmission through a quantum channel, and quantum error correction. Of particular note is the so-called hashing inequality, which states that a quantum state ρ_{AB} with $I_c(A\rangle B)_\rho > 0$ is distillable using one-way LOCC. We also note that

$$S(A|B) + S(A|C) \geq 0 \iff I_c(A\rangle B) + I(A\rangle C) \leq 0, \tag{5.56}$$

which can be interpreted as a statement of the no-cloning theorem.

Mutual information

Yet another crucially important entropic quantity is the *mutual information*:[7]

$$\begin{aligned} I(A;B)_\rho &= S(A) + S(B) - S(AB), && (5.57)\\ &= S(A) - S(A|B), && (5.58)\\ &= S(B) - S(B|A), && (5.59)\\ &= D(\rho_{AB} \| \rho_A \otimes \rho_B). && (5.60) \end{aligned}$$

[7] A useful way of understanding the mutual information is the relative entropy distance from being a product state.

It has the following operational interpretations:

1. Measure for total correlations (classical and quantum) in a bipartite state
2. entanglement-assisted classical communication
3. classical communication cost in state merging

The mutual information has many important features, several of which are summarized in the following proposition.

Proposition 46. *The mutual information satisfies the following properties:*

1. *Lower and upper bounds:*

$$0 \leq I(A;B)_\rho \leq 2 \log \min\{|A|, |B|\}, \tag{5.61}$$

   and if ρ_{XB} is classical-quantum, then

$$0 \leq I(X;B)_\rho \leq \log \min\{|X|, |B|\}. \tag{5.62}$$

2. *Data-processing:*

$$I(A;BC) \geq I(A;B) \qquad I(AB;C) \geq I(A;C). \tag{5.63}$$

3. *Holevo information: Let $\mathcal{E} = \{p_x, \rho_A^x\}$ be a quantum state ensem-*

ble. Then the Holevo information is equal to

$$\chi(\mathcal{E}) = S(\textstyle\sum_x p_x \rho_A^x) - \sum_x p_x S(\rho_A^x) = I(X;A)_\rho, \tag{5.64}$$

where $\rho_{XA} = \sum_x p_x |x\rangle\langle x|_x \otimes \rho_A^x$.

Proof. 1. By part 2) of Prop 39, we have

$$I(A;B)_\rho = D(\rho_{AB} \| \rho_A \otimes \rho_B) \geq 0. \tag{5.65}$$

We also have[8]

$$I(A;B) = S(A) - S(A|B) \leq S(A) + \log|A| \leq 2\log|A|, \tag{5.66}$$

and using the formula $I(A;B) = S(B) - S(B|A)$ instead gives

$$I(A;B) \leq 2\log|B|. \tag{5.67}$$

We have $S(A|B) = -\log|A|$ for a maximally entangled state Φ_{AB}^+, and hence Φ^+ achieves the maximal value of the mutual information:

$$I(A;B)_{\Phi^+} = S(A) - S(A|B) = 2\log|A|. \tag{5.68}$$

When $\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_A^x$ is classical-quantum, we have $S(A|X) \geq 0$ by part 6) of Prop. 44, and hence

$$I(X;A) = S(A) - S(A|X) \leq S(A) \leq \log|A|. \tag{5.69}$$

On the other hand,[9]

$$I(X;A) = S(X) - S(X|A) \leq S(X) \leq \log|X|. \tag{5.70}$$

2. To show that $I(AB;C) \geq I(A;C)$, we simply apply data-processing with respect to the channel $\mathrm{tr}_B(\cdot)$ to the relative entropy formula for $I(\cdot;\cdot)$:

$$I(AB;C) = D(\rho_{ABC} \| \rho_{AB} \otimes \rho_C) \geq D(\rho_{AC} \| \rho_A \otimes \rho_C) = I(A;C), \tag{5.71}$$

and similarly for $I(A;BC) \geq I(A;B)$.

3. Note that $\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_A^x$ implies $\rho_A = \sum_x p_x \rho_A^x$. Then,

$$\begin{aligned} I(X;A) &= S(A) - S(A|X) && (5.72)\\ &= S(\textstyle\sum_x p_x \rho_A^x) - \sum_x p_x S(\rho_A^x), && (5.73)\\ &= \chi(\{p_x, \rho_A^x\}), && (5.74) \end{aligned}$$

where the second equality follows from part 6) of Prop. 44. □

[8] Recall from Prop 44 that $S(A|B) \geq -\log|A|$.

[9] By Corollary 45 we have $S(X|A) \geq 0$ since ρ_{XA} is separable.

We now discuss a fundamental result in quantum information theory, the Holevo bound. To this end, let $X \sim p(x)$ be a classical, discrete random variable, $\{\rho_B^x\}$ be a set of quantum states, and let $M = \{M_B^y\}_y$ be a POVM ($M^y \geq 0, \sum_y M^y = \mathbb{1}_B$). Let Y be the random variable corresponding to the measurement outcome, i.e., defined by $p(x,y) = p(y|x)p(x)$ and the conditional probability distribution

$$p(y|x) = \mathrm{tr}(M_\rho^y \rho_B^x). \tag{5.75}$$

We define the *accessible information* of the quantum state ensemble as $E = \{p_x, \rho_B^x\}$ as

$$I_{\mathrm{acc}}(E) = \max_{M \text{ POVM}} I(X;Y), \tag{5.76}$$

where $I(X;Y) = H(X) + H(Y) - H(XY)$ is the classical mutual information of the classical state

$$\rho_{XY} = \sum_x p(x)|x\rangle\langle x| \otimes \sum_y p(y|x)|y\rangle\langle y|, \tag{5.77}$$

with $p(y|x)$ as defined above.

Proposition 47 (Holevo bound)**.** *With the above definitions, for every quantum state ensemble $E = \{p_x, \rho_B^x\}$ we have*

$$I_{acc}(E) \leq \chi(E) = I(X;B)_\rho, \tag{5.78}$$

where $\rho_{XB} = \sum_x p_x |x\rangle\langle x|_x \otimes \rho_B^x$

Proof. The measurement defined by the POVM $M = \{M_B^y\}_y$ can be written as a quantum channel $\mathcal{M}\colon B \to Y$,

$$\mathcal{M}(\sigma_B) = \sum_y \mathrm{tr}(M_y \sigma_B)|y\rangle\langle y|. \tag{5.79}$$

For a state ρ_B^x from the ensemble E, this leads via eq. (5.75) to

$$\mathcal{M}(\rho_B^x) = \sum_y p(y|x)|y\rangle\langle y|. \tag{5.80}$$

Now the Holevo bound is a simple application of data-processing: For any POVM M and channel $\mathcal{M}$ as defined above,

$$I(X;Y) \leq I(X;B). \tag{5.81}$$

Hence, this also holds for the optimal measurement achieving the accessible information, so that $I_{\mathrm{acc}}(E) \leq \chi(E)$. □

Corollary 48. *n qubits can encode at most n classical bits.*

Proof. We want to choose a POVM that defines a random variable Y such that the mutual information, $I(X;Y)$, is maximal.[10]

By the Holevo bound, Prop. 47,

$$I_{\text{acc}}(E) \leq I(X;B) \leq \log|B|. \tag{5.82}$$

The state space of n qubits is $\mathcal{H}_B = (\mathbb{C}^2)^{\otimes n}$, which has dimension $|B| = 2^n$. Thus, $\log|B| = n$ and the best we can do is encode X in an orthonormal basis of $|B|$. □

[10] Encoding classical information in qubits means to select quantum states $\{\rho_B^x\}$ such that measuring B reveals X.

Conditional mutual information

The next entropic quantity we will consider is the conditional mutual information (CMI):[11]

$$\begin{aligned} I(A;B|C) &= S(A|C) + S(B|C) - S(AB|C), && (5.83)\\ &= S(AC) + S(BC) - S(C) - S(ABC), && (5.84)\\ &= S(A|C) - S(A|BC), && (5.85)\\ &= I(AC;B) - I(C;B). && (5.86) \end{aligned}$$

[11] Recall that the mutual information is defined as $I(A;B) = S(A) + S(B) - S(AB)$.

Operational interpretations include:

1. quantum state redistribution
2. erasure of correlations
3. quantum Markov chains[12]

[12] A lot more on this in the coming lectures!

Proposition 49. *The CMI satisfies the following properties:*

1. *Dimension bound:*

$$0 \leq I(A;B|C) \leq 2\log\min\{|A|,|B|\} \tag{5.87}$$

2. *Chain rule:*

$$I(A;BC) = I(A;B) + I(A;C|B) \tag{5.88}$$

3. *Classical conditioning:*

$$I(A;B|X)_\rho = \sum_x p_x I(A;B)_{\rho^x} \tag{5.89}$$

for QQC-states $\rho_{ABX} = \sum_x p_x \rho_{AB}^x \otimes |x\rangle\langle x|_X$.

4. *Duality: Let* $|\psi\rangle_{ABCD}$ *be pure, then*

$$I(A;B|C)_\psi = I(A;B|D) \tag{5.90}$$

Proof. 1. We have

$$I(A;B|C) = S(AC) + S(BC) - S(C) - S(ABC) \geq 0 \tag{5.91}$$

by strong sub-additivity. For an upper bound, observe that

$$\begin{aligned} I(A;B|C) &= I(AC;B) - I(C;B), && (5.92)\\ &\leq I(AC;B), && (5.93)\\ &\leq 2\log|B|. && (5.94) \end{aligned}$$

Similarly,

$$\begin{aligned} I(A;B|C) &= I(A;BC) - I(A;C), && (5.95)\\ &\leq I(A;BC), && (5.96)\\ &\leq 2\log|A|. && (5.97) \end{aligned}$$

2. Holds by direct calculation.

3. This follows from writing $I(A;B|X) = S(A|X) + S(B|X) - S(AB|X)$ and using part 6) of Prop. 44.

4. Lastly, we have[13]

$$\begin{aligned} I(A;B|C)_\psi &= S(A|C) - S(A|BC), && (5.98)\\ &= -S(A|BD) + S(A|D), && (5.99)\\ &= I(A;B|D), && (5.100) \end{aligned}$$

which concludes the proof. □

[13] Note that the second equality holds by part 2) of Prop. 44.

A fundamental question for us is the following: what does $I(A;B|C) = 0$ imply for the state ρ_{ABC}? In order to get a feeling for this question, we first take a brief detour discussing classical Markov chains.

5.2 *Detour: classical Markov chains*

let X, Y, Z be classical random variables with joint probability distribution $p(x,y,z)$. The RVs X, Y, Z form a Markov chain, denoted $X \to Y \to Z$, if X and Z are independent when conditioned on Y. This is equivalent to the following conditions:[14]

1. $p(xz|y) = p(x|y)p(z|y)$, where $p(x|y) = \frac{p(xy)}{p(y)}$.
2. $I(X;Z|Y) = 0$, where $I(X;Z|Y) = \sum_y p_y I(X;Z)_{p(xz|y)}$.
3. There exists a stochastic matrix $W(z|y)$ such that

$$p(xyz) = w(z|y)p(xy) \tag{5.101}$$

[14] Recall Bayes' Theorem: $p_{xy} = p_{x|y}p_y$ and $p_{xyz} = p_{xz|y}p_y$. See 3Blue1Brown for a great refresher on Bayes' Theorem.

5.3 *Quantum Markov Chains*

In analogy with characterization 3 of a classical Markov chain, we make the following definition in the quantum setting:[15]

[15] Informally speaking, a tripartite quantum state ρ_{ABC} is a quantum Markov chain if we can recover C from B alone.

Definition 50. A tripartite quantum state ρ_{ABC} is called a *quantum Markov chain*, denoted $A \to B \to C$, if there exists a channel $\mathcal{R}\colon B \to BC$ such that

$$\rho_{ABC} = (\mathbb{1}_A \otimes \mathcal{R})(\rho_{AB}) \tag{5.102}$$

This definition allows us to state the following important theorem.

Theorem 51. *ρ_{ABC} is a quantum Markov chain $A \to B \to C$ iff $I(A;C|B) = 0$.*

The sufficiency of the Markov chain condition for $I(A;C|B) = 0$ is a simple application of data-processing:

Proof of ($\Rightarrow$) in Thm. 51. Let $\mathcal{R}\colon B \to BC$ be the quantum channel satisfying

$$\rho_{ABC} = (\mathbb{1}_A \otimes R)(\rho_{AB}), \tag{5.103}$$

which implies $\rho_{BC} = \mathcal{R}(\rho_B)$. We then have

$$\begin{aligned} D(\rho_{ABC}\|\rho_A \otimes \rho_{BC}) &\geq D(\rho_{AB}\|\rho_A \otimes \rho_B), &(5.104)\\ &\geq D(\rho_{ABC}\|\rho_A \otimes \rho_{BC}), &(5.105)\end{aligned}$$

where the first inequality holds by DPI with respect to $\mathcal{N} = \mathrm{tr}_C(\cdot)$ and the second due to DPI with respect to $\mathcal{N} = (\mathbb{1}_A \otimes \mathcal{R})(\cdot)$. This implies that

$$D(\rho_{ABC}\|\rho_A \otimes \rho_{BC}) = D(\rho_{AB}\|\rho_A \otimes \rho_B), \tag{5.106}$$

and expanding both sides as linear combination of entropies gives

$$-S(ABC) + S(A) + S(BC) = -S(AB) + S(A) + S(B), \tag{5.107}$$

and hence $I(A;C|B) = 0$. as desired. □

For the direction ($\Leftarrow$) in Thm. 51, note that $I(A;C|B) = 0$ holds if and only if

$$D(\rho_{ABC}\|\rho_A \otimes \rho_{BC}) = D(\rho_{AB}\|\rho_A \otimes \rho_B). \tag{5.108}$$

We have *inequality* in this equation due to data-processing, and hence we are interested in when *equality* holds in the data-processing inequality in general. We state the following result proved by Petz[16] giving a necessary and sufficient condition for equality in the data-processing inequality for the partial trace:

[16] Dénes Petz. Monotonicity of quantum relative entropy revisited. *Reviews in Mathematical Physics*, 15(01):79–91, mar 2003. DOI: 10.1142/S0129055X03001576

Proposition 52. *We have $D(\rho_{AB}\|\sigma_{AB}) = D(\rho_A\|\sigma_A)$ if and only if the channel $\mathcal{R}\colon A \to AB$ defined by*

$$\mathcal{R}_\sigma(X_A) = \sigma_{AB}^{1/2}\left(\sigma_A^{-1/2} X_A \sigma_A^{-1/2} \otimes \mathbb{1}_B\right)\sigma_{AB}^{1/2} \tag{5.109}$$

satisfies $\rho_{AB} = \mathcal{R}_\sigma(\rho_A)$.

Note that we always have $\mathcal{R}_\sigma(\sigma_A) = \sigma_{AB}$ by construction of $\mathcal{R}_\sigma$. The fact that $\mathcal{R}$ is a valid quantum channel (on the support of σ_A) is a simple exercise. The proof of Prop. 52 makes use of the modular operator formalism and functional calculus and can be found in either Petz's paper[17] or the lecture notes of a previous version of this course. Prop. 52 asserts the missing implication in the proof of Thm. 51:

[17] Dénes Petz. Monotonicity of quantum relative entropy revisited. *Reviews in Mathematical Physics*, 15(01):79–91, mar 2003. DOI: 10.1142/S0129055X03001576

Proof of ($\Leftarrow$) in Thm. 51. By the arguments given in the proof of ($\Rightarrow$) above (and in the proof of strong subadditivity in Prop. 43), the condition $I(A;C|B) = 0$ is equivalent to

$$D(\rho_{ABC}\|\rho_A \otimes \rho_{BC}) = D(\rho_{AB}\|\rho_A \otimes \rho_B). \tag{5.110}$$

Hence, Prop. 52 implies that the channel $\mathcal{R}\colon AB \to ABC$ acting as

$$\begin{aligned}\mathcal{R} = (\rho_A \otimes \rho_{BC})^{1/2}&\left((\rho_A \otimes \rho_B)^{-1/2} X_{AB} (\rho_A \otimes \rho_B)^{-1/2} \otimes \mathbb{1}_C\right)\\ &\times (\rho_A \otimes \rho_{BC})^{1/2}\end{aligned} \tag{5.111}$$

satisfies $\rho_{ABC} = \mathcal{R}(\rho_{AB})$. Assuming w.l.o.g. that ρ_A has full support, we observe that the powers of ρ_A in (5.111) cancel out. This gives us a map $\widetilde{\mathcal{R}}_{\rho_{BC}}\colon B \to BC$ defined just as in (5.109) satisfying $\rho_{ABC} = (\mathrm{id}_A \otimes \widetilde{\mathcal{R}}_{\rho_{BC}})(\rho_{BC})$. In other words, ρ_{ABC} forms a quantum Markov chain $A \to B \to C$. □

We can generalize the equality condition for DPI of Prop. 52 to arbitrary quantum channels:

Proposition 53. *Let $\rho \in B(\mathcal{H})$ be a quantum state, $\sigma \in B(\mathcal{H})$ a positive semidefinite operator, and $\mathcal{N}\colon B(\mathcal{H}) \to B(\mathcal{K})$ be a quantum channel. Then we have $D(\rho\|\sigma) = D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$ if and only if the channel $\mathcal{R}_{\sigma,\mathcal{N}}\colon B(\mathcal{K}) \to B(\mathcal{H})$ defined by*

$$\mathcal{R}_{\sigma,\mathcal{N}}(X) = \sigma^{1/2}\mathcal{N}^\dagger\left(\mathcal{N}(\sigma)^{1/2} X \mathcal{N}(\sigma)^{1/2}\right)\sigma^{1/2} \tag{5.112}$$

satisfies $\rho = \mathcal{R}_{\sigma,\mathcal{N}}(\mathcal{N}(\rho))$.

Proof. ($\Leftarrow$) Assuming w.l.o.g. that σ has full support, the channel $\mathcal{R}_{\sigma,\mathcal{N}}$ satisfies $\mathcal{R}_{\sigma,\mathcal{N}}(\mathcal{N}(\sigma)) = \sigma$ by construction and $\rho = \mathcal{R}_{\sigma,\mathcal{N}}(\mathcal{N}(\rho))$ by assumption. Hence $D(\rho\|\sigma) = D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$ follows from applying

the data-processing inequality twice, once with respect to $\mathcal{N}$ and another time with respect to $\mathcal{R}_{\sigma,\mathcal{N}}$.

($\Rightarrow$) Let $V\colon \mathcal{H} \to \mathcal{K} \otimes \mathcal{K}'$ be a Stinespring isometry for $\mathcal{N}$, that is, $\mathcal{N}(X) = \operatorname{tr}_{\mathcal{K}'} VXV^\dagger$. Then

$$D(\rho\|\sigma) = D(V\rho V^\dagger \| V\sigma V^\dagger) = D(\operatorname{tr}_{\mathcal{K}'} V\rho V^\dagger \| \operatorname{tr}_{\mathcal{K}'} V\sigma V^\dagger), \tag{5.113}$$

where the first equality follows from the isometric invariance of $D(\cdot\|\cdot)$ and the second equality is the assumed equality in the data-processing inequality. Applying Prop. 52 gives a map $\mathcal{R}\colon \mathcal{K} \to \mathcal{K} \otimes \mathcal{K}'$ defined as

$$\begin{aligned} Y \mapsto (V\sigma V^\dagger)^{1/2} \left(\left(\operatorname{tr}_{\mathcal{K}'} V\sigma V^\dagger \right)^{-1/2} Y \left(\operatorname{tr}_{\mathcal{K}'} V\sigma V^\dagger \right)^{-1/2} \otimes \mathbb{1}_{\mathcal{K}'} \right) \\ \times (V\sigma V^\dagger)^{1/2}. \end{aligned} \tag{5.114}$$

We have $(V\sigma V^\dagger)^{1/2} = V\sigma^{1/2}V^\dagger$. Furthermore, the adjoint of the map $\mathcal{N} = \operatorname{tr}_{\mathcal{K}'} V \cdot V^\dagger$ is given by $\mathcal{N}^\dagger = V^\dagger(\cdot \otimes \mathbb{1}_{\mathcal{K}'})$. Hence, the channel in (5.114) can be written as

$$\mathcal{R}\colon Y \mapsto V\sigma^{1/2}\mathcal{N}^\dagger(\mathcal{N}(\sigma)^{-1/2} Y \mathcal{N}(\sigma)^{-1/2})\sigma^{1/2}V^\dagger. \tag{5.115}$$

Defining $\mathcal{R}_{\sigma,\mathcal{N}}$ as in (5.112) then shows that

$$V\rho V^\dagger = \mathcal{R}(\mathcal{N}(\rho)) = V\mathcal{R}_{\sigma,\mathcal{N}}(\mathcal{N}(\rho))V^\dagger, \tag{5.116}$$

and canceling the isometry on each side yields the desired result. □

Prop. 53 has some interesting applications. We first use it to prove the "only if" direction in the statement that $D(\rho\|\sigma) = 0$ iff $\rho = \sigma$ (for quantum states ρ, σ):

Corollary 54. *Let ρ, σ be quantum states. Then $D(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$.*

Proof. We have $\operatorname{tr}\rho = 1 = \operatorname{tr}\sigma$, so $D(\rho\|\sigma) \geq D(1\|1) = 0$ by data-processing. In this case, the recovery map $\mathcal{R}_{\sigma,\mathrm{tr}}\colon \mathbb{C} \to B(\mathcal{H})$ from Prop. 53 assumes the simple form $\mathcal{R}_{\sigma,\mathrm{tr}}(c) = c\sigma$ for any $c \in \mathbb{C}$ (using the fact that $\mathrm{tr}^\dagger(c) = c\mathbb{1}_{\mathcal{H}}$). Hence, $D(\rho\|\sigma) = 0$ implies

$$\rho = \mathcal{R}_{\sigma,\mathrm{tr}}(\operatorname{tr}\rho) = \mathcal{R}_{\sigma,\mathrm{tr}}(1) = \sigma \tag{5.117}$$

by Prop. 53, which concludes the proof. □

A more interesting application is to determine when the accessible information of a quantum state ensemble achieves the maximal value given by the Holevo information:

Proposition 55. *The accessible information $I_{acc}(E)$ of a quantum state ensemble $E = (p_x, \rho_x)_x$ achieves its maximal value $I_{acc}(E) = \chi(E)$ iff the ensemble states commute pairwise, $[\rho_x, \rho_{x'}] = 0$ for all x, x'.*

Proof. ($\Leftarrow$) We first prove the easy direction. If the ρ_x commute pairwise, they can be simultaneously diagonalized with respect to some ONB $\{|y\rangle\}$ of the Hilbert space, giving

$$\rho_x = \sum_y p_{y|x} |y\rangle\langle y| \tag{5.118}$$

for some eigenvalues $p_{y|x}$. Defining $p_{x,y} = p_{y|x} p_x$, it then follows that the c-q ensemble state ρ_{XB} is actually classical,

$$\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \sum_y p_{y|x} |y\rangle\langle y| = \sum_{x,y} p_{x,y} |xy\rangle\langle xy|, \tag{5.119}$$

and hence $I_{\text{acc}}(E) = I(X;Y) = I(X;B) = \chi(E)$ using the obvious optimal measurement.

($\Rightarrow$) We start with the observation that any measurement given in terms of a POVM $M = \{M_y\}$ can be realized as a projective measurement on a larger space and an isometric embedding of the measured quantum state into that space.[18] Since the relative entropy is invariant with respect to isometries, we can w.l.o.g. assume that the measurement $M = \{M_y\}$ achieving $I_{\text{acc}}(E) = \chi(E)$ is projective, that is, $M_x M_y = \delta_{x,y} M_x$ and $\sum_x M_x = \mathbb{1}$. The corresponding quantum channel is the measurement channel

$$\mathcal{M}(\rho) = \sum_y \operatorname{tr}(M_y \rho) |y\rangle\langle y|, \tag{5.120}$$

and a straightforward calculation shows that the adjoint channel is given by

$$\mathcal{M}^\dagger(\tau) = \sum_y \langle y|\tau|y\rangle M_y. \tag{5.121}$$

Recall that equality in the Holevo bound can be rewritten as

$$\chi(E) = D(\rho_{XB} \| \rho_X \otimes \rho_B) = D(\rho_{XY} \| \rho_X \otimes \rho_Y) = I_{\text{acc}}(E), \tag{5.122}$$

where $\rho_{XB} = \sum p_x |x\rangle\langle x|_X \otimes \rho_B^x$ and $\rho_{XY} = (\mathrm{id}_X \otimes \mathcal{M})(\rho_{XB})$. The necessary and sufficient condition for equality in (5.122) from Prop. 53 gives

$$\begin{aligned} \rho_{XB} &= \mathcal{R}_{\rho_X \otimes \rho_B, \mathcal{M}^\dagger}(\mathrm{id}_X \otimes \mathcal{M}(\rho_{XB})) && (5.123) \\ &= (\rho_X \otimes \rho_B)^{1/2} (\mathrm{id}_X \otimes \mathcal{M}^\dagger)\Big((\rho_X \otimes \rho_Y)^{-1/2} \rho_{XY} (\rho_X \otimes \rho_Y)^{-1/2} \Big) \\ &\quad \times (\rho_X \otimes \rho_B)^{1/2}. && (5.124) \end{aligned}$$

[18] This is known as Naimark's dilation theorem and can be found e.g. in Michael Wolf's lecture notes on quantum channels.

Note that the factors of ρ_X cancel out on the RHS of this equation. For the middle term, we have

$$(\mathbb{1}_X \otimes \rho_Y^{-1/2})\rho_{XY}(\mathbb{1}_X \otimes \rho_Y^{-1/2}) = \sum_x p_x |x\rangle\langle x|_X \otimes \sum_y \frac{p_{y|x}}{p_y} |y\rangle\langle y|, \tag{5.125}$$

and using (5.121),

$$(\mathrm{id}_X \otimes \mathcal{M}^\dagger)\Big((\mathbb{1}_X \otimes \rho_Y)^{-1/2}\rho_{XY}(\mathbb{1}_X \otimes \rho_Y)^{-1/2}\Big) = \sum_x p_x |x\rangle\langle x|_X \otimes \sum_y \frac{p_{y|x}}{p_y} M_y. \tag{5.126}$$

Substituting this in (5.124) gives

$$\rho_{XB} = \sum_x p_x |x\rangle\langle x|_X \otimes \sum_y \frac{p_{y|x}}{p_y} \rho_B^{1/2} M_y \rho_B^{1/2}, \tag{5.127}$$

and hence

$$\rho_B^x = \sum_y \frac{p_{y|x}}{p_y} \rho_B^{1/2} M_y \rho_B^{1/2} \quad \text{for all } x. \tag{5.128}$$

Rearranging this, we see that

$$\rho_B^{-1/2}\rho_B^x\rho_B^{-1/2} = \sum_y \frac{p_{y|x}}{p_y} M_y \tag{5.129}$$

is a spectral decomposition of the operator $\rho_B^{-1/2}\rho_B^x\rho_B^{-1/2}$, and hence taking the square root we get

$$\left(\rho_B^{-1/2}\rho_B^x\rho_B^{-1/2}\right)^{1/2} = \sum_y \frac{p_{y|x}^{1/2}}{p_y^{1/2}} M_y. \tag{5.130}$$

Now recall that $p_{y|x} = \mathrm{tr}(M_y\rho_B^x)$ and

$$p_y = \sum p_x p_{y|x} = \mathrm{tr}\left(M_y \textstyle\sum_x p_x \rho_B^x\right) = \mathrm{tr}(M_y\rho_B). \tag{5.131}$$

Using (5.130), we calculate:

$$\mathrm{tr}\left(\rho_B(\rho_B^{-1/2}\rho_B^x\rho_B^{-1/2})^{1/2}\right) = \sum_y \frac{p_{y|x}^{1/2}}{p_y^{1/2}} \mathrm{tr}(\rho_B M_y) \tag{5.132}$$

$$= \sum_y p_{y|x}^{1/2} p_y^{1/2} \tag{5.133}$$

$$= \sum_y \left[\mathrm{tr}(M_y\rho_B^x)\right]^{1/2} \left[\mathrm{tr}(M_y\rho_B)\right]^{1/2}. \tag{5.134}$$

Writing $\mathrm{tr}(M_y\rho_B^x) = \mathrm{tr}[(\rho_B^x)^{1/2}M_yM_y(\rho_B^x)^{1/2}]$ and similarly for the other term, we may apply the matrix-version of the Cauchy-Schwarz inequality, $\mathrm{tr}(A^\dagger A)^{1/2}\,\mathrm{tr}(B^\dagger B)^{1/2} \geq |\mathrm{tr}(A^\dagger B)|$, to each term in the sum

in (5.134), leading to

$$\operatorname{tr}\left(\rho_B(\rho_B^{-1/2}\rho_B^x\rho_B^{-1/2})^{1/2}\right) = \sum_y \left[\operatorname{tr}(M_y\rho_B^x)\right]^{1/2}\left[\operatorname{tr}(M_y\rho_B)\right]^{1/2} \tag{5.135}$$

$$\geq \sum_y \left|\operatorname{tr}\left((\rho_B^x)^{1/2}M_yM_y\rho_B^{1/2}\right)\right| \tag{5.136}$$

$$\geq \left|\operatorname{tr}\left((\rho_B^x)^{1/2}\textstyle\sum_y M_y\rho_B^{1/2}\right)\right| \tag{5.137}$$

$$= \operatorname{tr}\left((\rho_B^x)^{1/2}\rho_B^{1/2}\right). \tag{5.138}$$

Since the square root is an operator-concave function,[19] the right-hand side of (5.138) is bounded from below by 1,

$$\operatorname{tr}\left((\rho_B^x)^{1/2}\rho_B^{1/2}\right) = \operatorname{tr}\left((\rho_B^x)^{1/2}\left(\textstyle\sum_x p_x\rho_B^x\right)^{1/2}\right) \tag{5.139}$$

$$\geq \textstyle\sum_x p_x \operatorname{tr}((\rho_B^x)^{1/2}(\rho_B^x)^{1/2}) \tag{5.140}$$

$$= \textstyle\sum_x p_x \operatorname{tr}(\rho_B^x) \tag{5.141}$$

$$= \textstyle\sum_x p_x \tag{5.142}$$

$$= 1. \tag{5.143}$$

[19] That is, $(\lambda X + (1-\lambda)Y)^{1/2} \geq \lambda X^{1/2} + (1-\lambda)Y^{1/2}$ for any $\lambda \in [0,1]$ and X, Y Hermitian.

On the other hand, the square root is also operator monotone,[20] and hence from $\rho_B^x \leq \mathbb{1}$ and $\rho_B^{-1/2}\rho_B^x\rho_B^{-1/2} \leq \rho_B^{-1}$ we then get

$$(\rho_B^{-1/2}\rho_B^x\rho_B^{-1/2})^{1/2} \leq \rho_B^{-1/2}, \tag{5.144}$$

implying an upper bound of 1 for the left-hand side of (5.138),

$$\operatorname{tr}\left(\rho_B(\rho_B^{-1/2}\rho_B^x\rho_B^{-1/2})^{1/2}\right) \leq \operatorname{tr}\rho_B^{1/2} \leq \operatorname{tr}\rho_B = 1. \tag{5.145}$$

[20] If $X \leq Y$ then $X^{1/2} \leq Y^{1/2}$.

It follows that we have equality in each of the Cauchy-Schwarz inequalities (5.136), which occurs if and only if

$$M_y(\rho_B^x)^{1/2} = \lambda_{xy}M_y\rho_B^{1/2} \tag{5.146}$$

for some $\lambda_{xy} \in \mathbb{C}$. Taking traces on both sides shows that actually $\lambda_{xy} \geq 0$ (and in particular they are real). Summing over y in (5.146) gives $(\rho_B^x)^{1/2} = \sum_y \lambda_{xy}M_y\rho_B^{1/2}$, and hence the operator $(\rho_B^x)^{1/2}\rho_B^{-1/2} = \sum_y \lambda_{xy}M_y$ is Hermitian (since the M_y are and $\lambda_{xy} \geq 0$). Therefore,

$$(\rho_B^x)^{1/2}\rho_B^{-1/2} = \rho_B^{-1/2}(\rho_B^x)^{1/2}. \tag{5.147}$$

We are almost done: Recall the operators

$$\sigma_x \equiv \rho_B^{-1/2}\rho_B^x\rho_B^{-1/2} = \sum_y \frac{p_{y|x}}{p_y}M_y, \tag{5.148}$$

which commute because of the orthogonality of the M_y. Then,

$$\sigma_x\sigma_{x'} = \rho_B^{-1/2}\rho_B^x\rho_B^{-1/2}\rho_B^{-1/2}\rho_B^{x'}\rho_B^{-1/2} \tag{5.149}$$

$$= \rho_B^{-1}\rho_B^x\rho_B^{x'}\rho_B^{-1} \tag{5.150}$$

by using (5.147) several times. Similarly, $\sigma_{x'}\sigma_x = \rho_B^{-1}\rho_B^{x'}\rho_B^x\rho_B^{-1}$. Then $\sigma_x\sigma_{x'} = \sigma_{x'}\sigma_x$ finally yields the claim that $\rho_B^x\rho_B^{x'} = \rho_B^{x'}\rho_B^x$ for all x, x'. □

Bibliography

Howard Barnum, Emanuel Knill, and Michael A. Nielsen. On quantum fidelities and channel capacities. *IEEE Transactions on Information Theory*, 46(4):1317–1329, 2000. DOI: 10.1109/18.850671.

Salman Beigi and Peter W. Shor. On the complexity of computing zero-error and holevo capacity of quantum channels. *arXiv preprint*, 2007.

Leonid Gurvits. Classical deterministic complexity of edmonds' problem and quantum entanglement. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 10–19, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581136749. DOI: 10.1145/780542.780545. URL https://doi.org/10.1145/780542.780545.

Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998. DOI: 10.1109/18.651037.

Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Mixed-state entanglement and distillation: Is there a "bound" entanglement in nature? *Physical Review Letters*, 80:5239–5242, 6 1998. DOI: 10.1103/PhysRevLett.80.5239. URL https://link.aps.org/doi/10.1103/PhysRevLett.80.5239.

Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269:107–136, 2007.

Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996. ISSN 0375-9601. DOI: 10.1016/S0375-9601(96)00706-2. URL http://www.sciencedirect.com/science/article/pii/S0375960196007062.

Dénes Petz. Monotonicity of quantum relative entropy revisited. *Reviews in Mathematical Physics*, 15(01):79–91, mar 2003. DOI: 10.1142/s0129055x03001576.

Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56: 131–138, 1997. DOI: 10.1103/PhysRevA.56.131.

Peter W. Shor. Additivity of the classical capacity of entanglement-breaking quantum channels. *Journal of Mathematical Physics*, 43(9): 4334–4340, 2002. DOI: 10.1063/1.1498000.

Peter W. Shor. Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics*, 246(3):473–473, 4 2004. ISSN 1432-0916. DOI: 10.1007/s00220-004-1071-1. URL https://doi.org/10.1007/s00220-004-1071-1.

Michael M Wolf and David Perez-Garcia. Quantum capacities of channels with small environment. *Physical Review A*, 75(1):012303, 2007.

William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

Index