MARCO TOMAMICHEL, RAHUL JAIN

TOPICS IN QUANTUM INFORMATION THEORY

QT5104

These notes can always be improved. Any comments that help to improve these notes are very much appreciated. We thank Ian T. George for revising the chapter on entropies and contributing the chapter on quantum key distribution. We would like to thank all the students of NUS QT5104 in 2024/25 Semester II who have contributed writing an early draft of these notes.

http://www.marcotom.info/teaching/

https://www.comp.nus.edu.sg/~rahul/

Latest update, May 2026

Contents

$\mathcal{L}(V,W)$	set of linear map from V to W
$\mathcal{L}(V)$	set $\mathcal{L}(V,V)$
$\mathcal{H}(V)$	set of self-adjoint operators in $\mathcal{L}(V)$
$\langle\cdot,\cdot\rangle$	inner product
$\otimes$	tensor product
span	vector span
vec	the isomorphism $\mathcal{L}(V,W) \to \mathcal{L}(V \otimes W)$
$\overline{z}$	complex conjugate of z
$X^\dagger$	adjoint/conjugate transpose of X
X^T	transpose (basis dependent)
Tr	trace
Tr_A	partial trace over the system A
det	matrix determinant
Pr	probability
$\mathbb{1}$	the identity matrix
$[A,B]$	commutator of A and B, $AB - BA$
$\{A,B\}$	anti-commutator of A and B, $AB + BA$
$\mathbf{1}\{x = y\}$	indicator function, 1 if $x = y$ and 0 otherwise, so that, for example, $\mathbf{1}\{x = y\} + \mathbf{1}\{x \neq y\} = 1$
δ_{xy}	shorthand for $\mathbf{1}\{x = y\}$
log	logarithm; to base 2 here, i.e. $\log = \log_2$
$\mathrm{F}(\cdot,\cdot)$	fidelity
$\mathrm{D}(\cdot\|\cdot)$	relative-entropy
$\mathrm{D}_{\max}(\cdot\|\cdot)$	max relative-entropy
$\mathrm{D}_H(\cdot\|\cdot)$	hypothesis testing relative-entropy
$\mathrm{H}(\cdot)$	entropy
$\mathrm{I}_{\max}(\cdot)$	max mutual information
$\rho \approx_\varepsilon \sigma$	$\mathrm{P}(\rho,\sigma) \leq \varepsilon$
$\langle\cdot,\cdot\rangle$	inner product
$\stackrel{\text{def}}{=}$	is defined as
$\mathcal{N}_{A\to B}$	a quantum channel from A to B

Table 1: Some basic notation used in these notes.

psd	positive semi-definite
td	trace distance
ONB	orthonormal basis
SVD	singaluar value decomposition
SD	spectral decomposition
CPTP	completely positive and trace-preserving
pmf	probability mass function
iff	if and only if

Table 2: Some abbreviations used in these notes.

1
Linear Algebra Foundations

In this lecture, we will cover the basics of linear algebra and Hilbert space, which are used to describe quantum objects such as states, density operators, observables, effects or channels. In fact, we will introduce a hierarchy of real and complex Hilbert spaces by constructing a new Hilbert space of linear operators acting on an underlying Hilbert space.

level	quantum object	field
1	pure states $\|\psi\rangle, \|\phi\rangle, \ldots$	$\mathbb{C}$
2	linear operators $X, L, K, \ldots$	$\mathbb{C}$
	observables, effects $M, N, \ldots$	$\mathbb{R}$
	quantum states $\rho, \sigma, \ldots$	$\mathbb{R}$
	pure states as linear operators $\|\psi\rangle, \|\phi\rangle, \ldots$	$\mathbb{C}$
3	channels $\mathcal{E}, \mathcal{F}, \ldots$	$\mathbb{R}$
4	super channels or combs (not covered)	$\mathbb{R}$

Table 1.1: Hierarchy of quantum objects. They will be introduced in Chapter 2.

1.1 Hilbert space

The general definition of a Hilbert space is a vector space with an inner product that makes it complete. Any finite-dimensional vector space with an inner product is complete, but this is not always true for infinite-dimensional vector space.

In this course, we will only deal with discrete systems and finite dimensions. A Hilbert space in finite dimension is simply a vector space with an inner product, or inner product space. However, since the name "Hilbert space" is so common in physics parlance we will stick with it.

The first ingredient is a vector space over a field $\mathbb{F}$. We will only use $\mathbb{R}$ or $\mathbb{C}$ as fields in this course.

Definition 1.1. *Let $\mathbb{F}$ be a field. A vector space over $\mathbb{F}$ is a set V with an addition operation $V \times V \to V$ and a scalar multiplication operation $\mathbb{F} \times V \to V$. The addition*

- *is associative:* $\forall u, v, w \in V,\ u + (v + w) = (u + v) + w$,
- *is commutative:* $\forall u, v \in V,\ u + v = v + u$,

- *has a* 0 *element:* $\forall v \in V,\ 0 + v = v$, *and*
- *has inverse elements:* $\forall v \in V, \exists(-v) \in V$ *such that* $v + (-v) = 0$.

The scalar multiplication

- *is distributive over the vector addition and the field addition:* $\forall \alpha, \beta \in \mathbb{F}, \forall u, v \in V,\ \alpha(u + v) = \alpha u + \alpha v$ *and* $(\alpha + \beta)v = \alpha v + \beta v$,
- *has the same identity element as the multiplicative identity of* $\mathbb{F}$*:* $\forall v \in V,\ 1v = v$, *and*
- *is compatible with scalar multiplication:* $\forall \alpha, \beta \in \mathbb{F}, \forall v \in V,\ \alpha(\beta v) = (\alpha\beta)v$.

The second ingredient is an inner product.

Definition 1.2. *An inner product of a vector space* V *over* $\mathbb{F}$ *is a map* $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{F}$ *which, following physics convention, satisfies:*

1. *Linearity in the second argument:* $\forall \alpha_1, \alpha_2 \in \mathbb{F},\ v_1, v_2, w \in V$, *we have* $\langle w, \alpha_1 v_1 + \alpha_2 v_2 \rangle = \alpha_1 \langle w, v_1 \rangle + \alpha_2 \langle w, v_2 \rangle$.
2. *Conjugate symmetry:* $\forall v, w \in V\ \langle w, v \rangle = \overline{\langle v, w \rangle}$.
3. *Positive definite:* $\langle v, v \rangle \geq 0,\ \forall v \in V$ *with equality iff* $v = 0$.

Conjugate linearity of the first element can be proved by applying Properties 1 and 2 from Definition 1.2, that is,

$$\langle \alpha_1 v_1 + \alpha_2 v_2, w \rangle = \bar{\alpha}_1 \langle v_1, w \rangle + \bar{\alpha}_2 \langle v_2, w \rangle . \tag{1.1}$$

Note that if the vector space is over $\mathbb{R}$, then the inner product is in fact linear in both arguments. This is because the conjugate symmetry is just the symmetry of the inner product as $\bar{r} = r$ iff $r \in \mathbb{R}$. The conjugate symmetry also ensures that $\langle v, v \rangle \in \mathbb{R}$ as $\langle v, v \rangle = \overline{\langle v, v \rangle}$ in this case.

The inner product allows for the concept of orthogonality and Hilbert spaces have orthonormal bases and a dimension.

Definition 1.3. *We say that two vectors* $v, w \in V$ *are* orthogonal *if* $\langle v, w \rangle = 0$. *A set of vectors* $\{v_i\}_i$ *is* orthonormal *if* $\forall i, j\ \langle v_i, v_j \rangle = \delta_{ij}$.

Moreover, $\{v_i\}_i$ *is an* orthonormal basis *(ONB) of* V *if it is orthonormal and if every element of* V *can be decomposed as a linear combination of elements of* $\{v_i\}_i$*, i.e., if*

$$V = \operatorname{span}\{v_i\}_i := \{v \in V :\ v = \sum_i \alpha_i v_i,\ \alpha_i \in \mathbb{F}\} . \tag{1.2}$$

The dimension *d of V is the number of elements of an ONB.*

There exist different ONBs for the same Hilbert space but all of them have the same number of elements. This follows from the property that if a vector space has a basis of n elements, any set of m vectors with $m > n$ is linearly dependent and thus is not orthonormal.

Using the inner product, we can also define a norm.

Definition 1.4. *A* norm *on a vector space V is a map $\|\cdot\| : V \to \mathbb{R}$ that satisfies:*

1. *Triangle inequality: $\|v + w\| \leq \|v\| + \|w\|$,*
2. *Absolute homogeneity: $\|\alpha v\| = |\alpha|\|v\|$, and*
3. *Positive definiteness: $\|v\| \geq 0$ with equality iff $v = 0$.*

for all $v, w \in V$ and $\alpha \in \mathbb{F}$.

From the inner product, we can define the canonical *norm*:

$$\|v\| = \sqrt{\langle v, v\rangle}. \tag{1.3}$$

The fact that this is indeed a norm follows immediately from the properties of the inner product. This is not the only norm that is useful, and we will get back to this in Section 1.3.

Beyond the triangular inequality, the *Cauchy-Schwarz* inequality holds:

$$\forall v, w \in V \; |\langle v, w\rangle| \leq \|v\| \cdot \|w\| . \tag{1.4}$$

The Cauchy-Schwarz inequality can be proven by looking at the polynomial $\|x + ty\|^2$ for $t \in \mathbb{R}, x, y \in V$ with $\langle x, y\rangle \in \mathbb{R}$ (we can ensure it by multiplying x and y with complex scalar). We can then use Cauchy-Schwarz inequality to prove the triangular inequality looking at the square of both sides of the inequality.

Example. *In quantum mechanics, it is common to denote an element of a Hilbert space as a ket $|\psi\rangle$. We introduce an ONB, the computational basis, using the notation $\{|i\rangle\}_{i=0}^{d-1}$ so we can write $|\phi\rangle = \sum_i \phi_i |i\rangle$ and $|\theta\rangle = \sum_i \theta_i |i\rangle$. This allows us to represent $|\psi\rangle$ as a column vector*

$$|\phi\rangle \sim \begin{pmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_{d-1} \end{pmatrix} \tag{1.5}$$

We also use row vectors, called bra, e.g.,

$$\langle\theta| \sim \begin{pmatrix} \overline{\theta_0}, & \overline{\theta_1}, & \ldots, & \overline{\theta_{d-1}} \end{pmatrix}. \tag{1.6}$$

The standard inner product for complex vectors is

$$\langle |\theta\rangle, |\phi\rangle \rangle = \langle \theta|\phi\rangle = \sum_i \overline{\theta}_i \phi_i \,, \tag{1.7}$$

where the second expression is simply writing the inner product in the so-called Dirac bracket notation.

1.2 Linear operators

An operator is a map from one vector space to another (possibly the same) mapping each element of the input space to an element of the output space. We are interested in the set of linear operators from V to W, which we denote by $\mathcal{L}(V, W)$ and use the shorthand $\mathcal{L}(V) = \mathcal{L}(V, V)$. A linear operator X acting on a vector space V over $\mathbb{F}$ satisfies

$$X(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 X v_1 + \alpha_2 X v_2 \tag{1.8}$$

for all $\alpha_1, \alpha_2 \in \mathbb{F}$ and $v_1, v_2 \in V$. As a consequence, a linear operator X is completely determined by its action on an ONB of V.

The space of linear operator is itself a vector space over the same field as the vector space it acts on. Namely, if $X_1, X_2 \in \mathcal{L}(V, W)$ and $\alpha_1, \alpha_2 \in \mathbb{F}$, we understand that linear combinations of linear operators act as $(\alpha_1 X_1 + \alpha_2 X_2) : V \ni v \mapsto \alpha_1 X_1 v + \alpha_2 X_2 v \in W$. This gives the set an additive structure. Moreover, linear operators in $X_1 \in \mathcal{L}(V, V')$ and $X_2 \in \mathcal{L}(V', V'')$ can be applied in sequence, i.e., $X_2 X_1 v$ is to be interpreted as first applying X_1 on v and then applying X_2 on the resulting vector in V'. This gives the set $\mathcal{L}(V)$ a multiplicative structure, and they in fact form a ring. We say that two operators $X, Y \in \mathcal{L}(V)$ commute if $XY = YX$.

Definition 1.5. *The* adjoint *operator of $X \in \mathcal{L}(V, W)$ is the unique operator, denoted $X^\dagger \in \mathcal{L}(W, V)$ satisfying:*

$$\forall w \in W,\ \forall v \in V,\ \langle w, Xv\rangle = \langle X^\dagger w, v\rangle.$$

An operator $H \in \mathcal{L}(V)$ is self-adjoint *or Hermitian if $H^\dagger = H$. We denote the set of Hermitian operators as $\mathcal{H}(V) \subset \mathcal{L}(V)$.*

For the uniqueness of the adjoint operator of X, suppose that there exist X_1 and X_2 such that $\forall v \in V,\ w \in W,\ \langle X_1 w, v\rangle = \langle X_2 w, v\rangle = \langle w, Xv\rangle$ then $\langle (X_1 - X_2)w, v\rangle = 0$ which by looking at a basis of each vector space gives $X_1 - X_2 = 0$ and $X_1 = X_2$.

The adjoint operators satisfy the following two properties, which follow immediately from the definition. For $X \in \mathcal{L}(V', V'')$ and $Y \in \mathcal{L}(V, V')$, we have

- $(X^\dagger)^\dagger = X$
- $(XY)^\dagger = Y^\dagger X^\dagger$

Example. *We can see kets as elements of $L(\mathbb{C}, V)$ and bras as elements of $L(V, \mathbb{C})$. That is, with an abuse of notation, we define the map $|\psi\rangle : \alpha \mapsto$*

$\alpha\,|\psi\rangle$, where the second $|\psi\rangle$ is an element of the Hilbert space. Similarly, the bras are maps of the form $\langle\psi| : |\phi\rangle \mapsto \langle\psi|\phi\rangle$. With this interpretation, the Dirac bracket notation becomes very natural. Moreover, since

$$\langle |\phi\rangle\,, |\psi\rangle\,\alpha\rangle = \alpha\langle\phi|\psi\rangle = \langle\langle\psi|\phi\rangle, \alpha\rangle, \tag{1.9}$$

we can deduce that $\langle\psi| = \big(\,|\psi\rangle\,\big)^\dagger$, compatible with our example in (1.6).

Definition 1.6. *The (multiplicative)* identity *operator on V, denoted 1_V is the map sending a vector to itself: $1_V : v \mapsto v$.*

The identity can be decomposed as $1_V|v\rangle = \sum_i |v_i\rangle\langle v_i, v\rangle$ for any ONB $\{|v_i\rangle\}_i$. This decomposition is extremely useful, as we will see.

Example. *In the bracket notation, the decomposition simply reads*

$$1_V = \sum_i |v_i\rangle\langle v_i|\,. \tag{1.10}$$

Next, we introduce isometries and unitaries, which are linear maps that preserve the inner product.

Definition 1.7. *An isometry $U \in \mathcal{L}(V,W)$ is a linear operator that satisfies $U^\dagger U = 1_V$. A unitary $U \in \mathcal{L}(V)$ is a linear operator that satisfies further $U^\dagger U = UU^\dagger = 1_V$. Unitaries are thus also isometries.*

Isometries preserve the inner product, i.e., $\langle Uv, Uw\rangle = \langle v, U^\dagger Uw\rangle = \langle v, w\rangle$. As a consequence, for any ONB $\{v_i\}_i$, an isometry U satisfies $\langle Uv_i, Uv_j\rangle = \langle v_i, U^\dagger Uv_j\rangle = \langle v_i, v_j\rangle = \delta_{ij}$. Thus, in particular, unitary operators map an ONB to another ONB on the same space, and conversely, every transformation from one ONB to another can be modelled as a unitary operator.

We have seen that the linear operators form a vector space, but to make it a Hilbert space we still need an inner product.

Definition 1.8. *We define the canonical* inner product *on $\mathcal{L}(V,W)$ as*

$$\langle X, Y\rangle = \sum_i \langle Xv_i, Yv_i\rangle$$

for any ONB $\{v_i\}_i$ of V.

This definition is a good inner product as the linearity in the second argument, the conjugate symmetry and the positive definiteness are inherited from the inner product over the vector space and the linearity of the operators. Positive definiteness is less obvious.

Proof (positive definiteness). Let $X \in \mathcal{L}(V,W)$ and $\{v_i\}_i$ be an ONB of the vector space. $\langle X,X\rangle = \sum_i \langle Xv_i, Xv_i\rangle = \sum_i \|Xv_i\|^2 \geq 0$ with equality iff $\forall i\ \|Xv_i\| = 0 \iff Xv_i = 0$. The action of X is defined by its action on the ONB $\{v_i\}_i$, so $X = 0$. □

The definition of the inner product appears to depend on our choice of ONB. However, we will see below that it is in fact basis-independent.

Closely related to the inner product is the trace.

Definition 1.9. *For $X \in \mathcal{L}(V)$ we define the* trace *as* $\mathrm{Tr}(X) = \langle 1_V, X\rangle$.

Using this and the definition of the adjoint, we can thus alternatively write $\langle X,Y\rangle = \mathrm{Tr}(X^\dagger Y)$. With the inner product again come the concepts of orthogonality and orthonormal bases.

What is $|i\rangle\langle j|$ in matrix representation? A matrix of all 0s except for 1 point at the ith row and jth column.

Example. *Given an ONB $\{|i\rangle\}_i$ of V, the linear operators $\{|i\rangle\langle j|\}_{i,j}$ form an ONB of $L(V)$. The coefficients in the decomposition $X = x_{i,j}|i\rangle\langle j|$ lead to a matrix representation*

$$X \sim \begin{pmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,d-1} \\ x_{1,0} & x_{1,1} & \cdots & \\ \vdots & \vdots & & \\ x_{d-1,0} & & & x_{d-1,d-1} \end{pmatrix}. \tag{1.11}$$

With the standard inner product in (1.7), the inner product for linear operators becomes $\langle X,Y\rangle = \mathrm{Tr}(X^\dagger Y)$ where $X^\dagger$ is the usual matrix conjugate transpose and Tr *the matrix trace. This is called the Hilbert-Schmidt inner product.*

The following lemma turns out to be useful. In the matrix case, it simply affirms the cyclicity of the trace.

Lemma 1.10. *For any $X,Y \in \mathcal{L}(V,W)$, we have $\langle X,Y\rangle = \langle Y^\dagger, X^\dagger\rangle$. In particular, the trace is cyclic, i.e., $\mathrm{Tr}(X^\dagger Y) = \mathrm{Tr}(YX^\dagger)$.*

Proof. Let $\{v_i\}$ be the ONB of the inner product.

$$\langle X,Y\rangle = \sum_i \langle Xv_i, Yv_i\rangle = \sum_i \langle Xv_i, \sum_j v_j\langle v_j, Yv_i\rangle\rangle \tag{1.12}$$

$$= \sum_{i,j} \langle Xv_i, v_j\rangle\langle v_j, Yv_i\rangle \tag{1.13}$$

$$= \sum_{i,j} \langle v_i, X^\dagger v_j\rangle\langle Y^\dagger v_j, v_i\rangle\,. \tag{1.14}$$

Examining the last expression, we see that $\langle X,Y\rangle = \langle Y^\dagger, X^\dagger\rangle$.

To show the cyclicity of the trace, we simply note that $\langle 1_V, X^\dagger Y\rangle = \langle X,Y\rangle = \langle Y^\dagger, X^\dagger\rangle = \langle 1_W, YX^\dagger\rangle$. □

We can use this lemma to show that the inner product is independent of the choice of the ONB.

Proof (basis-independence of canonical inner product). Let $\{v_i\}_i$ and $\{w_i\}_i$ be two ONBs and U the change-of-basis unitary: $v_i \mapsto w_i$. Defining the inner product with regards to w_i as $\langle\cdot,\cdot\rangle'$, we find that

$$\langle X, Y\rangle' := \sum_i \langle Xw_i, Yw_i\rangle = \sum_i \langle XUv_i, YUv_i\rangle \tag{1.15}$$

$$= \langle XU, YU\rangle = \mathrm{Tr}(U^\dagger X^\dagger Y U) = \mathrm{Tr}(X^\dagger Y) = \langle X, Y\rangle\,, \tag{1.16}$$

where we used the cyclicity of the trace from Lemma 1.10. □

This completes the construction: we have built a new Hilbert space of linear operators acting on an underlying Hilbert space.

1.3 Operator decompositions

Operator decompositions are a useful tool in analysing linear operators. For self-adjoint operators, we have a spectral decomposition.

Definition 1.11. *Any Hermitian operator $H \in \mathcal{H}(V)$ has a* spectral decomposition *(SD) of the form*

$$H : v \mapsto \sum_{i=0}^{d-1} \lambda_i v_i \langle v_i, v\rangle\,, \tag{1.17}$$

where $\lambda_i \in \mathbb{R}$ and $\|v_i\| = 1$ for all i. Here $\{\lambda_i\}_i$ are called eigenvalues *and $\{v_i\}_i$ are called* eigenvectors *and form an ONB.*

We use the convention $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{d-1}$. The spectral decomposition is closely related to the eigenvalue equation. Indeed, using the decomposition we find $Hv_i = \lambda_i v_i$.

Example. *In the bracket notation, we may write*

$$H = \sum_{i=0}^{d-1} \lambda_i \,|\psi_i\rangle\langle\psi_i|\,, \tag{1.18}$$

where $\{|\psi_i\rangle\}_i$ are the eigenvectors. Expressing everything in an ONB $\{|i\rangle\}_i$, we can write $H = UDU^\dagger$ where $D = \sum_i \lambda_i |i\rangle\langle i|$ is diagonal (when expressed as a matrix in this ONB) and $U = \sum_i |\psi_i\rangle\langle i|$ is the unitary operator mapping $|i\rangle$ to $|\psi_i\rangle$.

Proof sketch. Consider the optimisation $\lambda_0 := \max_{v,\|v\|=1} \langle v, Hv\rangle$ and any vector v_0 that achieves this maximum. We now look at the function

$$f(t) = \langle v_t, Hv_t\rangle \quad \text{for} \quad v_t = \left(\sqrt{1-t^2}\right)v_0 + tv_\perp \tag{1.19}$$

where $v_\perp$ is any vector orthogonal to v_0. By the defining property of the optimiser the derivative must satisfy $f'(0) = 0$, which yields

$$\langle v_\perp, Hv_0\rangle + \langle v_0, Hv_\perp\rangle = 2\Re\langle v_\perp, Hv_0\rangle = 0 \tag{1.20}$$

Since this also holds for $v_t' = iv_t$, we get the same condition for the imaginary part. We can thus conclude that $\langle v_\perp, Hv_0\rangle = 0$. Since this holds for any $v_\perp$, it implies the eigenvalue equation $Hv_0 = \lambda_0 v_0$.

We can consider the optimisation with the additional restriction that v is in the space orthogonal to v_0 to get λ_1 and v_1, and eventually arrive at the decomposition by iterating the procedure. □

For operators that are not self-adjoint we still find a singular value decomposition, but now we need to work with two ONBs.

Definition 1.12. *Any operator $X \in \mathcal{L}(V,W)$ has a* singular value decomposition *(SVD) of the form*

$$X : v \mapsto \sum_{i=0}^{r-1} s_i w_i \langle v_i, v\rangle\,, \tag{1.21}$$

where $\{v_i\}_i$ and $\{w_i\}_i$ are orthonormal sets in V and W, respectively. The reals $s_i > 0$ are called singular values.

- *The rank r of X is the number of elements in this decomposition and satisfies $r \leq \min\{d_V, d_W\}$.*
- $\mathrm{span}\{v_i\}_i$ *is called the* support *of X.*
- *Its orthogonal complement is called the* kernel *of X.*
- $\mathrm{span}\{w_i\}_i$ *is called the* image *of X.*

We will use the convention $s_0 \geq s_1 \geq \cdots \geq s_{r-1} > 0$. If X is self-adjoint then we can see that the singular values are given by the absolute values $|\lambda_i|$, appropriately ordered.

Example. *In bracket notation for $X \in \mathcal{L}(V,W)$, we may write*

$$X \mapsto \sum_{i=0}^{r-1} s_i |w_i\rangle\langle v_i|, \tag{1.22}$$

for orthonormal sets $\{|w_i\rangle\}_i$ and $\{|v_i\rangle\}_i$. In matrix form we can write $X = UDV^\dagger$, where $D = \sum_i s_i\,|i\rangle\langle i|$ is a diagonal matrix and U and V are isometries mapping $|i\rangle$ to $|w_i\rangle$ and $|v_i\rangle$, respectively.

1.4 Hermitian operators

Hermitian and positive semi-definite operators play a particularly important role in quantum information as both quantum states and effects are in this category.

1.4.1 Positive semi-definite operators

Definition 1.13. *A Hermitian operator $H \in \mathcal{H}(V)$ is* positive semi-definite *(psd) iff $\langle v, Hv\rangle \geq 0, \forall v \in V$. We write $H \geq 0$ if H is psd.*

This notation should not be confused with meaning that H has only non-negative entries.

When we write $H \geq 0$, we implicitly say that H is self-adjoint.

Lemma 1.14. *The following are equivalent for any $H \in \mathcal{H}(V)$:*

1. *$H \geq 0$, i.e., H is psd;*
2. *$H = L^\dagger L$ for some $L \in \mathcal{L}(V)$;*
3. *The eigenvalues of H satisfy $\lambda_i \geq 0$.*

Proof. It suffices to show the following implications:

1 $\implies$ 3*:* From the spectral decomposition we find $\lambda_i = \langle v_i, Hv_i\rangle$, which is positive due to the psd condition.

3 $\implies$ 2*:* From the spectral decomposition $H = \sum_i \lambda_i |v_i\rangle\langle v_i|$ we construct $L = \sum_i \sqrt{\lambda_i}|v_i\rangle\langle v_i|$ as we know that $\lambda_i \geq 0$.

2 $\implies$ 1*:* Observe that $\langle v, Hv\rangle = \langle v, L^\dagger Lv\rangle = \langle Lv, Lv\rangle \geq 0$, where we used the positive definiteness of the inner product. □

Positive semi-definiteness induces a partial order on Hermitian operators, namely, we say $A \geq B$ iff $A - B \geq 0$. For example, the notation $0 \leq M \leq I$ implies that the eigenvalues of M are in $[0, 1]$.

Lemma 1.15. *For any linear operator X, we have $M \geq 0 \implies X^\dagger MX \geq 0$.*

Proof. We have $\langle v, X^\dagger MXv\rangle = \langle Xv, MXv\rangle$ using the definition of the adjoint operator. Thus, $\langle Xv, MXv\rangle = \langle w, Mw\rangle \geq 0$ renaming $Xv = w$ and using the assumption that $M \geq 0$. □

Lemma 1.16. *We have $M, N \geq 0 \implies \langle M, N\rangle = \mathrm{Tr}(MN) \geq 0$.*

Proof. $\mathrm{Tr}(MN) = \mathrm{Tr}(M^{\frac{1}{2}} N M^{\frac{1}{2}}) \geq 0$ as $M^{\frac{1}{2}} N M^{\frac{1}{2}}$ is psd due to Lemma 1.15 and the trace is thus also positive. □

1.4.2 Projectors

Definition 1.17. *Projectors are psd operators P that satisfies $P^2 = P$.*

In particular, projectors have only eigenvalues 1 or 0.

Example. *The operator $|\psi\rangle\langle\psi|$ when $\langle\psi|\psi\rangle = 1$ is a projector since $|\psi\rangle\langle\psi| \cdot |\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|$. More generally, $\sum_i |v_i\rangle\langle v_i|$ is a projector when $\{v_i\}_i$ are orthonormal.*

1.4.3 Functions on Hermitian operators

Applying a function on a self-adjoint operator should be understood as applying to the eigenvalues in the spectral decomposition.

Definition 1.18. *Let $f : \mathcal{S} \to \mathbb{R}$ be a function on some support S. We define $f(H)$ for any $H \in \mathcal{H}(V)$ with $H = \sum_{i=0}^{d-1} \lambda_i |v_i\rangle\langle v_i|$ as*

$$\sum_{i:\lambda_i \in \mathcal{S}} f(\lambda_i) |v_i\rangle\langle v_i| \,, \tag{1.23}$$

where the sum is restricted to those i for which $\lambda_i \in \mathcal{S}$.

The first example is the inverse function, $f : t \to \frac{1}{t}$. Then

$$H^{-1} = \sum_{i=0,\lambda_i>0}^{r-1} \frac{1}{\lambda_i} |v_i\rangle\langle v_i| \,. \tag{1.24}$$

This is called the Penrose or generalized inverse. In general, if a matrix has an eigenvalue $\lambda = 0$, it is not invertible.

We find that $H^{-1}H = \sum_{i=0,\lambda_i>0}^{d-1} |v_i\rangle\langle v_i|$ is the projector on the support of H. It gives the identity if $r = d$, that is if H is of full rank.

The second example is $f : t \to t^2$. Then, $H^2 = \sum_i \lambda_i^2 |v_i\rangle\langle v_i|$, same as the product HH. More generally, our definition is consistent with how one would usually evaluate polynomials on matrices.

Finally, we define the *modulus* of a matrix X as

$$|X| = \sqrt{X^\dagger X}, \tag{1.25}$$

and note that $|X|$ is psd. Clearly, using the SVD in (1.22), we see that

$$|X| = \sum_{i=0}^{r-1} s_i \, |v_i\rangle\langle v_i| \tag{1.26}$$

and, thus, $X = U|X|$ with $U = \sum_{i=0}^{r-1} |w_i\rangle\langle v_i|$, which is called the *polar decomposition.*

1.5 Tensor product spaces

In physics, we are often quite happy to talk about the single quantum system (either open or closed) and local evaluation and measurements. However, in quantum information, we usually want to talk about multiple systems or composite systems, for example to describe entanglement between two physical systems. We deal with composite systems using tensor products and tensor spaces.

Definition 1.19. *Let V, W be two Hilbert spaces on $\mathbb{F}$ with inner products $\langle\cdot,\cdot\rangle_V$ and $\langle\cdot,\cdot\rangle_W$ and ONBs $\{v_i\}_i$ and $\{w_i\}_i$, respetively. We*

construct a vector space $V \otimes W$ on $\mathbb{F}$ as

$$V \otimes W = \text{span}\{v_i \otimes w_j\}_{ij} \tag{1.27}$$

with inner product

$$\langle v_i \otimes w_j, v_{i'} \otimes w_{j'} \rangle_{V\otimes W} = \langle v_i, v_{i'} \rangle_V \langle w_j, w_{j'} \rangle_W = \delta_{ii'}\delta_{jj'} \tag{1.28}$$

and its sesquilinear extension to $V \otimes W$.

The sesquilinear extension simply enforces that for two general vectors

$$u = \sum_{i,j} \alpha_{ij} v_i \otimes w_j, \quad u' = \sum_{i',j'} \beta_{i'j'} v_{i'} \otimes w_{j'} \tag{1.29}$$

in $\text{span}\{v_i \otimes w_j\}_{ij}$, their inner product satisfies

$$\langle u, u' \rangle = \langle \sum_{i,j} \alpha_{ij} v_i \otimes w_j, \sum_{i',j'} \beta_{i'j'} v_{i'} \otimes w_{j'} \rangle \tag{1.30}$$

$$= \sum_{i,j,i',j'} \overline{\alpha_{ij}} \beta_{i'j'} \delta_{ii'} \delta_{jj'} \tag{1.31}$$

$$= \sum_{i,j} \overline{\alpha_{ij}} \beta_{ij} \,. \tag{1.32}$$

A general tensor product of two vectors $v \otimes w$ is defined via its basis decomposition, which associates with it an element of $V \otimes W$:

$$v = \sum_i \alpha_i v_i, \; w = \sum_j \beta_j w_j \implies v \otimes w = \sum_{i,j} \alpha_i \beta_j v_i \otimes w_j \,. \tag{1.33}$$

This gives us some rules for working with tensor products:

- Taking a vector multiplied by the scalar gives the same result as the scalar multiplied on the other vector:

$$(\alpha v) \otimes w = v \otimes (\alpha w) = \alpha (v \otimes w) \,. \tag{1.34}$$

- Similarly, we have linearity and distributivity:

$$(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w \,, \tag{1.35}$$

$$v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2 \,. \tag{1.36}$$

When we define the tensor product for vectors in the above way, we realize that our construction is basis-dependent. However, all such constructions are isomorphic.

There is even a way to define a tensor product space it without going through a basis, but we will not do this here.

We will use this construction not only for vectors but also for linear operators and channels. The construction will be exactly the same. However, we need to be a bit careful

Example. *If we take a tensor product of $\mathcal{L}(V)$ and $\mathcal{L}(W)$, we get tensors of linear maps, which are isomorphic to linear maps on the tensor product space $V \otimes W$, i.e.,*

$$\mathcal{L}(V) \otimes \mathcal{L}(W) \sim \mathcal{L}(V \otimes W) \tag{1.37}$$

When we construct $\mathcal{L}(V) \otimes \mathcal{L}(W)$, we require that $(L \otimes K)(v \otimes w) = Lv \otimes Kw$, which now tells us how elements in $\mathcal{L}(V) \otimes \mathcal{L}(W)$ act as linear operators on $V \otimes W$. This also implies that

It is sufficient to enforce that for a basis.

$$(L_1 \otimes K_1)(L_2 \otimes K_2) = (L_1 L_2) \otimes (K_1 K_2)\,. \tag{1.38}$$

Similarly we can also define the action of channels $\mathcal{L}(\mathcal{L}(V), \mathcal{L}(W)) \otimes \mathcal{L}(\mathcal{L}(V'), \mathcal{L}(W'))$ as an element of $\mathcal{L}(\mathcal{L}(V \otimes V'), \mathcal{L}(W \otimes W'))$ by looking at its action on tensors. We will do this implicitly often, e.g., when we consider a channel acting only on one subsystem of a composite system.

1.5.1 A useful isomorphism

Given an ONB $\{v_i\}_i$ of V, we introduce the isomorphism vec from $\mathcal{L}(V, W)$ to $W \otimes V$ as

$$\mathsf{vec} : X \mapsto \sum_i X v_i \otimes v_i = (X \otimes 1)\Omega\,, \tag{1.39}$$

where $\Omega = \sum_i v_i \otimes v_i$. Note that the inner product is preserved. Indeed,

$$\langle \mathsf{vec}(X), \mathsf{vec}(Y) \rangle = \sum_{i,j} \langle X v_i, Y v_j \rangle \langle v_i, v_j \rangle = \sum_i \langle X v_i, Y v_j \rangle = \langle X, Y \rangle\,. \tag{1.40}$$

This isomorphism vec is called the Choi-Jamiołkowski isomorphism when V and W are themselves spaces of linear operators and is useful when describing quantum channels.

1.5.2 Transpose

We also define the transpose X^T of X via the relation

$$\sum_i X v_i \otimes v_i = \sum_i v_i \otimes X^T v_i, \tag{1.41}$$

which evidently depends on the ONB chosen. This reduces to the usual matrix transpose when X and X^T are expressed in this basis (see below). Using the vec isomorphism, this allows us to write

$$\mathrm{Tr}(AB) = \langle 1, AB \rangle = \langle \Omega, (A \otimes B^T)\Omega \rangle\,. \tag{1.42}$$

The relation in (1.41) is also called the *transpose trick*, and we want to verify that if $X \in \mathcal{L}(A, A)$ then X^T above is indeed the usual matrix transpose of X.

Example. *Let us explore this in the example where $A \sim B$ are two isomorphic Hilbert spaces and $|\Omega\rangle_{AB} = \sum_x |x\rangle_A \otimes |x\rangle_B$ with two ONBs $\{|x\rangle_A\}_x$ of A and $\{|x\rangle_B\}_x$ of B. We then find*

$$
\begin{aligned}
(M_A \otimes I_B)\,|\Omega\rangle_{AB} &= \sum_j M_A\,|j\rangle_A \otimes |j\rangle_B && (1.43)\\
&= \sum_j \left(\sum_{i,k} M_{i,k}\,|i\rangle\langle k|_A \right) |j\rangle_A \otimes |j\rangle_B && (1.44)\\
&= \sum_{i,j} M_{i,j}\,|i\rangle_A \otimes |j\rangle_B && (1.45)\\
&= \sum_i |i\rangle_A \otimes \sum_j M_{i,j}\,|j\rangle_B && (1.46)\\
&= \sum_i |i\rangle_A \otimes \left(\sum_{j,k} M_{k,j}\,|j\rangle\langle k|_B \right) |i\rangle_B && (1.47)\\
&= \sum_i |i\rangle_A \otimes M_B^T\,|i\rangle_B && (1.48)\\
&= (I_A \otimes M_B^T)\,|\Omega\rangle_{AB}\ . && (1.49)
\end{aligned}
$$

2
Quantum formalism

2.1 States

Definition 2.1. *Quantum states (also called density operators) are psd operators with unit trace.*

In spectral decomposition, we can write $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$.

In other words, $\{\lambda_i\}_i$ is a probability mass function

- States are usually denoted with Greek letters such as $\rho, \sigma, \omega, \pi$.
- Pure sates are states with rank 1. We can write $\psi = |\psi\rangle\langle\psi|$, where $|\psi\rangle\langle\psi|$ is the outer product. There is no need to write the zero part of the spectral decomposition. Pure states are projectors $P^2 = P$.

Example. *The most common quantum system is the qubit. The set of* $\{|0\rangle, |1\rangle\}$ *is the computational basis,* $\{|+\rangle, |-\rangle\}$ *is the diagonal basis, where*

$$|0\rangle \sim \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle \sim \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{2.1}$$

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle) \sim \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} \tag{2.2}$$

The Hadamard operator maps between computational and diagonal basis. $H|0\rangle = |+\rangle$ *and* $H|1\rangle = |-\rangle$*. The backward conversion* $H|+\rangle = |0\rangle$ *and* $H|+\rangle = |1\rangle$ *follows from linearity and the decomposition in* (2.2)*. In matrix form:*

$$H \sim \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{2.3}$$

Let us consider the 4 Pauli matrices:

$$P_0 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.4}$$

$$P_1 = X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2.5}$$

$$P_2 = Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{2.6}$$

$$P_3 = Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.7}$$

To form a basis of Hermitian operators on qubits those operators require normalization: $\{\frac{1}{\sqrt{2}}P_i\}_i$ is an ONB. For any Hermitian operator H on a qubit we can write

$$H = \frac{1}{2}(a+b)1 + \frac{1}{2}(a-b)Z + \frac{1}{2}cX + \frac{1}{2}dY \sim \begin{pmatrix} a & c-id \\ c+id & b \end{pmatrix}, \tag{2.8}$$

where $a, b, c, d \in \mathbb{R}$ are the 4 real degrees of freedom.

2.1.1 Bloch representation

If ρ is a state we must have 1. $\mathrm{Tr}(\rho) = 1$ and 2. $\rho \geq 0$. From 1, we have $a + b = 1$ and we can rewrite $\rho = \frac{1}{2}(I + \sum_{x=1}^{3} r_x P_x)$, where $r_1, r_2, r_3 \in \mathbb{R}$. This gives

$$\rho = \frac{1}{2}(I + \sum_{x=1}^{3} r_x P_x) \sim \frac{1}{2}\begin{pmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{pmatrix}. \tag{2.9}$$

To satisfy 2, we notice that $\rho \geq 0 \iff \mathrm{Tr}(\rho) \geq 0$ and $\det(\rho) \geq 0$ for 2x2 matrices. The trace is already positive so we just need to compute the determinant as

To see this, note that from the determinant, we know that the product of eigenvalues is positive so they have the same sign and from the trace we know that they are positive.

$$\det(\rho) = \frac{1}{4}(1 - r_x^2 - r_y^2 - r_z^2) \tag{2.10}$$

Hence, $\det(\rho) \geq 0 \iff r_x^2 + r_y^2 + r_z^2 \leq 1 \iff \|r\| \leq 1$.

This gives the Bloch representation of qubit states, where pure states are located on the surface of the sphere $\|r\| = 1$ due to $\lambda_0 = 1, \lambda_1 = 0$ and consequently $\det \rho = 0$.

2.2 Measurements

Definition 2.2. *A positive operator valued measure (POVM) is a set of $\{M_x\}_{x\in\mathcal{X}}$ such that $M_x \geq 0$ and $\sum_x M_x = I$.*

- *A POVM is projective if M_x are projector: $M_x^2 = M_x$.*

- *The operators M_x are called the effects.*
- *The set $\mathcal{X}$ is called the alphabet.*

Born's rule associates probabilities to states and effects. The probability of measuring $x \in \mathcal{X}$ on a state ρ is given by

$$p_x = \Pr[[] X = x] = \langle M_x, \rho \rangle . \tag{2.11}$$

- We have $p_x \geq 0$ because of lemma 1.16.
- We have $\sum_x p_x = 1$ because of linearity $\sum_x p_x = \langle \sum_x M_x, \rho \rangle = \langle I, \rho \rangle = \mathrm{Tr}(\rho) = 1$.

Example. *For a pure state $|\psi\rangle\langle\psi|$, we have*

$$p_x = \langle M_x, |\psi\rangle\langle\psi| \rangle = \mathrm{Tr}(M_x |\psi\rangle\langle\psi|) = \mathrm{Tr}(\langle\psi| M_x |\psi\rangle) \tag{2.12}$$

$$= \langle \psi | M_x | \psi \rangle = \langle |\psi\rangle, M_x |\psi\rangle \rangle . \tag{2.13}$$

Naimark's dilation tells us that we can see every POVMas a projective measurement on a larger space.

Theorem 2.3 (Naimark dilation)**.** *Given a POVM $\{M_x\}_x$ on A, there exists a projective POVM $\{P_x\}_x$ and an isometry $U : A \to AX$ such that $\langle M_x, \rho \rangle = \langle P_x, U\rho U^\dagger \rangle$.*

Proof. Let $P_x = I_A \otimes |x\rangle\langle x|$ and $U = \sum_x \sqrt{M_x} \otimes |x\rangle$ so that $U^\dagger = \sum_{x'} \sqrt{M_x} \otimes \langle x'|$. One can verify that $U^\dagger U = I$. Moreover, we have

$$\langle P_x, U\rho U^\dagger \rangle = \langle U^\dagger P_x U, \rho \rangle. \tag{2.14}$$

Hence, it is only left to verify $U^\dagger P_x U = M_x$. We find

$$U^\dagger P_x U = \sum_{x', x''} \sqrt{M_{x'}} I \sqrt{M_{x''}} \, \langle x' | \cdot |x\rangle\langle x| \cdot |x''\rangle = M_x. \tag{2.15}$$

Therefore, $\langle M_x, p \rangle = \langle P_x, U\rho U^\dagger \rangle$. □

2.3 Composite systems

Every physical system is associated with a Hilbert space. We usually use letters $A, B, C, \ldots$ to denote these Hilbert spaces. Our states and objects live in the tensor products of these spaces: $A \otimes B \otimes C$, which are shortened to ABC. We usually use subscripts to indicate where operators live, e.g., the notation ρ_{AB} indicates that the quantum state $\rho_{AB} \geq 0$, $\mathrm{Tr}(\rho_{AB}) = 1$ is an element of $\mathcal{H}(AB)$.

If we talk about cryptography, which we will, then A is usually Alice, B is Bob, C is Charlie. These letters make it easier to discuss them sometimes.

2.3.1 Marginal state

We will present here the mathematical definition of the partial trace and come back to a more physical motivation later.

Definition 2.4. *Given state $\rho_{AB} \in \mathcal{H}(AB)$ and an ONB $\{|i\rangle_B\}_i$ of B, we define the marginal state:*

$$\rho_A = \sum_i (I_A \otimes \langle i|_B)\rho_{AB}(I_A \otimes |i\rangle_B) =: \mathrm{Tr}_B(\rho_{AB})$$

where I_A is the identity operator in $\mathcal{L}(A)$.

The operation that gives us the marginal (a channel, we will get back to that later) is also known as the partial trace (on the space B):

Again, this construction is basis independent.

$$(\mathcal{I}_A \otimes \mathrm{Tr}_B)(\rho_{AB}) = \mathrm{Tr}_B(\rho_{AB}) \tag{2.16}$$

To compute the partial trace we will first decompose ρ_{AB} into tensors and then use linearity to compute the inner products.

2.3.2 Purification

Definition 2.5. *A* purification *of a state ρ_A is a pure state $|\rho\rangle\langle\rho|_{AB}$ such that $\mathrm{Tr}_B(|\rho\rangle\langle\rho|_{AB}) = \rho_A$.*

We present one way to construct a purification here, but this is not unique. The state on A written into eigenvalue decomposition:

$$\rho_A = \sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i|_A \,. \tag{2.17}$$

We now introduce a new Hilbert space A' which is isomorphic to A, which means they have the same dimension and we can map the basis from A to basis in A'. Then we define the vector

$$|\rho\rangle_{AA'} = \sum_i \sqrt{\lambda_i}|\varphi_i\rangle_A \otimes |i\rangle_{A'} \tag{2.18}$$

Clearly $\rho_{AA'}$ is a pure state and we will next verify that its partial trace over B gives back ρ_A. We find

$$\begin{aligned}
\mathrm{Tr}_{A'}(\rho_{AA'}) &= \mathrm{Tr}_{A'}\left(\sum_{i,j}\sqrt{\lambda_i\lambda_j}|\varphi_i\rangle\langle\varphi_j| \otimes |i\rangle\langle j|\right) && (2.19)\\
&= \sum_{i,j}\sqrt{\lambda_i\lambda_j}|\varphi_i\rangle\langle\varphi_j| \cdot \langle i|j\rangle && (2.20)\\
&= \sum_{i,j}\sqrt{\lambda_i\lambda_j}|\varphi_i\rangle\langle\varphi_j| \cdot \delta_{ij} && (2.21)\\
&= \sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i| && (2.22)\\
&= \rho_A && (2.23)
\end{aligned}$$

Hence, we have verified that $\rho_{AA'} = |\rho\rangle\langle\rho|_{AA'}$ is a purification of ρ_A.

2.3.3 Entanglement

The fundamental differences between quantum and classical information stem from the properties, implications, and uses of quantum entanglement.

We will start with a formal definition of a separable state and entangled state of a composite system.

Definition 2.6. *A state is called* separable *if it can be written as*

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i$$

where $\{\rho_A^i\}$ and $\{\rho_B^i\}$ are states and $\{p_i\}$ are probabilities. Otherwise, it is called entangled.

A decomposition into tensors is always possible, but the point here is that it has to be psd tensors! In other words, linear operators on a joint system can always be written as a linear combination of tensors, but the coefficients have to be positive for it to be a separable state!

As a consequence, pure entangled states are states of a composite system that cannot be expressed as a tensor of pure states of individual subsystems.

Let us go through some examples.

Example. *For two qubits, the maximally entangled states are known as the Bell states. They can be expressed as*

$$|\psi_i\rangle = \big(P_i \otimes I\big)\frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big) \tag{2.24}$$

where P_i are the Pauli matrices. Here, $|00\rangle$ is the standard notation for $|0\rangle \otimes |0\rangle$, and likewise for $|11\rangle$.

Simplifying from this notation, the four Bell states can be written as

$$|\psi_0\rangle = \big(I \otimes I\big)\frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big) = \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big) \tag{2.25}$$

The X Pauli matrix flips the 0's and 1's.

$$|\psi_1\rangle = \big(X \otimes I\big)|\psi_0\rangle = \frac{1}{\sqrt{2}}\big(|10\rangle + |01\rangle\big) \tag{2.26}$$

The Y Pauli matrix enacts a bit and a phase flip (and adds a global phase i).

$$|\psi_2\rangle = \big(Y \otimes I\big)|\psi_0\rangle = \frac{i}{\sqrt{2}}\big(|10\rangle - |01\rangle\big) \tag{2.27}$$

The Z Pauli matrix enacts a phase flip operation.

$$|\psi_3\rangle = \big(Z \otimes I\big)|\psi_0\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle - |11\rangle\big) \tag{2.28}$$

We will encounter these states often.

Example. *Classical-quantum (CQ) states are a special class of bipartite quantum states where one subsystem remains classical while the other retains*

quantum properties. Mathematically, a CQ state for a bipartite system takes the form

$$\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \rho_B^x \tag{2.29}$$

To indicate classicality, we use the labels X, Y, Z and for quantum, we use A, B, C.

where p_x are probabilities from a probability mass function. The state $\rho_x = \sum_x p_x |x\rangle\langle x|$ is classical as it is diagonal in the computational basis.

These CQ states describe classical information correlated to a quantum system. They arise, for instance, if we flipped a classical coin and prepared a quantum state depending on the outcome of that coin. Another example would be having a quantum composite state, measuring one subsystem, recording the outcome of the measurement in a classical register, and then throwing away the state.

2.4 Channels

Now we have covered states and measurements. The last thing we need is evolution. The time evolution of quantum systems is described by quantum channels. Quantum channels are linear maps that take states to states.

Definition 2.7. *A linear map $\mathcal{E}_{A\to B} \in \mathcal{L}(\mathcal{H}(A), \mathcal{H}(B))$ is a* quantum channel *if it satisfies:*

- *$\mathcal{E}_{A\to B}$ is trace-preserving:*

$$\mathrm{Tr}(\mathcal{E}_{A\to B}(H_A)) = \mathrm{Tr}(H_A), \quad \forall H_A \in \mathcal{H}(A).$$

- *$\mathcal{E}_{A\to B}$ is positive:*

$$H_A \geq 0 \implies \mathcal{E}_{A\to B}(H_A) \geq 0.$$

- *$\mathcal{E}_{A\to B}$ is completely positive: $\mathcal{E}_{A\to B} \otimes \mathcal{I}_C$ is positive for all C, or*

$$\rho_{AC} \geq 0 \Rightarrow (\mathcal{E}_{A\to B} \otimes \mathcal{I}_C)(\rho_{AC}) \geq 0.$$

Channels are also called completely positive trace-preserving (CPTP) maps.

2.4.1 Example channels

Let us look at some examples.

- Let U be a unitary operator. A unitary channel is the map

$$\mathcal{U}_{A\to A}(\cdot) = U_A(\cdot)U_A^\dagger \tag{2.30}$$

Unitary channels $\mathcal{U}$ are indeed CPTP maps. To see this, note that

We used that $X\rho X^\dagger \geq 0$ if $\rho \geq 0$.

$$\mathrm{Tr}(\mathcal{U}(\rho)) = \mathrm{Tr}(U\rho U^\dagger) = \mathrm{Tr}(\rho) \qquad \text{and} \tag{2.31}$$

$$(\mathcal{U}_{A\to A} \otimes I_B)(\rho_{AB}) = (U_A \otimes I_B)(\rho_{AB})(U_A^\dagger \otimes I_B) \geq 0 \tag{2.32}$$

For a pure state $\psi_A = |\psi\rangle\langle\psi|_A$, we have $\mathcal{U}_{A\to A}(\psi_A) = U|\psi\rangle\langle\psi|_A U^\dagger = |\psi'\rangle\langle\psi'|_A$, where $|\psi'\rangle = U|\psi\rangle$ is the evolved state.

- Pauli channels are defined as

$$\mathcal{P}(H) = \sum_{i=0}^{3} p_i P_i H P_i^\dagger \tag{2.33}$$

with Pauli operators P_i and a probability distribution $\{p_i\}_i$.

- The (partial) trace, as a channel, maps a Hermitian H in some space B to a real number in $\mathbb{R}$:

$$\mathrm{Tr}_B : H \in \mathcal{H}(B) \to \mathrm{Tr}_B(H) \in \mathbb{R}$$

The (partial) trace is linear and is trace-preserving by definition. It is positive and completely positive:

$$I_A \otimes \mathrm{Tr}_B \in \mathcal{H}(AB) \to \mathcal{H}(A) \tag{2.34}$$

We can also write down its adjoint channel as

$$(\mathrm{Tr}_B)^\dagger : H_A \in \mathcal{H}(A) \to H_A \otimes I_B \in \mathcal{H}(AB) \tag{2.35}$$

Indeed, we can verify that

$$\begin{aligned} \langle H_A, I_A \otimes \mathrm{Tr}_B(M_{AB})\rangle &= \mathrm{Tr}(H_A \,\mathrm{Tr}_B(M_{AB})) && (2.36)\\ &= \mathrm{Tr}(H_A \otimes I_B \cdot M_{AB}) && (2.37)\\ &= \langle H_A \otimes I_B, M_{AB}\rangle_{AB}\,. && (2.38) \end{aligned}$$

- The transpose map is a linear map (not a channel) that is useful in studying entanglement properties. It is defined by transposing the matrix representation of a quantum channel in a given basis:

$$\tau(|i\rangle\langle j|) = |j\rangle\langle i| \tag{2.39}$$

It is trace-preserving, positive, but not completely positive. To verify the last point we see what it does on a Bell state of a composite system. We find

$$X_{AA'} = (\tau_A \otimes \mathcal{I}_{A'})(|\psi_0\rangle\langle\psi_0|) = \frac{1}{d}\sum_{i,j} |ji\rangle\langle ij| \tag{2.40}$$

where d is the dimension of the Hilbert space. To check that $X_{AA'}$ has a negative eigenvalue, we compute the overlap with $|\psi_2\rangle = \frac{i}{\sqrt{2}}(|10\rangle -$

$|01\rangle$) by computing $\langle\psi_2|X_{AA'}|\psi_2\rangle$. This yields

$$\langle\psi_2|X_{AA'}|\psi_2\rangle \tag{2.41}$$
$$= \frac{1}{2}\langle\psi_2|(|00\rangle\langle 00| + |11\rangle\langle 11| + |01\rangle\langle 10| + |10\rangle\langle 01|)|\psi_2\rangle \tag{2.42}$$
$$= \frac{1}{2}\langle\psi_2|(|01\rangle\langle 10| + |10\rangle\langle 01|)|\psi_2\rangle \tag{2.43}$$
$$= \frac{1}{4}(-1-1) = -\frac{1}{2}. \tag{2.44}$$

Note that $(i)\cdot(-i) = 1$, so the phase cancels out.

The transpose map is thus not a completely positive map.

2.4.2 *Partial trace via Born's rule*

We have some joint state between Alice and Bob. A POVM $\{M'_A\}_i$ is a measurement in lab A. We can construct another POVM $\{M'_A \otimes I_B\}_i$ on AB. Born's rule says that the probability of output $x = i$

$$Pr[x = i] = \langle M'_A \otimes I_B, \rho_{AB}\rangle \tag{2.45}$$
$$= \langle M'_A, \mathrm{Tr}_B(\rho_{AB})\rangle \tag{2.46}$$
$$= \langle M'_A, \rho_A\rangle \tag{2.47}$$

This is the same expression as in the definition of the marginal in the previous section.

If we only care about the probability of Alice's outcomes, we do not need to work with joint states. We can get it from the measurements done only on Alice's side by using the marginal state.

This is important because this tells us that there is no spooky action at the distance. What happens or what we see on Alice's side is completely determined by the state ρ_A. The probability determined by ρ_A will not change even if Bob applies some channel on B.

However, when we concern ourselves with correlations between measurement outcomes in Alice's lab and Bob's lab, then it is not sufficient to just look at the marginal states.

2.4.3 *Choi-Jamiołkowski isomorphism*

The Choi-Jamiołkowski isomorphism represents a quantum channel as a matrix by means of the vec isomorphism we introduced previously, effectively encoding its action in a larger Hilbert space of linear operators. This works because the channel is a linear map on basis elements, meaning its effect on any operator can be understood through its effect on a chosen basis. By vectorizing operators and recording them in an orthogonal basis of the other system, we obtain a structured representation that can help in analysing and computing properties of quantum channels.

Definition 2.8. *Let $\mathcal{E}_{A\to B} \in \mathcal{L}(\mathcal{H}(A), \mathcal{H}(B))$ be a linear map. Define*

$$X_{A'B} = \mathsf{vec}(\mathcal{E}_{A\to B}) = (\mathcal{E}_{A\to B} \otimes \mathcal{I}_{A'})(|\Omega\rangle\langle\Omega|), \tag{2.48}$$

where $|\Omega\rangle = \sum_{i=0}^{d-1} |i\rangle \otimes |i\rangle$. The state $X_{A'B}$ is called the Choi state. *The following holds:*

- *$X_{A'B}$ is psd iff $\mathcal{E}_{A\to B}$ is completely positive.*
- *$\mathrm{Tr}_B(X_{A'B}) = I_{A'}$ iff $\mathcal{E}_{A\to B}$ is trace-preserving.*

With qubits, for instance, the state $|\Omega\rangle$ is proportional to the Bell state $|\psi_0\rangle$, so the Choi state is just the result of applying the channel to one part of a Bell state. Operationally, the isomorphism tells us that if we put one subsystem of the Bell state through a noisy channel, perform full tomography of the quantum state, then we will be able to fully reconstruct the properties of the noisy channel.

2.4.4 Measurement channels

We can see measurements as channels.

Definition 2.9 (Measurement channel)**.** *If we have a POVM $\{M_x\}_{x\in X}$ on A , we can define a map between A and AX,*

$$\mathcal{M}_{A\to AX}(\cdot) = \sum_x |x\rangle\langle x| \otimes \sqrt{M_x}(\cdot)\sqrt{M_x}\,, \tag{2.49}$$

which is called instrument. By tracing out system A, we get

$$\mathcal{M}_{A\to X} := \mathrm{Tr}_A(\mathcal{M}_{A\to AX}), \tag{2.50}$$

which is called measurement.

- This is a trace-preserving map.
- When applying the measurement map to a state ρ_A,

$$\mathcal{M}_{A\to X}(\rho_A) = \sum_x |x\rangle\langle x| \cdot \underbrace{\mathrm{Tr}(M_x\rho_A)}_{p_x} \tag{2.51}$$

  Here, p_x is the probability of getting outcome x.

- When applying the measurement map to system A of a state ρ_{AB},

$$\begin{aligned}\rho_{XB} = (\mathcal{M}_{A\to X}\cdot\mathcal{I}_B)(\rho_{AB}) &= \sum_x |x\rangle\langle x| \otimes \mathrm{Tr}_A(\sqrt{M_x}\rho_{AB}\sqrt{M_x}) \\ &= \sum_x p_x |x\rangle\langle x| \otimes \rho_B^x \end{aligned} \tag{2.52}$$

  where $\rho_B^x = \frac{\mathrm{Tr}_A(\sqrt{M_x}\rho_{AB}\sqrt{M_x})}{\mathrm{Tr}(M_x\rho_{AB})}$ is the post-measurement state.

It is also called the collapsed state, but we will not use this expression.

This viewpoint on measurements is very useful because sometimes we have multiple systems but only want to measure some of them, and we can express this process easily using measurement channels.

2.5 Norms and metrics on states

Norms and metrics for quantum states are essential to quantify the similarity or distinguishability between two quantum states. This is useful, for example, to determine how well we are able to implement some quantum process in experiment.

We start with some norms. Given $X = \sum_i s_i |w_i\rangle\langle v_i|$ from the SVD, the canonical norm is defined as

$$\begin{aligned} ||X|| &= \sqrt{\langle X, X\rangle} = \sqrt{\sum_j \langle Xv_j, Xv_j\rangle} \\ &= \sqrt{\sum_j \langle s_j w_j, s_j w_j\rangle} = \sqrt{\sum_j s_j^2} = ||X||_2 \end{aligned}$$

This is also called the Frobenius or Schatten-2 norm.

Definition 2.10. *The* Schatten p-norm *is defined as*

$$||X||_p = \left(\sum_i (s_i)^p\right)^{\frac{1}{p}}$$

If the norm of X is 0, then $X = 0$.

We mostly use the trace norm $||X||_1$ and the operator norm $||X||_\infty$.

$$||X||_1 = \sum_i s_i = \operatorname{Tr}|X| = \operatorname{Tr}\left(\sqrt{X^\dagger X}\right), \tag{2.53}$$

$$||X||_\infty = s_0\,. \tag{2.54}$$

Here, s_0 is the maximum singular value of X. The Schatten p-norm are unitarily invariant in that $||UXV^\dagger||_p = ||X||_p$ for all isometries U, V. Hölder's inequality can be stated as

$$|\langle X, Y\rangle| \leq ||X||_p ||Y||_q \quad \text{for} \quad \frac{1}{p} + \frac{1}{q} = 1. \tag{2.55}$$

The Cauchy-Schwartz inequality is a special case of the Hölder's inequality where $p = q = 2$. Another commonly used case is $p = 1, q = \infty$, which, applied to the probability in Born's rule, for example, tells us that

$$\Pr[[]\, X = x] = \langle M_x, \rho\rangle \leq \|M_x\|_\infty \|\rho\|_1 = 1\,.$$

Before we define the common metrics for quantum states, let's review the necessary properties of a metric.

Definition 2.11. *A* metric d *statisifes the following:*

1. $d(\rho,\sigma) \geq 0$ *with equality iff* $\rho = \sigma$.
2. $d(\rho,\sigma) \leq d(\rho,\tau) + d(\tau,\sigma)$ *(triangle inequality)*
3. $d(\rho,\sigma) = d(\sigma,\rho)$

The trace distance between two quantum states has an operational interpretation as the maximum probability of successfully distinguishing the two states in a single-shot measurement scenario (as discussed in the homework).

Definition 2.12. *The trace distance is defined as*

$$\Delta(\rho,\sigma) = \frac{1}{2}||\rho - \sigma||_1 \,. \tag{2.56}$$

We have $\Delta(\rho,\sigma) \in [0,1]$. In the trivial case of distinguishing two copies of the same state, $\Delta(\rho,\rho) = 0$ where there is zero probability of distinguishing between the two states. The maximum value 1 is achieved if the two states are orthogonal.

Fidelity is another metric to compare two quantum states.

Definition 2.13. *The fidelity between two quantum states is defined as*

$$F(\rho,\sigma) = ||\sqrt{\rho}\sqrt{\sigma}||_1^2 = (\mathrm{Tr}\,|\sqrt{\rho}\sqrt{\sigma}|)^2 = \left(\mathrm{Tr}\,\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2 . \tag{2.57}$$

We have $F(\rho,\sigma) \in [0,1]$. Here the highest value is achieved with two copies of the same state $F(\rho,\rho) = 1$, and the lowest value is achieved when the two states are orthogonal.

If we compare a state ρ with a pure state $|\psi\rangle\langle\psi|$, we obtain $F(\rho,|\psi\rangle\langle\psi|) = \langle\psi|\rho|\psi\rangle$. In the case of comparing two pure states, $F(|\theta\rangle\langle\theta|,|\psi\rangle\langle\psi|) = |\langle\psi|\theta\rangle|^2$.

The fidelity is not a metric. However, metrics can be defined using the fidelity. For example, the purified distance.

Definition 2.14. *The purified distance or infidelity between two quantum states is defined as*

$$P(\rho,\sigma) = \sqrt{1 - F(\rho,\sigma)} \,. \tag{2.58}$$

The trace distance provides upper and lower bounds on the fidelity as quantified by the Fuchs–van de Graaf inequality.

Lemma 2.15. *The Fuchs-van de Graaf inequality states that*

$$1 - \sqrt{F(\rho,\sigma)} \leq \Delta(\rho,\sigma) \leq \sqrt{1 - F(\rho,\sigma)} = P(\rho,\sigma)\,. \tag{2.59}$$

Moreover, the second inequality is an equality if ρ, σ are both pure states.

Both Δ and F, P are monotonic under quantum channels. This is described by the data-processing inequality (DPI). Intuitively, this tells us that applying noise via a quantum channel only makes states closer and harder to distinguish.

Lemma 2.16. *The data-processing inequality (DPI) states that*

- $\Delta(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq \Delta(\rho,\sigma)$
- $F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq F(\rho,\sigma)$
- $P(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq P(\rho,\sigma)$

for any pair of states ρ and σ and any channel $\mathcal{E}$.

We saw that for two pure states, their fidelity coincides with the overlap. Uhlmann's theorem generalizes this statement to mixed states, in terms of their purifications.

Lemma 2.17. *Uhlmann's theorem states that*

$$F(\rho,\sigma) = \max_{|\rho\rangle, |\sigma\rangle} F(|\rho\rangle\langle\rho|, |\sigma\rangle\langle\sigma|) = |\langle\rho|\sigma\rangle|^2$$

where $|\rho\rangle, |\sigma\rangle$ are purifications of ρ, σ, respectively.

Then we can proceed with a sketch of the proof of Uhlmann's theorem.

Proof. Now, for one direction, using the fact that the partial trace is a quantum channel and Lemma 2.16, we find

$$F(\rho,\sigma) \geq F(|\rho\rangle\langle\rho|, |\sigma\rangle\langle\sigma|) \tag{2.60}$$

for any purifications of ρ and σ. Thus, the inequality also holds when we maximise over purifications on the right-hand side.

For the other direction, let us first construct specific purifications of the quantum states ρ, σ. A useful property we will take advantage of is $X \otimes \mathbb{1}|\Omega\rangle = \mathbb{1} \otimes X^T|\Omega\rangle$ where $|\Omega\rangle = \sum_i |i\rangle \otimes |i\rangle$. We find

$$|\rho\rangle = \mathrm{vec}(\sqrt{\rho}) = \sum_i \sqrt{\rho}|i\rangle \otimes |i\rangle \tag{2.61}$$

$$|\sigma\rangle = (\mathbb{1} \otimes U)\mathrm{vec}(\sqrt{\sigma}) = \mathrm{vec}(\sqrt{\sigma}U^T) \tag{2.62}$$

Using this, we have

$$\max_U |\langle\rho|\sigma\rangle| = \max_U \langle\sqrt{\rho}, \sqrt{\sigma}U^T\rangle \tag{2.63}$$

$$= \max_U \mathrm{Tr}(\sqrt{\rho}\sqrt{\sigma}U^T) \tag{2.64}$$

$$= \max_U \mathrm{Tr}(V|\sqrt{\rho}\sqrt{\sigma}|U^T) \tag{2.65}$$

$$\geq \mathrm{Tr}\,|\sqrt{\rho}\sqrt{\sigma}| = \sqrt{F(\rho,\sigma)}\,. \tag{2.66}$$

Here, we used the polar decomposition and the inequality on the last line follows from the choice of U that makes $U^T = V^\dagger$. □

3
Quantum correlation and games

A natural way to explore the power of quantum correlations (i.e., entanglement) is via two-party games where the players are not allowed to communicate (after receiving their respective inputs) and may or may not employ quantum entanglement to increase their winning probability.

3.1 Two-party games

Let us first introduce such games very generically.

Definition 3.1. *A two-party game of Alice and Bob is determined by parameters $G = (X, Y, A, B, P_{XY}, V)$, where*

- *X, Y are discrete and finite input alphabets.*
- *A, B are discrete and finite output alphabets.*
- *P_{XY} is the probability mass function on inputs.*
- *$V : A \times B \times X \times Y \to \{0, 1\}$ is the verification function.*

The game is played as follows: The inputs x, y are sampled from P_{XY}. Alice gets $x \in X$ and outputs $a \in A$, Bob gets $y \in Y$ and outputs $b \in B$. They win if their inputs and outputs satisfy $V(a, b, x, y) = 1$.

It is important that Alice and Bob cannot communicate, but they can prepare a strategy in advance.

Example. *The Clauser-Horne-Shimony-Holt (CHSH) game is the best known two-party game. In this case, inputs and outputs are bits, which we can see as elements of the field $\mathbb{F}_2$, where addition corresponds to the exclusive or of bits and multiplication corresponds to the logical and of bits.*

Definition 3.2. *CHSH game is a two-party game with $X = Y = A = B =$*

$\mathbb{F}_2$, *uniform input distribution* $P_{XY}(x,y) = \frac{1}{4}, \forall x, y$ *and*

$$V(a,b,x,y) = \begin{cases} 1, & \text{if } a+b = xy \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

3.2 Classical strategies

A general classical strategy S for Alice and Bob is given by a joint distribution of a random variable $Q_{X'Y'}$ and two functions $f_A : X \times X' \to A$ and $f_B : Y \times Y' \to B$, determining Alice's and Bob's strategy, respectively.

The winning probability is then

Here, the superscript cl stands for classical.

$$\begin{aligned} p^{\text{cl}}_{\text{win}}(G,S) &= \Pr[[]\, V(f_A(X,X'), f_B(Y,Y'), X, Y) = 1] && (3.2)\\ &= \sum_{x,x',y,y'} P_{XY}(x,y) Q_{X'Y'}(x',y') \\ &\qquad \cdot V(f_A(x,x'), f_B(y,y'), x, y) && (3.3)\\ &= \sum_{x,y,a,b} P_{XY}(x,y) S(a,b|x,y) V(a,b,x,y)\,, && (3.4) \end{aligned}$$

where

$$S(a,b|x,y) = \sum_{x',y'} Q_{X'Y'}(x',y') \mathbf{1}\{f_A(x,x') = a\} \cdot \mathbf{1}\{f_B(y,y') = b\} \tag{3.5}$$

encapsulates the strategy.

We next show that there is always a deterministic optimal strategy.

Definition 3.3. *A strategy is deterministic if, for all* $x, x' y, y'$,

$$f_A(x,x') = f_A(x), \quad f_B(y,y') = f_B(y). \tag{3.6}$$

Lemma 3.4. *The maximum winning probability,* $p^{\text{cl}}_{\text{win}}(G) = \sup_S p^{\text{cl}}_{\text{win}}(G,S)$, *is achieved by a deterministic strategy.*

Proof. The idea here is that the winning probability is averaged over the distribution $Q_{X'Y'}$, and thus we can always just bound this average by its maximum, which is achieved for a particular pair (x^*, y^*). Formally,

$$\begin{aligned} P^{cl}_{win}(G,S) &\leq \max_{x',y'} \sum_{x,y} P_{XY}(x,y) V(f_A(x,x'), f_B(y,y'), x, y) && (3.7)\\ &= \sum_{x,y} P_{XY}(x,y) V(f_A(x,x^*), f_B(y,y^*), x, y), && (3.8) \end{aligned}$$

where (x^*, y^*) is any pair that achieves the maximum in (x', y'). □

Example. *Revisiting the CHSH game, we can try all deterministic strategies to find the best. Considering all deterministic strategies, we can get* $P^{cl}_{win}(\text{CHSH}) = \frac{3}{4}$.

3.3 *Quantum strategy*

A general quantum strategy S is given by a state $\rho_{A'B'}$ and a collection of POVMs $\{M_a^x\}_a$ on A' and $\{N_b^y\}_b$ on B' for all x, y, with $\sum_a M_a^x = I_A$ and $\sum_b N_b^y = I_B$. Alice and Bob will apply different POVMs depending on the input.

The winning probability is given by

$$p_{\text{win}}^{\text{q}}(G, S) = \sum_{x,y,a,b} P_{XY}(x,y) S(a,b|x,y) V(a,b,x,y) \tag{3.9}$$

with

$$S(a,b|x,y) = \text{Tr}((M_a^X \otimes N_b^Y)\rho_{A'B'})\,. \tag{3.10}$$

We need to argue why this is indeed the most general strategy for Alice and Bob. Indeed, Alice (or Bob) could also apply some channel $\mathcal{E}^x_{A'\to A'}$ (or $\mathcal{E}^y_{B'\to B'}$) locally, after receiving input x (or y). This would lead to outcome probabilities

$$\begin{aligned}&\text{Tr}((M_a^x \otimes N_b^y)(\mathcal{E}^x_{A'\to A'} \otimes I)(\rho_{A'B'}))\\ &\qquad = \text{Tr}(((\mathcal{E}^x_{A'\to A'})^\dagger M_a^x \otimes N_b^y)(\rho_{A'B'})).\end{aligned} \tag{3.11}$$

We thus want to show that $\{(\mathcal{E}^x_{A'\to A'})^\dagger(M_a^x)\}_a$ is also a POVM, so we could have just applied that directly. To do this, we introduct the notion of unital maps.

Definition 3.5 (Unital). *A linear map $\mathcal{F} \in \mathcal{L}(\mathcal{H}(A), \mathcal{H}(B))$ is called unital if $\mathcal{F}(I_A) = I_B$.*

Lemma 3.6. *The maps $(\mathcal{E}^x_{A'\to A'})^\dagger$ are unital.*

Proof. We have

$$\forall \rho:\ \langle \mathcal{E}^\dagger(I), \rho\rangle = \langle I, \mathcal{E}(\rho)\rangle = \text{Tr}(\mathcal{E}(\rho)) = \text{Tr}(\rho) = \langle I, \rho\rangle.$$

□

Lemma 3.7. *The PSD matrices $\{(\mathcal{E}^x_{A'\to A'})^\dagger(M_a^x)\}_a$ form a POVM.*

Proof. We have

$$\sum_a (\mathcal{E}^x_{A'\to A'})^\dagger(M_a^x) = (\mathcal{E}^x_{A'\to A'})^\dagger(\sum_a M_a^x) = (\mathcal{E}^x_{A'\to A'})^\dagger(I_{A'}) = I_{A'}. \tag{3.12}$$

□

Hence, we can assume without loss of generality that any quantum strategy is of the form given at the beginning of this section.

Moreover, Naimark's dilation theorem implies that P^q_{win} is achieved with a pure state and projective measurements since we can simply enlarge the space to accommodate for these measurements.

We start with a simple observation about quantum strategies.

Lemma 3.8. *The maximum of probability to win, $p^{\text{q}}_{\text{win}}(G) := \sup_S p^{\text{q}}_{\text{win}}(G,S)$, satisfies*

$$p^{\text{q}}_{\text{win}}(G) \geq p^{\text{cl}}_{\text{win}}(G)\,. \tag{3.13}$$

Proof. We can write every classical strategy as a quantum strategy. Concretely, choose S as

$$\rho_{A'B'} = 1 \in \mathcal{L}(\mathcal{C}\otimes\mathcal{C}) \tag{3.14}$$

$$M^x_a = \begin{cases} 1, & \text{if } f_A(x) = a \\ 0, & \text{otherwise} \end{cases} \tag{3.15}$$

$$N^y_b = \begin{cases} 1, & \text{if } f_B(y) = b \\ 0, & \text{otherwise} \end{cases} \tag{3.16}$$

then $p^{\text{q}}_{\text{win}}(G,S) = p^{\text{cl}}_{\text{win}}(G)$, and the inequality thus holds. □

3.4 CHSH Game

We have already introduced the CHSH game and explored classical strategies. Let us now expore a particular quantum strategy, that we will later prove to be optimal.

3.4.1 Optimal strategy

In the following, we use the notation Let $|\alpha\rangle := \cos\alpha\,|0\rangle + \sin\alpha\,|1\rangle$ for some angle $\alpha \in [-\pi,\pi]$. Our quantum strategy S is as follows. The parties share a maximally entangled state

$$\rho_{A'B'} = |\psi_0\rangle\langle\psi_0|, \tag{3.17}$$

Recall that $|\psi_0\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.

and peform measurements

$$M^0 = \{|0\rangle\langle 0|, |\frac{\pi}{2}\rangle\langle\frac{\pi}{2}|\}, \quad M^1 = \{|\frac{\pi}{4}\rangle\langle\frac{\pi}{4}|, |-\frac{\pi}{4}\rangle\langle-\frac{\pi}{4}|\}, \tag{3.18}$$

$$N^0 = \left\{|\frac{\pi}{8}\rangle\langle\frac{\pi}{8}|, |\frac{5\pi}{8}\rangle\langle\frac{5\pi}{8}|\right\}, \tag{3.19}$$

$$N^1 = \left\{|-\frac{\pi}{8}\rangle\langle-\frac{\pi}{8}|, |\frac{3\pi}{8}\rangle\langle\frac{3\pi}{8}|\right\}. \tag{3.20}$$

We now compute $P^q_{\text{win}}(G,S)$ for this strategy S. The probability that

Alice's output corresponds to $|\alpha\rangle$ and Bob's corresponds to $|\beta\rangle$ is

$$P_{\alpha,\beta} = \mathrm{Tr}\left(|\alpha\rangle\langle\alpha|_{A'} \otimes |\beta\rangle\langle\beta|_{B'} |\psi_0\rangle\langle\psi_0|_{A'B'}\right) \tag{3.21}$$
$$= \mathrm{Tr}\left(|\alpha\rangle\langle\alpha|_{A'} (|\beta\rangle\langle\beta|)^T_{A'} \otimes I_{B'} |\psi_0\rangle\langle\psi_0|_{A'B'}\right) \tag{3.22}$$
$$= \frac{1}{2}\mathrm{Tr}\left(|\alpha\rangle\langle\alpha|_{A'} |\beta\rangle\langle\beta|_{A'}\right) \tag{3.23}$$
$$= \frac{1}{2}|\langle\alpha,\beta\rangle|^2 \tag{3.24}$$
$$= \frac{1}{2}\cos^2(\alpha-\beta). \tag{3.25}$$

For Equation (3.22), note that $(I_{A'} \otimes M_{B'})|\psi_0\rangle = (M^T_{A'} \otimes I_{B'})|\psi_0\rangle$ by the transpose trick for maximally entangled states. For Equation (3.23), notice that $|\beta\rangle$ is a real vector so $(|\beta\rangle\langle\beta|)^T = |\beta\rangle\langle\beta|$, and

$$\mathrm{Tr}\big(M_{A'} \otimes I_{B'} |\psi_0\rangle\langle\psi_0|_{A'B'}\big) \tag{3.26}$$
$$= \langle\psi_0|_{A'B'} (M_{A'} \otimes I_{B'}) |\psi_0\rangle_{A'B'} \tag{3.27}$$
$$= \frac{1}{2}\sum_{i,j}(\langle i|_{A'} \otimes \langle i|_{B'})(M_{A'} \otimes I_{B'})(|j\rangle_{A'} \otimes |j\rangle_{B'}) \tag{3.28}$$
$$= \frac{1}{2}\sum_{i,j}\langle i|_{A'} M_{A'} |j\rangle_{A'} \langle i\rangle j_{B'} \tag{3.29}$$
$$= \frac{1}{2}\sum_{i}\langle i|_{A'} M_{A'} |i\rangle_{A'} \tag{3.30}$$
$$= \frac{1}{2}\mathrm{Tr}(M_{A'}). \tag{3.31}$$

Now we can get

$$p^{\mathrm{q}}_{\mathrm{win}}(\mathrm{CHSH}) \geq p^{\mathrm{q}}_{\mathrm{win}}(\mathrm{CHSH}, S) \tag{3.32}$$
$$= \frac{1}{4}\Bigg(\underbrace{P_{0,\frac{\pi}{8}} + P_{\frac{\pi}{2},\frac{5\pi}{8}}}_{\text{for input } x=y=0} + \underbrace{P_{0,-\frac{\pi}{8}} + P_{\frac{\pi}{2},\frac{3\pi}{8}}}_{x=0,y=1} + \underbrace{P_{\frac{\pi}{4},\frac{\pi}{8}} + P_{-\frac{\pi}{4},\frac{5\pi}{8}}}_{x=1,y=0} + \underbrace{P_{\frac{\pi}{4},\frac{3\pi}{8}} + P_{\frac{-\pi}{4},-\frac{\pi}{8}}}_{x=y=1}\Bigg) \tag{3.33}$$
$$= 2\cdot\frac{1}{2}\cos\left(\frac{\pi}{8}\right)^2 = \frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.85 > 0.75. \tag{3.34}$$

The next thing is to verify that this is the best strategy.

3.4.2 Tsirelson's bound

Assume that the optimal strategy is given by the state $|\varphi\rangle$, and projective measurements used by Alice and Bob (on input x, y) $\{P^x_a\}_a$ and $\{Q^y_b\}_b$. We have the following probability of obtaining outcomes a and b:

$$\Pr[(]\, a,b|x,y) = \langle\varphi|\, P_a^x \otimes Q_b^y \,|\varphi\rangle\,. \tag{3.35}$$

For example, for inputs $x = y = 0$, the probability of winning is

$$\sum_{a,b} \mathbf{1}\{a=b\}\Pr[(]\, a,b|0,0) = \frac{1}{2} + \frac{1}{2}\sum_{a,b}(-1)^{a+b}\,\langle\varphi|\, P_a^0 \otimes Q_b^0 \,|\varphi\rangle \tag{3.36}$$

$$= \frac{1}{2} + \frac{1}{2}\,\langle\varphi|\underbrace{\sum_{a,b}(-1)^{a+b} P_a^0 \otimes Q_b^0}_{C^0\otimes D^0}|\varphi\rangle\,. \tag{3.37}$$

Here we define

$$C^x := \sum_a (-1)^a P_a^x, \tag{3.38}$$

$$D^y := \sum_b (-1)^b Q_b^y, \tag{3.39}$$

for $x, y \in \{0,1\}$. It's straightforward to see that $(C^x)^2 = (D^y)^2 = I$. The winning probability is as follows:

$$p_{\text{win}}^{\text{q}}(\text{CHSH}) = \frac{1}{2} + \frac{1}{8}\,\langle\varphi|\underbrace{C^0\otimes D^0 + C^0\otimes D^1 + C^1\otimes D^0 - C^1\otimes D^1}_{=:B,\ \text{called Bell operator}}|\varphi\rangle \tag{3.40}$$

$$\leq \frac{1}{2} + \frac{1}{8}\,\|B\|_\infty \tag{3.41}$$

$$= \frac{1}{2} + \frac{1}{8}\sqrt{\|B^2\|_\infty}. \tag{3.42}$$

By direct computation, we have

$$B^2 = 4I + [C^0, C^1]\otimes[D^1, D^0]. \tag{3.43}$$

Using $\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty$ and $\|A\otimes B\|_\infty = \|A\|_\infty \|B\|_\infty$, we get

$$\left\|[C^0, C^1]\right\|_\infty \leq 2\left\|C^0\right\|_\infty \left\|C^1\right\|_\infty = 2, \tag{3.44}$$

$$\left\|B^2\right\|_\infty \leq 4 + 2\cdot 2 = 8. \tag{3.45}$$

It follows that

$$p_{\text{win}}^{\text{q}}(\text{CHSH}) \leq \frac{1}{2} + \frac{1}{8}\sqrt{\|B^2\|_\infty} \leq \frac{1}{2} + \frac{1}{2\sqrt{2}}. \tag{3.46}$$

This shows that the strategy we used earlier is already optimal, and so $p_{\text{win}}^{\text{q}}(\text{CHSH}) = \frac{1}{2} + \frac{1}{2\sqrt{2}}$.

3.5 Robustness

One property of the local measurements that achieve the maximum CHSH value is that they anti-commute as observables. In other words, the optimal measurements M_0 and M_1 in (3.18) correspond to the eigenvectors of $C^0 = Z$ and $C^1 = X$, respectively, and we have $\{X, Z\} = 0$, where $\{X, Z\} = XZ + ZX$. This is in fact a defining property of measurements that achieve maximal CHSH violation.

Lemma 3.9. *For any strategy that reaches the maximum quantum value for CHSH, we have* $(I \otimes \{D^0, D^1\}) |\psi\rangle = 0$.

Proof. We can expand the operator $2\sqrt{2}I - B$ with respect to the measurement projectors. This yields

$$2\sqrt{2}I - B = \frac{1}{\sqrt{2}} \left(\frac{C^0 + C^1}{\sqrt{2}} \otimes I - I \otimes D^0 \right)^2 + \frac{1}{\sqrt{2}} \left(\frac{C^0 - C^1}{\sqrt{2}} \otimes I - I \otimes D^1 \right)^2 . \tag{3.47}$$

To show this decomposition, we evaluate the right-hand side to find

$$\begin{aligned}
&\frac{1}{\sqrt{2}} \left[\left(\frac{C^0 + C^1}{\sqrt{2}} \otimes I - I \otimes D^0 \right)^2 + \left(\frac{C^0 - C^1}{\sqrt{2}} \otimes I - I \otimes D^1 \right)^2 \right] \\
&= \frac{1}{\sqrt{2}} \left[\frac{(C^0 + C^1)^2}{2} \otimes I - \sqrt{2}(C^0 + C^1) \otimes D^0 + I \otimes (D^0)^2 \right] \\
&\quad + \frac{1}{\sqrt{2}} \left[\frac{(C^0 - C^1)^2}{2} \otimes I - \sqrt{2}(C^0 - C^1) \otimes D^1 + I \otimes (D^1)^2 \right] && (3.48) \\
&= \frac{1}{\sqrt{2}} \left(4I - \sqrt{2}((C^0 + C^1) \otimes D^0 + (C^0 - C^1) \otimes D^1) \right) && (3.49) \\
&= \frac{1}{\sqrt{2}} \left(4I - \sqrt{2}B \right) && (3.50) \\
&= 2\sqrt{2}I - B. && (3.51)
\end{aligned}$$

Recall now that $B |\psi\rangle = 2\sqrt{2} |\psi\rangle$ for any state that achieves the maximal violation. Combining this with Eq. (3.47), we know that if the winning probability reaches the maximum, then we have

$$\left(\frac{C^0 + C^1}{\sqrt{2}} \otimes I - I \otimes D^0 \right) |\psi\rangle = 0; \tag{3.52}$$

$$\left(\frac{C^0 - C^1}{\sqrt{2}} \otimes I - I \otimes D^1 \right) |\psi\rangle = 0. \tag{3.53}$$

Hence we have

$$\left(\frac{C^0 + C^1}{\sqrt{2}} \otimes I \right) |\psi\rangle = \left(I \otimes D^0 \right) |\psi\rangle ; \tag{3.54}$$

$$\left(\frac{C^0 - C^1}{\sqrt{2}} \otimes I \right) |\psi\rangle = \left(I \otimes D^1 \right) |\psi\rangle . \tag{3.55}$$

Thus we have

$$
\begin{aligned}
&\left\{I \otimes D^0, I \otimes D^1\right\} |\psi\rangle \\
&= \left(I \otimes D^0 D^1 + I \otimes D^1 D^0\right) |\psi\rangle \\
&= \left[\left(\frac{C^0 + C^1}{\sqrt{2}} \otimes I\right)\left(\frac{C^0 - C^1}{\sqrt{2}} \otimes I\right) + \left(\frac{C^0 - C^1}{\sqrt{2}} \otimes I\right)\left(\frac{C^0 + C^1}{\sqrt{2}} \otimes I\right)\right] |\psi\rangle \\
&= \left[\left((C^0)^2 - (C^1)^2\right) \otimes I\right] |\psi\rangle \\
&= 0.
\end{aligned}
$$

The last equality above is since $(C^0)^2 = (C^1)^2 = I$. □

3.6 Mermin magic squares

Definition 3.10. *Merlin magic squres game is a two-party game with*

- $X = Y = \{1, 2, 3\}$
- $A = B = \{-1, 1\}^3$
- $P_{XY}(x, y) = \frac{1}{9}, \forall x, y$
- $V(a, b, x, y) = \begin{cases} 1, & \text{if } a_1 a_2 a_3 = 1 \text{ and } b_1 b_2 b_3 = -1 \text{ and } a_y = b_x \\ 0, & \text{otherwise} \end{cases}$

3.6.1 Classical strategy

The winning probability of the classical strategy is $p^{\text{cl}}_{\text{win}}(\text{MS}) = \frac{8}{9}$. One of the strategies to achieve it is as in Table 3.1.

x \ y	1	2	3
1	1	1	1
2	−1	1	−1
3	1	−1	Alice: −1 Bob: 1

Table 3.1: classical strategy

3.6.2 Quantum strategy

Lemma 3.11. *Let H, K be commuting observables, e.g. $[H, K] = 0$. There exists an orthonormal basis $\{|\varphi_i\rangle\}$ s.t.*

$$H |\varphi_i\rangle = \lambda_i |\varphi_i\rangle , \tag{3.56}$$

$$K |\varphi_i\rangle = \mu_i |\varphi_i\rangle . \tag{3.57}$$

The quantum strategy is as Table 3.2. Alice and Bob measure observables arranged in the square (sequentially since they commute) with the shared maximally entangled state $|\psi\rangle_{A_1B_1A_2B_2} = |\psi_0\rangle_{A_1B_1} \otimes |\psi_0\rangle_{A_2B_2}$.

x \ y	1	2	3
1	$I \otimes Z$	$Z \otimes I$	$Z \otimes Z$
2	$X \otimes I$	$I \otimes X$	$X \otimes X$
3	$-X \otimes Z$	$-Z \otimes X$	$Y \otimes Y$

Table 3.2: quantum strategy

Here $a_1a_2a_3 = 1$ and $b_1b_2b_3 = -1$ are already satisfied due to Lemma 3.11, as the product of each row is $I \otimes I$ and the product of each column is $-I \otimes I$.

Notice that $1 = \frac{1}{2} + \frac{1}{2} \times 1$, and $0 = \frac{1}{2} + \frac{1}{2} \times (-1)$.

Thus, we only have to verify the probability that $\forall(x, y) : \ a_y = b_x$ (below O_{xy} is the observable on the xth row and the yth column of the square):

$$
\begin{aligned}
P_{\text{win}|x,y} &= \Pr[[]\, a_y = b_x] && (3.58)\\
&= \frac{1}{2}\left(1 + \langle O_{xy}^{A_1A_2} \otimes O_{xy}^{B_1B_2}, |\psi\rangle\langle\psi|_{A_1B_1A_2B_2}\rangle\right) && (3.59)\\
&= \frac{1}{2}\left(1 + \frac{1}{4}\ \mathrm{Tr}(O_{xy}O_{xy}^T)\right) && (3.60)\\
&= 1. && (3.61)
\end{aligned}
$$

Equation (3.60) is due to the transpose trick. The last equality is since $\mathrm{Tr}(O_{xy}O_{xy}^T) = \mathrm{Tr}(I \otimes I) = 4$. Therefore, we have $p_{\text{win}}^{\text{q}}(\text{MS}) = 1$ and this is clearly already optimal.

4
Entropy, Uncertainty, and Randomness

In this chapter, we introduce basic measures of (quantum) information. We begin with the entropy and related measures, which we motivate as measures of uncertainty. We then discuss how these are all derived from the relative entropy. We next turn to the min-entropy and show that it characterizes the optimal ability for someone to guess the value x correctly when given access to quantum side-information according to classical-quantum state ρ_{XB}; a very concrete operational notion of uncertainty. Finally, we show that the amount of uniform, private randomness in the X register of ρ_{XB} is characterized by the *smooth* min-entropy. In other words, this chapter is about how the formalization of uncertainty via entropies allows us to characterize the (im)possibility of information processing tasks.

4.1 Surprisal and Shannon Entropy

4.1.1 Surprisal

We start by looking into how the uncertainty is captured in the classical scenario. Consider a random variable $X \in \mathcal{X}$ following the pmf $\Pr_X[x] = p_x$. We want to find a notion capturing how surprised we are to see a particular outcome $x \in \mathcal{X}$. It has to satisfy the following properties:

- It is a function of p_x. Let us call it $S(p_x)$.
- We have $S(1) = 0$, i.e., for a completely certain event we are not surprised at all.
- *Monotonicity:* $S(p_x)$ monotonically decreases as p_x increases.
- *Additivity:* $S(p_x \cdot p_y) = s(p_x) + s(p_y)$ for two independent events.

We notice that the only positive function that satisfies all the above properties is the logarithm function. Thus, we consider

$$S(p_x) = \log \frac{1}{p_x}, \tag{4.1}$$

where the logarithm is of the base 2 (so that we can count in bits). We call this notion *surprisal*. We can also see it as a random variable that takes the value $\log \frac{1}{p_x}$ with probability p_x. Since it is a function of the random variable X, we usually represent this new random variable as

$$S(X) = \log \frac{1}{\Pr_X[X]}. \tag{4.2}$$

4.1.2 Entropy

Now we can define Shannon *entropy*, which is the expected value of the surprisal of X.

Definition 4.1. *Given a random variable X, the entropy of X is defined as*

$$H(X) := \mathbb{E}[S(X)] = \mathbb{E}\left[\log \frac{1}{\Pr_X[X]}\right] = \sum_{x \in \mathcal{X}} \Pr_X[x] \log \frac{1}{\Pr_X[x]}. \tag{4.3}$$

Note that we use the convention that $0 \log 0 = 0$ throughout this chapter since $\lim_{\varepsilon \to 0} \varepsilon \log \varepsilon = 0$.

Consider the coin toss example. If the coin is fair (i.e., both heads and tails occur with probability $1/2$), then the entropy of the outcome of the coin toss is

$$H(X) = 2\left(\frac{1}{2} \log 2\right) = 1. \tag{4.4}$$

If the coin is biased such that the head occurs with the probability p, then the entropy is

$$H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} =: h(p). \tag{4.5}$$

We call $h(p)$ the *binary entropy*. It is easy to verify that $h(p)$ is symmetric around $1/2$, i.e., $h(p) = h(1-p)$. It also achieves the maximum value for $p = 1/2$. The binary entropy is plotted in Figure 4.1.

4.2 Von Neumann (Quantum) Entropy

Consider a quantum state ρ in the Hilbert space $\mathcal{H}$. We want to define the entropy of ρ in a way that generalizes the classical entropy. First, we want it to be unitarily invariant, that is,

$$H(U\rho U^\dagger) = H(\rho). \tag{4.6}$$

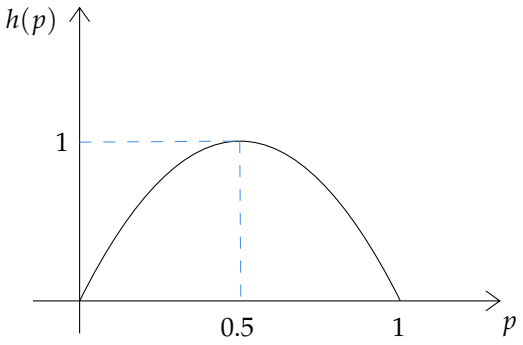

Figure 4.1: The binary entropy $h(p)$.

Intuitively, the unitary will not bring any noise to a quantum state, so our uncertainty about it and thus the entropy should not change. Furthermore, we also want that the entropy of a pure state is 0 since there is no uncertainty in a pure state. This leads to the following definition.

Definition 4.2. *Given a quantum state ρ, the von Neumann entropy of ρ is defined as*

$$H(\rho) := -\mathrm{Tr}(\rho \log \rho) = \sum_i \lambda_i \log \frac{1}{\lambda_i}. \tag{4.7}$$

Here, $\{\lambda_i\}_i$ are the eigenvalues of ρ, i.e., $\rho = \sum_i \lambda_i |\psi_i\rangle \langle\psi_i|$. Note that von Neumann entropy is simply the Shannon entropy of the eigevalues of ρ. It is thus non-negative. Moreover, if the dimension of the Hilbert Space is d it can be shown that

$$0 \leq H(\rho) \leq \log(d) \tag{4.8}$$

Given a state ρ_A in the Hilbert Space A, we usually denote $H(A)_\rho = H(\rho_A)$. If the context is clear, we may omit the subscript ρ, i.e., we simply write $H(A)$.

We next discuss several properties of von Neumann entropy.

1. *Monotone* under noise quantum channels (unital channels): For any mixing quantum channel $\mathcal{E}$ with $\mathcal{E}(I) = I$, we have

   $$H(\rho) \leq H(\mathcal{E}(\rho)). \tag{4.9}$$

   Intuitively, noise can increase the randomness in the state, so quantum entropy should increase.

2. Given a rank-1 projective measurement (von Neumann measurement) channel

   $$\mathcal{M}_{A\to X}(\cdot) = \sum_x |x\rangle \langle\psi_x| \cdot |\psi_x\rangle \langle x|, \tag{4.10}$$

where $\{|\psi_x\rangle\}_x$ is an ONB. We can easily verify that this is a unital channel.

$$\mathcal{M}_{A\to X}(I_A) = \sum_x |x\rangle\langle x| = I_X. \tag{4.11}$$

By the monotonicity of quantum entropy, we have

$$H(A) = \min_{\mathcal{M}_{A\to X}} H(X), \tag{4.12}$$

where the entropy $H(A)$ is achieved by the measuring in the eigenbasis of ρ_A. In other words, the entropy of a quantum state is the minimum Shannon entropy of the classical state obtained by measuring this quantum state.

3. *Duality:* Given a pure state $|\rho\rangle_{AB}$, and denote ρ_A and ρ_B as its marginals. Then, we have

$$H(A) = H(B). \tag{4.13}$$

We show this by applying Schmidt decomposition to $|\rho\rangle_{AB}$, then we have

$$|\rho\rangle_{AB} = \sum_i \sqrt{\lambda_i}\,|\phi_i\rangle_A\,|\psi_i\rangle_B\,, \tag{4.14}$$

where λ_is are eigenvalues of ρ_A and ρ_B, and $|\phi_i\rangle$s and $|\psi_i\rangle$s are the corresponding eigenvectors. It follows that $H(A) = H(B)$.

This property yields a different behaviour from Shannon entropy. In the classical case, the entropy of marginal distribution is always less than or equal to the entropy of joint distribution, i.e.,

$$H(XY) \geq H(X) \tag{4.15}$$
$$H(XY) \geq H(Y). \tag{4.16}$$

However, quantum entropy violates this property: When $|\rho\rangle_{AB}$ is pure, we have

$$H(AB) = 0 \leq H(A) = H(B). \tag{4.17}$$

4.3 Conditional Entropy and Mutual Information

Consider a bipartite state on $A \otimes B$. We want to quantify the relation between the system A and the system B.

4.3.1 Conditional Entropy

Definition 4.3. *Given a bipartite state ρ_{AB}, the conditional entropy of A given B is defined as*

$$H(A|B)_\rho := H(AB)_\rho - H(B)_\rho. \tag{4.18}$$

Note that $H(A|B)$ can be negative, but it's non-negative if ρ_{AB} is separable. Intuitively, conditional entropy is a measure of uncertainty of A from the perspective of an observer with access to B. We call B as the side-information. Here are several properties of conditional entropy.

1. *Data processing inequality (DPI):* Given a unital quantum channel $\mathcal{E}_{A\to A'}$ and a quantum channel $\mathcal{F}_{B\to B'}$, we have

$$H(A|B)_\rho \leq H(A'|B')_{\mathcal{E}\otimes\mathcal{F}(\rho)}. \tag{4.19}$$

   It follows that the randomness of a quantum state is only increasing by meddling with the side information.

2. *Strong sub-additivity:* Consider a special case of DPI when the given channel is $\mathbb{1}_A \otimes \mathrm{Tr}_C$, we have

$$H(A|BC) \leq H(A|B). \tag{4.20}$$

   If we rewrite this inequality in terms of von Neumann entropy, we have

$$H(ABC) - H(BC) \leq H(AB) - H(B). \tag{4.21}$$

   It follows the so-called *strong sub-additivity*:

$$H(AC|B) \leq H(A|B) + H(C|B), \tag{4.22}$$

   and it automatically implies the *sub-additivity*:

$$H(AC) \leq H(A) + H(C). \tag{4.23}$$

3. *Duality*: Given a pure state ρ_{ABC}, the duality also holds for conditional entropy, i.e.,

$$H(A|B) = -H(A|C). \tag{4.24}$$

   One can simply see it from the duality of quantum entropy.

$$H(A|B) = H(AB) - H(B) = H(C) - H(AC) = -H(A|C). \tag{4.25}$$

The second equality is true because due to the purity condition of ρ_{ABC} we have $H(AB) = H(C)$ and $H(B) = H(AC)$.

4.3.2 Mutual Information

Definition 4.4. *Given a bipartite state ρ_{AB}, the mutual information between A and B is defined as*

$$I(A:B)_\rho := H(A)_\rho + H(B)_\rho - H(AB)_\rho \tag{4.26}$$
$$= H(A)_\rho - H(A|B)_\rho \tag{4.27}$$
$$= H(B)_\rho - H(B|A)_\rho. \tag{4.28}$$

The definition shows that mutual information is symmetric. Mutual information measures the (classical and quantum) correlation between A and B. Here are several properties of mutual information.

1. We can bound the value of mutual information.

$$0 \leq I(A:B) \leq 2\log d, \tag{4.29}$$

   where d is the minimum dimension between Hilbert Spaces A and B. The non-negativity of mutual information is due to the sub-additivity. The upper bound is achieved when ρ_{AB} is a maximally entangled state, i.e., $|\rho\rangle_{AB} = \sum_{i=1}^{d} d^{-1/2} |i\rangle_A |i\rangle_B$. We can verify

$$I(A:B) = H(A) + H(B) - H(AB) = 2H(A) = 2\log d, \tag{4.30}$$

   where the last-second equality comes from the duality of quantum entropy.

2. *Data processing inequality:* Given a quantum channel $\mathcal{E}_{A\to A'}$ and a quantum channel $\mathcal{F}_{B\to B'}$, we have

$$I(A:B)_\rho \geq I(A':B')_{\mathcal{E}\otimes\mathcal{F}(\rho)}. \tag{4.31}$$

   It implies that local operations on A and B cannot increase correlation.

4.4 Relative Entropy

Definition 4.5. *Given two quantum states ρ and σ, the relative entropy between ρ and σ is defined as*

$$D(\rho\|\sigma) := \mathrm{Tr}\left(\rho(\log\rho - \log\sigma)\right). \tag{4.32}$$

Relative entropy serves as a parent quality to other entropic measures. The data processing inequality (DPI) also holds for relative entropy, i.e., for any quantum channel $\mathcal{E}$, we have

$$D(\rho\|\sigma) \geq D(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)). \tag{4.33}$$

It implies all other DPIs we have seen in this chapter. We will explore these implications further in the homework, but here we focus on the data-processing inequality itself by proving it.

4.4.1 Proof of DPI for Relative Entropy

We start with an overview of the technical results that we need for the proof. We then move on to proving each component separately and to stitching them properly that yields the final proof. The proof we use follows what has become a standard method for establishing data processing inequality of quantum divergences: we use matrix analysis to prove joint convexity of the relative entropy and then we use joint convexity to conclude data processing of the quantity.

Part 1: Matrix Analysis

1. *Condition for Positivity of Block Matrices:* For $A, B \geq 0$

$$\begin{pmatrix} A & X \\ X^\dagger & B \end{pmatrix} \geq 0 \iff A \geq XB^{-1}X^\dagger \tag{4.34}$$

2. *Joint Concavity of Matrix Harmonic Mean:* For $A, B \geq 0$,

$$(A, B) \mapsto 2(A^{-1} + B^{-1})^{-1} =: A!B \tag{4.35}$$

   is jointly concave. $A!B$ is known as the Matrix Harmonic Mean.

Part 2: Joint Convexity from Integral Representation

3. *Integral Representation of Logarithm:*

$$\log t = \int_0^\infty \left(\frac{1}{\mu + 1} - \frac{1}{\mu + t} \right) \mathrm{d}\mu \,, \tag{4.36}$$

   where the log is in base e.

4. *Integral Representation of Matrix Logarithm:* For $A \geq 0$,

$$\log(A) = \int_0^\infty \left[\frac{I}{\mu + 1} - (\mu I + A)^{-1} \right] \mathrm{d}\mu \,. \tag{4.37}$$

5. *Integral Representation of Relative Entropy*

$$D(\rho \| \sigma) = \int_0^\infty \langle \Omega | \, \frac{I}{\mu + 1} \otimes I - \left((\mu^{-1}\rho \otimes I)^{-1} + (I \otimes \sigma)^{-1} \right)^{-1} | \Omega \rangle \, \mathrm{d}\mu \,. \tag{4.38}$$

   (a) Applying Item 2 to (4.38) gives the joint convexity of $D(\rho \| \sigma)$.

Part 3: DPI from Joint Convexity

6. *Isometric Invariance:* For all states ρ_A, σ_A such that $\rho \ll \sigma$ and isometry $V_{A \to A'}$,

$$D(V\rho V^\dagger \| V\sigma V^\dagger) = D(\rho \| \sigma) \,. \tag{4.39}$$

7. *1-Designs from Heisenberg-Weyl Operators:* There exists unitary 1-designs $(p_X = \frac{1}{|\mathcal{X}|}, \{U_x\}_{x\in\mathcal{X}})$ in any dimension such that

$$\mathbb{E}_{x\sim p_X}[U_x A U_x^\dagger] = \frac{1}{|\mathcal{X}|}\sum_x U_x A U_x^\dagger = \mathrm{Tr}(A)\cdot\frac{I}{d} \quad \forall A \in \mathcal{L} \tag{4.40}$$

8. *DPI Under Partial Trace:* For all ρ_{AB} and σ_{AB},

$$D(\rho_{AB}\|\sigma_{AB}) \geq D(\rho_A\|\sigma_A) = D(\mathrm{Tr}_B[\rho_{AB}]\|\mathrm{Tr}_B[\sigma_{AB}])\,. \tag{4.41}$$

   (a) By Steinspring representation of a quantum channel, DPI under partial trace implies DPI under all quantum channels.

Proof of Item 1 (Positivity Condition). We start by noticing the following identity:

$$\begin{pmatrix} I & -XB^{-1} \\ 0 & I \end{pmatrix}\begin{pmatrix} I & XB^{-1} \\ 0 & I \end{pmatrix} = I \tag{4.42}$$

where we abuse notation and use the same symbol I for identity, with the dimension being clear from context. This shows that the upper triangular matrix $Y := \begin{pmatrix} I & -XB^{-1} \\ 0 & I \end{pmatrix}$ is invertible. We know that $M \geq 0 \iff X^\dagger M X \geq 0$ for all X. Using this identity and that $Y^{-1}YMY^\dagger {Y^{-1}}^\dagger = M$ by invertibility, we conclude

$$\begin{pmatrix} A & X \\ X^\dagger & B \end{pmatrix} \geq 0 \iff Y\begin{pmatrix} A & X \\ X^\dagger & B \end{pmatrix}Y^\dagger \geq 0 \tag{4.43}$$

$$\iff \underbrace{\begin{pmatrix} A - XB^{-1}X^\dagger & 0 \\ 0 & B \end{pmatrix}}_{\text{Diagonal Operator}} \geq 0 \iff A \geq XB^{-1}X^\dagger\,. \tag{4.44}$$

This establishes (4.34). □

Proof of Item 2 (Concavity of Matrix Harmonic Mean).

We start by showing that $(A, B) \mapsto BA^{-1}B$ is jointly convex. Let's assume $A_1, B_1, A_2, B_2 \geq 0$, $\lambda \in [0, 1]$.
We have that for $i \in \{0, 1\}$

$$\begin{pmatrix} B_i A_i^{-1} B_i & B_i \\ B_i & A_i \end{pmatrix} \geq 0 \tag{4.45}$$

(4.46)

which can be easily verified using property 1 shown above. Taking the convex sum of these positive operators we have

$$\lambda \begin{pmatrix} B_1 A_1^{-1} B_1 & B_1 \\ B_1 & A_1 \end{pmatrix} + (1-\lambda) \begin{pmatrix} B_2 A_2^{-1} B_2 & B_2 \\ B_2 & A_2 \end{pmatrix} \geq 0 \tag{4.47}$$

$$\implies \begin{pmatrix} \lambda B_1 A_1^{-1} B_1 + (1-\lambda) B_2 A_2^{-1} B_1 & \lambda B_1 + (1-\lambda) B_2 \\ \lambda B_1 + (1-\lambda) B_2 & \lambda A_1 + (1-\lambda) A_2 \end{pmatrix} \geq 0 \tag{4.48}$$

$$\implies \lambda B_1 A_1^{-1} B_1 + (1-\lambda) B_2 A_2^{-1} B_1 \tag{4.49}$$

$$\geq (\lambda B_1 + (1-\lambda) B_2)(\lambda A_1 + (1-\lambda) A_2)^{-1} (\lambda B_1 + (1-\lambda) B_2) \tag{4.50}$$

where the last inequality is again due to Property 1 proved earlier. The final line entails the fact that $(A, B) \mapsto BA^{-1}B$ is jointly convex. We now use the Woodbury matrix identity

$$(A^{-1} + B^{-1})^{-1} = \underbrace{B - \underbrace{B(A+B)^{-1}B}_{\text{jointly convex}}}_{\text{jointly concave}} \tag{4.51}$$

Using this identity we finally deduce that $(A, B) \mapsto 2(A^{-1} + B^{-1})^{-1} := A!B$ is jointly concave as described using the under braces. □

Proof of Woodbury Matrix Identity: We have

$$(A^{-1} + B^{-1})^{-1} = (A^{-1}(A+B)B^{-1})^{-1} \tag{4.52}$$

$$= A(A+B)^{-1}B \tag{4.53}$$

$$= (A+B)(A+B)^{-1}B - B(A+B)^{-1}B \tag{4.54}$$

$$= B - B(A+B)^{-1}B\,. \tag{4.55}$$

□

Proof of Item 3 (Integral Representation of Logarithm). We simply write down an integral and show that it simplifies to the correct thing:

$$\int_0^\infty \left(\frac{1}{\mu+1} - \frac{1}{\mu+t}\right) \mathrm{d}\mu \tag{4.56}$$

$$= \lim_{k\to\infty} \int_0^k \left(\frac{1}{\mu+1} - \frac{1}{\mu+t}\right) \mathrm{d}\mu \tag{4.57}$$

$$= \lim_{k\to\infty} \log(k+1) - (\log(k+t) - \log(t)) \tag{4.58}$$

$$= \log(t)\,. \tag{4.59}$$

This establishes (4.36). □

The above is nice because we may use this integral representation to obtain an integral representation of the matrix logarithm.

Proof of Item 4. (Matrix Logarithm Integral Representation). Consider the spectral decomposition $A = \sum_i \lambda_i |\nu_i\rangle\langle\nu_i|$ where $\lambda_i > 0$ for all i. By spectral calculus (Definition 1.18),

$$\begin{aligned}
\log(A) &= \sum_i \log(\lambda_i)|\nu_i\rangle\langle\nu_i| && (4.60)\\
&= \sum_i \int \left(\frac{1}{\mu+1} - (\mu+\lambda_i)^{-1}\right)|\nu_i\rangle\langle\nu_i| \mathrm{d}\mu && (4.61)\\
&= \int \left[\sum_i \left(\frac{1}{\mu+1} - (\mu+\lambda_i)\right)^{-1} |\nu_i\rangle\langle\nu_i|\right] \mathrm{d}\mu && (4.62)\\
&= \int \left[\frac{1}{\mu+1} I - \sum_i (\mu+\lambda_i^{-1})|\nu_i\rangle\langle\nu_i|\right] \mathrm{d}\mu && (4.63)\\
&= \int \left[\frac{1}{\mu+1} I - (\mu I + A)^{-1}\right] \mathrm{d}\mu\,, && (4.64)
\end{aligned}$$

where the second equality is the integral representation of the logarithm, the third is linearity of the integral[1], the fourth is linearity, and the final equality is spectral calculus again. Finally, one can generalize this to $A \geq 0$ by continuity. □

[1] If you are really concerned, note that Fubini's theorem applies.

Therefore, we can replace the matrix logarithm with the integral representation above to get an integral representation of the relative entropy.

Proof of Item 5 (Integral Representation of Relative Entropy). This is effectively a series of manipulations now:

$$\begin{aligned}
D(\rho||\sigma) &= \mathrm{Tr}[\rho(\log\rho - \log\sigma)]\\
&= \langle\Omega|\ (\rho\otimes I)(\log(\rho)\otimes I - I\otimes\log(\sigma))\ |\Omega\rangle\\
&= \langle\Omega|\ (\rho\otimes I)\log(\rho\otimes\sigma^{-1})\ |\Omega\rangle\\
&= \int_0^\infty \langle\Omega|\ (\rho\otimes I)\Big(\frac{I\otimes I}{\mu+1} - (\mu I\otimes I + \rho\otimes\sigma^{-1})^{-1}\Big)\ |\Omega\rangle\ \mathrm{d}\mu\\
&= \int_0^\infty \langle\Omega|\ \frac{1}{\mu+1}\rho\otimes I - (\mu\rho^{-1}\otimes I + I\otimes\sigma^{-1})^{-1}\ |\Omega\rangle\ \mathrm{d}\mu\\
&= \underbrace{\int_0^\infty \frac{1}{\mu+1}\mathrm{Tr}(\rho) - \langle\Omega|\ \underbrace{((\mu^{-1}\rho\otimes I)^{-1} + (I\otimes\sigma)^{-1})^{-1}}_{\text{jointly concave by Item 2}}\ |\Omega\rangle\ \mathrm{d}\mu}_{\text{jointly convex}}\,,
\end{aligned}$$

where the second equality may be verified using the transpose trick assuming that $|\Omega\rangle$ is defined in the eigenbasis of ρ, the third equality you will justify in the homework, the fourth is the integral representation of the matrix logarithm, the fifth is by things commuting, and the sixth is linearity. □

As described by the under braces $D(\rho||\sigma)$ is therefore jointly convex, i.e.,

$$\lambda \cdot D(\rho_1||\sigma_1) + (1-\lambda)\cdot D(\rho_2||\sigma_2) \geq D(\lambda\rho_1 + (1-\lambda)\rho_2||\lambda\sigma_1 + (1-\lambda)\sigma_2)\,. \tag{4.65}$$

The rest of the proof is now trying to find a way to turn joint convexity into data processing. To this end, we will heavily rely upon isometric invariance.

Proof of Item 6 (Isometric Invariance). This follows from a direct calculation using the convention $0 \cdot \log(0) = 0$ so that $\mathrm{Tr}[V\rho V^\dagger \log(V\rho V^\dagger)] = \mathrm{Tr}[\rho\log(\rho)]$ by the spectral calculus (Definition 1.18). The same argument works for $\mathrm{Tr}[\rho\log(\sigma)]$ as we assumed $\rho \ll \sigma$. □

We thus want powerful results about unitaries. The following shows how to build a unitary 1-design, which will be useful due to the isometric invariance.

Proof of Item 7 (1-Designs from Heisenberg-Weyl Operators). For qubits, the Heisenberg-Weyl operators are the Pauli matrices $\{P_i\}_i$ given in (2.4)-(2.7). We will show these are a 1-design Let, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then, we have

$$XAX = \begin{pmatrix} d & c \\ b & a \end{pmatrix},\ YAY = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix},\ ZAZ = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}\,.$$

A direct calculation thus gives

$$\frac{1}{4}\sum_{i=0}^{3} P_i A P_i = \frac{1}{4}[A + XAX + YAY + ZAZ] \tag{4.66}$$

$$= \frac{1}{4}\begin{pmatrix} 2a+2d & 0 \\ 0 & 2a+2d \end{pmatrix} \tag{4.67}$$

$$= \begin{pmatrix} \frac{1}{2}(a+d) & 0 \\ 0 & \frac{1}{2}(a+d) \end{pmatrix} = \mathrm{Tr}(A)\cdot\frac{I}{2} \tag{4.68}$$

Heisenberg-Weyl operators do the same in higher dimensions but we omit the construction here. □

Now observe that this means the 1-design acting on a single marginal is equivalent to tracing out that subsystem and preparing the maximally mixed state. To see this, note that if one decomposes ρ_{AB} into tensor products of linear operators acting on the A and B spaces individually,

$\rho_{AB} = \sum_k X_A^k \otimes Y_B^k$, then $\rho_A = \mathrm{Tr}_B[\rho_{AB}] = \sum_k X_A^k \cdot \mathrm{Tr}[Y_B^k]$. Thus,

$$\frac{1}{|\mathcal{X}|}\sum_x (I \otimes U_x)\rho_{AB}(I \otimes U_x^\dagger) = \sum_k X_A^k \otimes \sum_x U_x Y_B^k U_x^\dagger \tag{4.69}$$

$$= \sum_k X_A^k \otimes \mathrm{Tr}[Y_B^k]\frac{I}{d} \tag{4.70}$$

$$= \left(\sum_k \mathrm{Tr}[Y_B^k] X_A^k\right) \otimes \frac{I}{d} = \rho_A \otimes \frac{I}{d}\,. \tag{4.71}$$

We will use this to prove DPI under partial trace.

Proof of DPI Under Partial Trace. This uses the properties we have identified:

$$\begin{aligned} D(\rho_{AB}\|\sigma_{AB}) &= \frac{1}{|\mathcal{X}|}\sum_x D((I \otimes U_x)\rho_{AB}(I \otimes U_x^\dagger)\|(I \otimes U_x)\sigma_{AB}(I \otimes U_x^\dagger)) \\ &\geq D\Big(\frac{1}{|\mathcal{X}|}\sum_x (I \otimes U_x)\rho_{AB}(I \otimes U_x^\dagger))\|\frac{1}{|\mathcal{X}|}\sum_x (I \otimes U_x)\sigma_{AB}(I \otimes U_x^\dagger)) \\ &= D\Big(\rho_A \otimes \frac{I}{d}\|\sigma_A \otimes \frac{I}{d}\Big) \\ &= D(\rho_A\|\sigma_A) + D\Big(\frac{I}{d}\|\frac{I}{d}\Big) = D(\rho_A\|\sigma_A)\,, \end{aligned}$$

where the first equality is isometric invariance, the inequality is joint convexity, the second equality is (4.71), and the fourth equality is additivity over tensor products. □

4.5 Guessing Probability and Min-Entropy

In this section, we explore the concept of guessing probability and its relation to min-entropy in the context of classical and quantum systems. Suppose that we are given a quantum system B that is correlated with a classical system X. The joint state is denoted as $P_{XB} = \sum_x P_X(x)|x\rangle\langle x| \otimes \rho_B^x$, where $P_X(x)$ is the probability distribution of X, and ρ_B^x is the quantum state of B conditioned on $X = x$. The goal is to guess X by measuring B.

The guessing probability $P_{\text{guess}}(X|B)$ is defined as the maximum probability of correctly guessing X by performing a measurement on B. This involves maximizing over all possible measurements $\{M_x\}$ on B, where $M_x \geq 0$ and $\sum_x M_x = I$. Formally,

$$P_{\text{guess}}(X|B) = \max_{\{M_x\}_x} \sum_x P_X(x)\mathrm{Tr}(M_x\rho_B^x) \tag{4.72}$$

Notice that $\mathrm{Tr}(M_x\rho_B^x)$ represents the probability of measuring outcome x given the state ρ_B^x.

The maximization program defined by equation 4.72 is an example of a something called a Semidefinite program (SDP). To see this, we first rewrite

the guessing probability by using the linearity of the trace operation. The expression can be reformulated as follows:

$$P_{\text{guess}}(X|B) = \max_{\substack{M_{XB}\geq 0 \\ \text{Tr}_X M_{XB}\leq 1_B}} \underbrace{\text{Tr}(M_{XB}\rho_{XB})}_{(1)}. \tag{4.73}$$

Note that the equality has been relaxed in the above inequality because the program is maximized when the equality is achieved. Additionally, note that the expression (1) above can be rewritten as,

$$\text{Tr}(M_{XB}\rho_{XB}) = \text{Tr}\left(M_{XB}\sum_x P_X(x)|x\rangle\langle x|\otimes\rho_B^x\right) = \sum_x P_X(x)\text{Tr}(M_B^x\rho_B^x),$$

where $M_B^x := (\langle x|\otimes I_B)M_{XB}(|x\rangle\otimes I_B)$. We will now introduce semidefinite programs to formally see that indeed (4.73) is an SDP.

4.5.1 *Semidefinite Programming*

This section introduces semidefinite programs (SDPs) and their application to optimization problems in quantum information theory. An SDP is defined as a triple $\{K, L, \mathcal{E}\}$, where $K \in \mathcal{H}(A)$, $L \in \mathcal{H}(B)$, and $E \in \mathcal{L}(\mathcal{H}(A), \mathcal{H}(B))$, i.e. it is a Hermitian-preserving map. Here, $\mathcal{H}(A)$ and $\mathcal{H}(B)$ Hermitian operators on Hilbert spaces A and B respectively, and $\mathcal{L}$ denotes the set of linear operators between them.

Consider the following optimization problem. The primal problem seeks to minimize the trace of KX subject to certain constraints, while the dual problem seeks to maximize the trace of LY under complementary constraints.

$$\text{primal: } \alpha = \inf_{\substack{Y\geq 0 \\ \mathcal{E}(Y)\geq L}} \text{Tr}(KY) \tag{4.74}$$

$$\text{dual: } \beta = \sup_{\substack{Z\geq 0 \\ \mathcal{E}^\dagger(Z)\leq K}} \text{Tr}(LZ) \tag{4.75}$$

SDPs are useful in quantum information because, as we have seen in earlier lectures, many mathematical objects in quantum information theory are positive semidefinite. Thus, often optimizing over a set of quantum objects is an SDP. Moreover, SDPs are useful because they can be solved efficiently (in polynomial time in the dimension of the problem).

We now state two theorems related to SDPs.

Theorem 4.6 (Weak Duality). *For any SDP* (K, L, E), $\alpha \geq \beta$.

This theorem implies that every dual feasible point provides a lower bound on the value of the primal problem, α. This can be useful for building certificates of lower bounds on a function of interest if it is an SDP.

Given weak duality, the natural question is when is it the case $\alpha = \beta$, which is known as *strong duality*. The following is a sufficient condition for strong duality to hold.

Theorem 4.7 (Slater's Criterion for Strong Duality)**.** *If α is finite and there exists $Z \geq 0$ with $\mathcal{E}^\dagger(Z) < K$, then $\alpha = \beta$ and there exists a Y such that $\mathrm{Tr}(KY) = \alpha$.*

4.5.2 Dual SDP formulation of Guessing Probability

Note that by re-labeling the variables in (4.75) according to

$$K \to I_B \quad L \to \rho_{XB} \quad \mathcal{E}^\dagger \to \mathrm{Tr}_X \quad Z \to M_{XB}\,, \tag{4.76}$$

we may conclude that (4.73) is indeed the dual problem of a semidefinite program. Using the adjoint of the partial trace Tr_X is the map $\mathcal{E}(Y_B) = I_X \otimes Y_B$ along with the above identifications, we may conclude the primal problem of the guessing probability is

$$P_{\text{guess}}(X|B) = \min_{\substack{Y_B \geq 0 \\ I_X \otimes Y \geq \rho_{XB}}} \mathrm{Tr}(Y)\,. \tag{4.77}$$

By re-parametering $Y_B = \lambda' \sigma_B$ for some quantum state σ_B and $\lambda' \geq 0$, we can rewrite the SDP as

$$P_{\text{guess}}(X|B) = \min \left\{ \lambda' : \begin{matrix} \rho_{XB} \leq \lambda' \cdot I_X \otimes \sigma_B\,, \mathrm{Tr}(\sigma_B) = 1 \\ \sigma_B \geq 0\,, \lambda' \geq 0 \end{matrix} \right\}. \tag{4.78}$$

This expression will prove useful subsequently.

4.5.3 Min-Entropy

A family of quantities motivated by the von Neumann entropy are the *(optimized) sandwiched Rényi entropies*, defined for $\alpha \in [1/2, 1) \cup (1, \infty)$ and quantum state ρ_{AB} as follows:

$$\widetilde{H}^{\uparrow}_{\alpha}(A|B)_\rho := \max_{\substack{\sigma_B \geq 0 \\ \mathrm{Tr}(\sigma_B)=1}} \frac{1}{1-\alpha} \log \mathrm{Tr}\left[\left(I_A \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}} \rho_{AB} I_A \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right]. \tag{4.79}$$

That these are motivated by the von Neumann entropy follows from the fact

$$\lim_{\alpha \to 1} \widetilde{H}^{\uparrow}_{\alpha}(A|B)_\rho = H(A|B)_\rho\,. \tag{4.80}$$

We may consider the other two limits of the parameterization:

$$\lim_{\alpha \to \infty} \widetilde{H}^{\uparrow}_{\alpha}(A|B)_\rho := H_{\min}(A|B)_\rho \tag{4.81}$$

$$\widetilde{H}^{\uparrow}_{1/2}(A|B)_\rho := H_{\max}(A|B)_\rho\,, \tag{4.82}$$

where the choice of 'min' and 'max' labels follows from the fact (which we do not prove) that sandwiched Rényi entropy is monotonically decreasing in the parameter α. By an argument that we omit, one will find the min-entropy has the following simple form:

$$H_{\min}(A|B)_\rho = \min\{\lambda \in \mathbb{R} : \rho_{AB} \leq 2^{-\lambda} I_A \otimes \sigma_B \,,\, \sigma_B \geq 0 \,,\, \mathrm{Tr}(\sigma_B) = 1\} \,. \tag{4.83}$$

By letting the A system be a classical system X, one concludes from (4.83) and (4.78) that

$$P_{\text{guess}}(X|B)_\rho = \exp(-H_{\min}(X|B)_\rho) \,. \tag{4.84}$$

This establishes a connection between min-entropy and the guessing probability.

As a bonus, we prove the following relation, which is easier to prove than the monotonicity of the sandwiched Rényi entropies in α and justifies our labels of min- and max-entropy.

Lemma 4.8.

$$H_{\min}(A|B) \leq H(A|B) \leq H_{\max}(A|B) \tag{4.85}$$

Proof. First, note that by (4.83), $\rho_{AB} \leq 2^{-\lambda} I_A \otimes \sigma_B$ where $\lambda = H_{\min}(A|B)_\rho$. We use this to bound the conditional entropy $H(A|B)$:

$$\begin{aligned} H(A|B) &= -\mathrm{Tr}\left(\rho_{AB}\left(\log \rho_{AB} - \log I_A \otimes \rho_B\right)\right) && (4.86)\\ &\geq \lambda - \mathrm{Tr}\left(\rho_B\left(\log \sigma_B - \log \rho_B\right)\right) && (4.87)\\ &= \lambda + D(\rho_B \| \sigma_B) \geq \lambda \,, && (4.88) \end{aligned}$$

where we used the non-negativity of the relative entropy between states. This implies $H(A|B) \geq H_{\min}(A|B)$.

As you will prove in your homework, for a pure state ψ_{ABC}, $H_{\min}(A|B)_\psi = -H_{\max}(A|C)_\psi$. Thus, let ρ_{ABC} be the purification of ρ_{AB} and one has

$$H(A|B)_\rho = -H(A|C)_\rho \leq -H_{\min}(A|C) = H_{\max}(A|B) \,, \tag{4.89}$$

where we used the duality of the von Neumann entropy and that we had already established $H(A|C) \geq H_{\min}(A|C)$. This concludes the proof. □

4.6 *Randomness Extraction*

In this section, we discuss randomness extraction. Above, we showed that the min-entropy quantifies the guessing probability of a classical-quantum state. If the B register is correlated with the X register, the guessing probability might be quite high. The goal of randomness extraction is to take a correlated ρ_{XB} and act solely on the X register to obtain a new classical register Z such that the Z register is approximately decorrelated or 'decoupled' from the B register. More formally, given a classical-quantum state ρ_{XB}, The goal is to produce a state ρ_{ZB} that has the following properties:

- Z is (close to) uniformly random.
- Z is (close to) independent of B.

This can be captured by the trace distance.

$$\Delta(\rho_{ZB}, \pi_Z \otimes \rho_B) \leq \epsilon, \tag{4.90}$$

where $\pi_Z = \frac{I_Z}{d_Z}$ is the uniform distribution, and $\rho_B = \sum_z \rho_z \rho_B^z$ is the marginal of ρ_{ZB}. As it's natural to want as much randomness as you can get from ρ_{XB}, the goal is to maximize d_Z such that the inequality 4.90 holds.

Before continuing, we show that the above is operationally relevant in the sense that if we perform randomness extraction, then anyone with access to B can do at most ϵ better than randomly guessing the value of Z.

Lemma 4.9. *Condition 4.90 implies the following bound:*

$$P_{guess}(Z|B) \leq \frac{1}{d_Z} + \epsilon \tag{4.91}$$

Proof. We start with the trace distance:

$$\epsilon \geq \Delta(\rho_{ZB}, \pi_Z \otimes \rho_B) = \max_{I \geq M \geq 0} \mathrm{Tr}[M(\rho_{ZB} - \pi_Z \otimes \rho_B)]$$

Choose

$$M = \sum_z |z\rangle\langle z| \otimes M_z,$$

where $\{M_z\}_z$ is a POVM. Then,

$$\begin{aligned} \epsilon &\geq \max_{\{M_z\}_z} \left(\sum_z P_z \mathrm{Tr}(M_z \rho_B^z) - \sum_z \frac{1}{d_Z} \mathrm{Tr}(M_z \rho_B) \right) \\ &= P_{\text{guess}}(Z|B) - \frac{1}{d_Z}\,. \end{aligned}$$

Re-ordering the inequality establishes the desired bound. □

4.6.1 Extractors & Two-Universal Families of Functions

Given ρ_{XB}, to convert the register X into a register Z such that it is harder to guess Z from B, we need to compute some function of X. However, any single function $f : \mathcal{X} \to \mathcal{Z}$ won't work as then guessing $f(Z)$ can only be easier than guessing X.[2] As such, we need to inject some extra randomness by choosing a function f_s from a family of functions $\{f_s\}_{s \in \mathcal{S}}$ according to some distribution q_S. The idea is then to one of these functions according to q_S and show that

$$\Delta(\rho_{ZBS}, \pi_Z \otimes \rho_B \otimes q_s) = \mathbb{E}_s[\Delta(\rho_{f_s(X)B}, \pi_Z \otimes \rho_B)] \leq \epsilon\,, \tag{4.92}$$

[2] Indeed, for any function $f : \mathcal{X} \to \mathcal{Z}$ and state ρ_{XB}, $H_{\min}(f(X)|B)_\rho \leq H(X|B)_\rho$ (Lemma 4.15).

where the right hand side writes the distance as an expectation over the value of the seed s and the states are defined as

$$\rho_{ZBS} = \sum_{x,s} p_x q_s \left|f_s(x)\right\rangle \left\langle f_s(x)\right|_z \otimes \rho_B^x \otimes \left|s\right\rangle \left\langle s\right| \tag{4.93}$$

$$\rho_{f_s(X)B} = \sum_z \sum_x \mathbb{1}\{f_s(x) = z\} p_x \left|z\right\rangle \left\langle z\right| \otimes \rho_B^x \,. \tag{4.94}$$

We call the S register the 'seed,' because it's (generally) a small amount of uniform randomness that helps grow the weak randomness of X into a lot of uniform randomness Z. The procedure of using a seed to extract near private randomness from a weakly random source is called an 'extractor,' i.e. we identify the extractor with distribution q_S and the family of functions $\{f_s\}$.

For us, we will use a two-universal familiy of functions as an extractor.

Definition 4.10. *A two-universal family of functions is a set of functions $\{f_s\}_{s\in\mathcal{S}}$ from $\mathcal{X} \to \mathcal{Z}$ equipped with a probability mass function q_S such that*

$$\Pr_S[f_S(x) = f_S(x')] \leq \frac{1}{|\mathcal{Z}|} \quad \forall x \neq x'. \tag{4.95}$$

Note that what (4.95) says is that the probability (over the choice of 'seed' $s \in \mathcal{S}$) of two distinct inputs taking the same value (colliding) is at most inverse in the output dimension. We may re-write (4.95) as

$$\mathbb{E}_S\left[\mathbb{1}\{f_s(x) = f_s(x')\}\right] \leq \frac{1}{d_z}. \tag{4.96}$$

Example. *Consider field $\mathbb{F}_{2^n}$ and seed $s \in \mathbb{F}_{2^n}\backslash\{0\}$, P_s is uniform. We can define the family $f_s : \{0,1\}^n \mapsto \{0,1\}^l$ for $l \leq n$ as*

$$x \mapsto x \cdot s \pmod{2^l}. \tag{4.97}$$

For this family, we see that $f_s(x) = f_s(x') \iff xs = x's \pmod{2^l} \iff (x - x') \cdot s = k \cdot 2^l$ for some $k \in \mathbb{N}$. However, since $(x - x') \cdot s$ for $x \neq x'$ generates the whole field as we iterate over s, we can deduce that $\Pr[f_s(x) = f_s(x')] = \frac{1}{2^n-1}(2^{n-l} - 1) \leq \frac{2^{n-l}}{2^n} = 2^{-l}$, where in the last inequality we used that $\frac{a-1}{b-1} \leq \frac{a}{b}$ for $1 < a \leq b$.

4.6.2 Leftover Hashing lemma

We now turn to proving that we can use a family of two-universal functions as an extractor. To show such a result, we will need to following inequality:

Theorem 4.11 (Hölder's inequality)**.**

$$||XY||_r \leq ||X||_q ||Y||_p \quad \text{for} \quad \frac{1}{q} + \frac{1}{p} = \frac{1}{r}. \tag{4.98}$$

In particular, we have

$$||ABC||_1 \leq ||AB||_{\frac{4}{3}} ||C||_4 \leq ||A||_4 ||B||_2 ||C||_4 . \tag{4.99}$$

Theorem 4.12. *Let ρ_{XB} be a classical-quantum state and $(q_S, \{f_s : \mathcal{X} \to \mathcal{Z}\}_s)$ be a two-universal family of hash functions (Definition 4.10). Then,*

$$\mathbb{E}_s[\Delta(\rho_{f_s(x)B}, \pi_Z \otimes \rho_B)] \leq \frac{1}{2}\sqrt{d_z 2^{-\mathrm{H}_{\min}(X|B)}} . \tag{4.100}$$

Proof. We start by using the Hölder inequality to split the trace norm into three parts, i.e.

$$\begin{aligned}
&2\Delta(\rho_{f_s(x)B}, \pi_Z \otimes \rho_B) \\
&= ||\rho_B^{\frac{1}{4}} \rho_B^{-\frac{1}{4}} (\rho_{f_s(x)B} - \pi_Z \otimes \rho_B) \rho_B^{\frac{-1}{4}} \rho_B^{\frac{1}{4}}||_1 && (4.101)\\
&\leq ||I_Z \otimes \rho_B^{\frac{1}{4}}||_4 \, ||\rho_B^{-\frac{1}{4}} (\rho_{f_s(x)B} - \pi_Z \otimes \rho_B) \rho_B^{-\frac{1}{4}}||_2 \, ||I_Z \otimes \rho_B^{\frac{1}{4}}||_4 && (4.102)\\
&= d_z^{\frac{1}{4}} \sqrt{\mathrm{Tr}(\rho_B^{-\frac{1}{2}} (\rho_{f_s(x)B} - \pi_Z \otimes \rho_B) \rho_B^{-\frac{1}{2}} (\rho_{f_s(x)B} - \pi_Z \otimes \rho_B))} \, d_z^{\frac{1}{4}} && (4.103)\\
&= \sqrt{d_z \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_{f_s(x)B} \rho_B^{-\frac{1}{2}} \rho_{f_s(x)B}) - 1} , && (4.104)
\end{aligned}$$

where in the first equality we used $\rho_B^{\frac{1}{4}} \rho_B^{-\frac{1}{4}} = I_B$. Using Jensen's inequality, we can furthermore pull the expectation into the square, which yields

$$\mathbb{E}_s[2\Delta(\rho_{f_s(x)B}, \pi_Z \otimes \rho_B)] \leq \sqrt{d_z \mathbb{E}_s \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_{f_s(x)B} \rho_B^{-\frac{1}{2}} \rho_{f_s(x)B}) - 1}$$

It remains to bound the expectation term. We do this by expanding the term using (4.94) and then using the property of two-universal function family as written in (4.96):

$$\begin{aligned}
&\mathbb{E}_s[\mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_{f_s(x)B} \rho_B^{-\frac{1}{2}} \rho_{f_s(x)B})] \\
&= \mathbb{E}_s[\sum_z \sum_{x,x'} p_x p_{x'} \mathbb{1}\{f_s(x) = z\} \, \mathbb{1}\{f_s(x') = z\} \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_B^x \rho_B^{-\frac{1}{2}} \rho_B^{x'})] \\
&= \mathbb{E}_s[\sum_{x,x'} p_x p_{x'} \, \mathbb{1}\{f_s(x) = f_s(x')\} \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_B^x \rho_B^{-\frac{1}{2}} \rho_B^{x'})] \\
&= \sum_x p_x^2 \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_B^x \rho_B^{-\frac{1}{2}} \rho_B^x) \\
&\quad + \sum_{x \neq x'} p_x p_{x'} \mathbb{E}_s[\mathbb{1}\{f_s(x) = f_s(x')\}] \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_B^x \rho_B^{-\frac{1}{2}} \rho_B^{x'}) \\
&\leq \sum_x p_x^2 \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_B^x \rho_B^{-\frac{1}{2}} \rho_B^x) + \frac{1}{d_z} \sum_{x \neq x'} p_x p_{x'} \cdot \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_B^x \rho_B^{-\frac{1}{2}} \rho_B^{x'}) \\
&= \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_{XB} \rho_B^{-\frac{1}{2}} \rho_{XB}) + \frac{1}{d_z} \sum_{x \neq x'} p_x p_{x'} \cdot \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_B^x \rho_B^{-\frac{1}{2}} \rho_B^{x'}) \\
&\leq \mathrm{Tr}(\rho_B^{-\frac{1}{2}} \rho_{XB} \rho_B^{-\frac{1}{2}} \rho_{XB}) + \frac{1}{d_z} \mathrm{Tr}[\rho_B^{-\frac{1}{2}} (\sum_x p_x \rho_B^x) \rho_B^{-\frac{1}{2}} (\sum_{x'} p_{x'} \rho_B^{x'})]
\end{aligned}$$

$$= \mathrm{Tr}(\rho_B^{-\frac{1}{2}}\rho_{XB}\rho_B^{-\frac{1}{2}}\rho_{XB}) + \frac{1}{d_z} \tag{4.105}$$

$$= 2^{-\lambda} + \frac{1}{d_z} \tag{4.106}$$

Here, to get the last equality, we bound the trace using the relation $\rho_{XB} \leq 2^{-\lambda}\mathbb{I}_x \otimes \sigma_B$ with $\lambda = H_{\min}(X|B)$ on one of the copies of ρ_{XB} and simplifying. □

Interpretation: We want this expectation to be small, i.e.,

$$\frac{1}{2}\sqrt{d_z 2^{-\mathrm{H}_{\min}(X|B)}} \leq \epsilon \iff d_z 2^{-\mathrm{H}_{\min}(X|B)} \leq (2\epsilon)^2 \tag{4.107}$$

$$\iff d_z \leq (2\epsilon)^2 2^{\mathrm{H}_{\min}(X|B)} \tag{4.108}$$

$$\iff \log d_z \leq \mathrm{H}_{\min}(X|B) - 2\log\frac{1}{2\epsilon} \tag{4.109}$$

where $\log d_z$ is the number of random bits. If $\mathrm{H}_{\min}(X|B)$ is large compared to $\log\frac{1}{\epsilon}$, this essentially means that we can extract $\mathrm{H}_{\min}(X|B)$ random bits as the $\log\frac{1}{2\epsilon}$ term can be neglected. Then what it tells us is that as long as we choose the number of bits that we extract using this method to be strictly smaller than the min-entropy, then we get an output that indeed satisfies this condition.

Finally, it will be useful to extend the above to the smooth min-entropy to show near optimality as well as for our application of the result to QKD in the next chapter.

Definition 4.13. *Let ρ_{AB} be a quantum state. Consider the ball of subnormalized states*

$$\mathcal{B}^\varepsilon(\rho) := \{\widetilde{\rho}_{AB} \geq 0 : P(\widetilde{\rho}, \rho) \leq \varepsilon\} . \tag{4.110}$$

Then the smooth min-entropy is

$$H_{\min}^\varepsilon(A|B)_\rho := \max_{\widetilde{\rho}\in\mathcal{B}^\varepsilon(\rho)} H_{\min}(A|B)_\rho . \tag{4.111}$$

We note that without loss of generality the optimizer is a quantum state. Using this, we can establish the following.

Proposition 4.14. *Let $\varepsilon, \delta > 0$ such that $\varepsilon > 2\delta$. Let ρ_{XB} be a classical-quantum state. It is the case that $\Delta(\rho_{ZBS}, \pi_Z \otimes \rho_B \otimes q_S) \leq \varepsilon$ if*

$$\log(d_Z) \leq H_{\min}^\delta(X|B)_\rho - 2\log\frac{1}{2(\varepsilon - 2\delta)} . \tag{4.112}$$

Proof. Let $\widetilde{\rho}$ be the state such that $H_{\min}^\delta(A|B)_\rho = H_{\min}(A|B)_{\widetilde{\rho}}$. Then by the previous theorem, $\Delta(\tilde{\rho}_{ZBS}, \pi_Z \otimes \tilde{\rho}_B \otimes q_S) \leq \frac{1}{2}\sqrt{d_z 2^{-\mathrm{H}_{min}(A|B)_{\widetilde{\rho}}}}$. As $\Delta(\widetilde{\rho}, \rho) \leq \delta$ by definition of the smoothing ball, we may apply the triangle inequality twice to conclude

$$\Delta(\rho_{ZBS}, \pi_Z \otimes \rho_B \otimes q_S) \leq 2\delta + \frac{1}{2}\sqrt{d_z 2^{-\mathrm{H}_{min}^\delta(A|B)_\rho}} . \tag{4.113}$$

Re-ordering completes the proof. □

A particularly convenient version of the above is if $\delta = \varepsilon/4$ so that one has

$$\log(d_Z) \leq H_{\min}^{\varepsilon/4}(X|B)_\rho - 2\log(1/\varepsilon) \,. \tag{4.114}$$

4.6.3 *Optimality*

The bound on the size of the extracted randomness in (4.114) is roughly tight as we can show that $\log(d_z) \geq H_{\min}^{\varepsilon'}(A|B)$ for $\varepsilon' = \sqrt{2\varepsilon - \varepsilon^2}$ is necessary for any extractor to satisfy the trace distance criteria (4.90). The argument is based on the following lemma, which we will state here without proof.

Lemma 4.15. *Let $f : \mathcal{X} \to \mathcal{Z}$ be a function, ρ_{XB} be a classical-quantum state, and $\varepsilon \geq 0$. Then,*

$$H_{\min}^{\varepsilon}(f(X)|B) \leq H_{\min}^{\varepsilon}(X|B) \,. \tag{4.115}$$

In other words, applying a function can at most reduce the conditional entropy of a random variable.

Proposition 4.16. *For a classical-quantum state ρ_{XB}, an extractor satisfies $\Delta(\rho_{ZBS}, \pi_Z \otimes \rho_B \otimes q_S) \leq \varepsilon$ only if $\log(d_z) \leq H_{\min}^{\varepsilon'}(A|B)$ for $\varepsilon' = \sqrt{2\varepsilon - \varepsilon^2}$.*

Proof. We prove the contrapositive. Let $\log(d_Z) > H_{\min}^{\varepsilon'}(X|B)_\rho$. By the previous lemma and our assumed smooth min-entropy bound,

$$\log(d_Z) > H_{\min}^{\varepsilon'}(Z|B)_{\rho_{f_s(X)B}} \quad \forall s \in \mathcal{S} \,. \tag{4.116}$$

Now observe that if $P(\rho_{f_s(X)B}, \pi_Z \otimes \rho_B) \leq \varepsilon'$, then, as the smooth min-entropy is a maximization,

$$H_{\min}^{\varepsilon'}(Z|B)_{\rho_{f_s(X)B}} \geq H_{\min}(Z|B)_{\pi_Z \otimes \rho_B} = \log(d_Z) \,. \tag{4.117}$$

Combining (4.116) and (4.117), we can conclude

$$P(\rho_{f_s(X)B}, \pi_Z \otimes \rho_B) > \varepsilon' \quad \forall s \in \mathcal{S} \,. \tag{4.118}$$

Using that, by the Fuchs-van de Graaf inequality, $\sqrt{2\Delta(\rho,\sigma) - \Delta(\rho,\sigma)^2} \geq P(\rho,\sigma)$, we conclude

$$\Delta(\rho_{f_s(X)B}, \pi_Z \otimes \rho_B) > \varepsilon \quad \forall s \in \mathcal{S} \,. \tag{4.119}$$

However, then

$$\Delta(\rho_{ZBS}, \pi_Z \otimes \rho_B \otimes q_S) = \sum_s q_S(s) \Delta(\rho_{f_s(X)B}, \pi_Z \otimes \rho_B) > \varepsilon \,. \tag{4.120}$$

Thus, an extractor cannot satisfy $\Delta(\rho_{ZBS}, \pi_Z \otimes \rho_B \otimes q_S) \leq \varepsilon$ if $\log(d_Z) > H_{\min}^{\varepsilon'}(X|B)$. □

Note the above shows even if the seed were kept private, one would still need sufficient min-entropy.

4.7 BONUS: Entropic Uncertainty Relations

We now discuss a specific application of entropy and relative entropy, namely an entropic formulation of the famous Heisbenberg uncertainty principle.

Consider two rank-1 projective measurements on a system A

$$\{M_x\}_x, \quad M_x = |\phi_x\rangle \langle\phi_x| \tag{4.121}$$

$$\{N_z\}_z, \quad N_z = |\psi_z\rangle \langle\psi_z| \tag{4.122}$$

with the corresponding measurment channels

$$\mathcal{M}(\cdot) = \sum_x |x\rangle\langle\phi_x| \cdot |\phi_x\rangle\langle x|, \quad \mathcal{N}(\cdot) = \sum_z |z\rangle\langle\phi_z| \cdot |\phi_z\rangle\langle z|. \tag{4.123}$$

Theorem 4.17. *Given a tripartite state ρ_{ABC}, define*

$$\rho_{XBC} = (\mathcal{M}_A \otimes I_{BC})\, \rho_{ABC}, \quad \rho_{ZBC} = (\mathcal{N}_A \otimes I_{BC})\, \rho_{ABC}. \tag{4.124}$$

Then we have that

$$H(X|B)_\rho + H(Z|C)_\rho \geq \log \frac{1}{c} \tag{4.125}$$

where $c = \max_{x,z} |\langle\phi_x|\psi_z\rangle|^2$ is the maximum overlap between two the ONBs $\{\phi_x\}_x, \{\psi_z\}_z$.

The above theorem says that either Bob (B) has high uncertainty about X or Charlie (C) has high uncertainty about Z. In a typical cryptographic application, if Alice (A) and Bob can convince themselves that B could know X then they can rest assured that Charlie, the eavesdropper, does not know Z. We shall use this more formally later in this course.

In case of qubits, if we consider measurements in the computational and the diagonal bases, we have

$$c = |\langle +|0\rangle|^2 = |\langle +|1\rangle|^2 = |\langle -|0\rangle|^2 = |\langle -|1\rangle|^2 = \frac{1}{2} \tag{4.126}$$

$$\implies H(X|B) + H(Z|C) \geq 1 \tag{4.127}$$

As a special case if B, C are trivial we have the *Maassen-Uffink relation*

$$H(X) + H(Z) \geq \log \frac{1}{c}. \tag{4.128}$$

Proof. We need the following technical instruments:

- Duality of entropy;
- Operator monotonicity of log i.e., $A \geq B \implies \log A \geq \log B\,, \forall\, A, B > 0$; (We shall give a proof of this fact after we conclude the main proof.)
- the DPI for relative entropy.

We start by assuming ρ_{ABC} to be pure. Now, as the Steinspring dilation of the measurement $\mathcal{N}$, we use $V : A \mapsto A'Z$ with

$$V = \sum_z |z\rangle_z |z\rangle_{A'} \langle\psi_z|_{A'} \tag{4.129}$$

where A' is isomorphic to A. Thus we can verify that

$$V^\dagger V = \sum_{z,z'} \langle z|\, z'\rangle \cdot \langle z|\, z'\rangle \cdot |\psi_z\rangle \langle\psi_{z'}| = \sum_z |\psi_z\rangle \langle\psi_z| = I, \tag{4.130}$$

which makes V an isometry. We now define the states

$$\tilde{\rho}_{ZABC} = (V \otimes I_{BC})\rho_{ABC}(V^\dagger \otimes I_{BC}), \tag{4.131}$$

$$\tilde{\rho}_{ZC} = \sum_z |z\rangle \langle z| \otimes (\langle\psi_z| \otimes I_c)\rho_{AC}(|\psi_z\rangle \otimes I_c) \tag{4.132}$$

Using the entropic duality relation in the first step, we find

$$H(Z|C)_{\tilde{\rho}} = -H(Z|AB)_{\tilde{\rho}} \tag{4.133}$$

$$= D(\tilde{\rho}_{ZA'B} \| I_Z \otimes \tilde{\rho}_{A'B}) \tag{4.134}$$

$$\geq D\Big(\rho_{AB} \| V^\dagger (I_Z \otimes \tilde{\rho}_{A'B}) V\Big) \tag{4.135}$$

$$= D\Big(\rho_{AB} \| \sum_z |\psi_z\rangle \langle\psi_z| \otimes (\langle z| \otimes I_B)\tilde{\rho}_{A'B}(|z\rangle \otimes I_B)\Big) \tag{4.136}$$

$$\geq D\Big(\rho_{XB} \| \sum_z \mathcal{M}(|\psi_z\rangle \langle\psi_z|)) \otimes (\langle z| \otimes I_B)\tilde{\rho}_{A'B}(|z\rangle \otimes I_B)\Big) \tag{4.137}$$

$$= D\Big(\rho_{XB} \| \sum_{x,z} |\langle\phi_x|\, \psi_z\rangle|^2 |x\rangle \langle x| \otimes (\langle z| \otimes I_B)\tilde{\rho}_{A'B}(|z\rangle \otimes I_B)\Big) \tag{4.138}$$

$$\geq D\Big(\rho_{XB} \| \sum_{x,z} K \cdot |x\rangle \langle x| \otimes (\langle z| \otimes I_B)\tilde{\rho}_{A'B}(|z\rangle \otimes I_B)\Big) \tag{4.139}$$

$$= D(\rho_{XB} \| I_x \otimes \mathrm{Tr}_{A'}(\tilde{\rho}_{A'B})) - \log c \tag{4.140}$$

$$= D(\rho_{XB} \| I_x \otimes \rho_B) - \log c \tag{4.141}$$

$$= -H(X|B) + \log \frac{1}{c} \tag{4.142}$$

This finishes the proof for a pure ρ_{ABC}. For a generic state we use purification and DPI for partial trace to obtain the same result. □

Proof of Operator Monotonicity of log for $A, B > 0$: We have $A \geq B \implies A^{-1} \leq B^{-1}$, because

$$A - B \geq 0 \tag{4.143}$$

$$\implies B^{-1/2}(A - B)B^{-1/2} \geq 0 \tag{4.144}$$

$$\implies B^{-1/2}AB^{-1/2} \geq I \tag{4.145}$$

$$\implies B^{1/2}A^{-1}B^{1/2} \leq I \tag{4.146}$$

$$\implies A^{-1} \leq B^{-1} \tag{4.147}$$

Now,

$$\log A - \log B = \int_o^\infty (\mu I + B)^{-1} - (\mu I + A)^{-1} \, \mathrm{d}\mu \tag{4.148}$$

Since $A \geq B$, we have $(\mu I + B) \leq (\mu I + A)$ and therefore, $(\mu I + B)^{-1} \geq (\mu I + A)^{-1}$. Since the integrand is non-negative we have $\log A \geq \log B$.

5
Quantum Key Distribution

In this chapter, we will bring together some of the concepts already introduced to prove security of quantum key distribution. This has three motivations. First, it will show how the concepts introduced so far in this course are relevant for current quantum information processing. Second, it will introduce you to some key ideas in information theory that are relevant to quantum “Shannon" theory in general. Third, some of the ideas introduced in this chapter get at key ideas used more broadly in defining and analyzing quantum cryptographic protocols, and thus this chapter also serves as an introduction to quantum cryptography. The chapter will culminate with you (in the homework) calculating the asymptotic key rate of the BB84 quantum key distribution protocol.

5.1 Motivation: Secret Keys and Entanglement

This section builds up motivation for the rest of the chapter. We begin by explaining why we should care about quantum key distribution. We will see secret key encryption gives perfect privacy over an authenticated classical channel, but that it is not possible to establish a secret key over an authenticated classical channel. Then we will show if Alice and Bob shared a maximally entangled state, they could establish a single bit of secret key. This shows if Alice could distribute entanglement to Bob in a manner where they could check whether or not they share the entanglement, then they could generate a secret key. Alice and Bob can check whether or not they share entanglement, so they can generate key. This is what QKD does. In other words, we care about quantum key distribution because it uses quantum mechanics to achieve a cryptographic task that classically cannot be done.

5.1.1 Secret Key Encryption

We begin by explaining what secret key encryption is and a problem with using it that quantum mechanics will resolve.

Alice and Bob share a communication channel where any message that is transmitted cannot be altered, but can in principle be listened to. Such a classical channel is known as an authenticated channel. Alice would like to send a message of ℓ-bits, $m \in \{0,1\}^\ell$, to Bob over the authenticated channel. She would like anyone who looks at the message over the classical channel to know nothing more about the message than they knew previously (perfect secrecy) and she would like Bob to learn the message m (perfect correctness). Secret key encryption is a method for achieving this. Like all encryption algorithms, it is made up of an encryption algorithm and a decryption algorithm. The algorithms for secret key encryption will make use of a secret key (hence the name). The algorithms are as follows.

1. **Encryption:** The algorithm produces a cipher text by bit-wise XORing the message with the key:

$$\mathrm{Enc}(k,m) = c := k \oplus m = (k_1 \oplus m_1, k_2 \oplus m_2, \ldots, k_\ell \oplus m_\ell)\,. \tag{5.1}$$

2. **Decryption:** The algorithm decrypts by bit-wise XORing the received cipher text with the key:

$$\mathrm{Dec}(k,c) = \hat{k} := c \oplus k = (k_1 \oplus c_1, k_2 \oplus c_2, \ldots, k_\ell \oplus c_\ell)\,. \tag{5.2}$$

The secret key encryption scheme is then as follows:

1. Alice computes ciphertext $c = \mathrm{Enc}(k,m)$.
2. Alice sends c over the authenticated channel to Bob.
3. Bob receives c as the channel is authenticated.
4. Bob decrypts the ciphertext obtaining a guess of the message, $\hat{m} = \mathrm{Dec}(k,c)$.

Using the definitions of the encryption and decryption algorithms and the assumption on the classical channel (so that Bob's received c is equal to the c Alice computed), one can conclude $\hat{m} = m$ always. That is, this encryption algorithm is perfectly correct.

To formalize perfect secrecy, we require a formal definition of a secret key.

Definition 5.1. *Let $\ell \in \mathbb{N}$ and E be a quantum register. A ℓ-long binary secret key is*

$$\rho_{KK'E} = \sum_{k\in\{0,1\}^\ell} \frac{1}{|\{0,1\}^\ell|} |k\rangle\langle k|_K \otimes |k\rangle\langle k|_{K'} \otimes \rho_E := \chi_{KK'} \otimes \rho_E\,. \tag{5.3}$$

This definition shows that the E system, which we can view as held by an eavesdropper, is independent of the KK' systems. This partially captures

what it means for the key to be "secret." It also shows that the value of the registers K and K' are always the same and uniformly random. That it is always the same is what makes it a *shared* key. That the value of the key is uniformly random is the other property that makes the key "secret." We now have enough to establish perfect secrecy.

Proposition 5.2. *Secret key encryption is perfectly secure.*

Proof. We will use a proof that uses some of the modeling tools we use throughout this chapter. Let $q_M = \sum_{m\in\{0,1\}^\ell} q(m)|m\rangle\langle m|_M$ be a distribution over Alice's messages. The total state generated over the protocol is $\rho_{MCC_E\hat{M}E}$ where M is the initial message, C is the ciphertext, C_E is Eve's copy of the cipher text, $\hat{M}$ is Bob's recovered version of the message, and E is Eve's initial system from the secret key. The relevant marginal is the joint state between the message M, the copy of the ciphertext that Eve acquires when Alice sends the ciphertext over the channel C_E, and the E register of the secret key, which Eve also has. That is,

$$\rho_{MC_EE} = \sum_{m\in\{0,1\}^\ell, k\in\{0,1\}^\ell} \frac{q(m)}{2^\ell}|m\rangle\langle m|_M \otimes |m\otimes k\rangle\langle m\otimes k|_{C_E}\otimes\rho_E \quad (5.4)$$

$$= q_M \otimes \frac{1}{2^\ell}\mathbb{1}\otimes\rho_E \quad (5.5)$$

$$\coloneqq q_M\otimes\pi_{C_E}\otimes\rho_E\,, \quad (5.6)$$

where the first equality uses that C_E is a copy of C and how C is generated and the second equality we used that for fixed m, summing over all possible values of k results in all possible values in $\{0,1\}^\ell$. The above shows the ciphertext and E register are independent of the message register, and thus Eve cannot guess the value of M any better with the copy of the ciphertext.[1] □

The problem: The above shows that a secret key is extremely useful for private communication. However, how are Alice and Bob to share a secret key? If they do not already share a secret key, then Alice could try to build a secret key by uniformly randomly picking an ℓ-bit string and sending a copy of the value to Bob. However, an eavesdropper could copy the value of k, so that it is not a secret key.[2] Indeed, there is no way in classical information theory for Alice and Bob to establish a shared secret key over a channel without placing further limitations on what Eve learns that is communicated over the channel.

5.1.2 Secret Key from Maximally Entangled State

Now imagine Alice and Bob share the same authenticated channel as previously, but also share a maximally entangled state $|\Phi^+\rangle_{AB} = \sum_{i\in\{0,1\}}\frac{1}{\sqrt{2}}|i\rangle|i\rangle$. Consider the following simple scheme:

[1] A more classical proof, ignoring the E register, is to note what Eve would do is guess the value of the message m conditioned on the value of the ciphertext c. Thus, what matters is the probability of m given a value of the ciphertext c, $p(m|c)$. Then one has

$$p(m|c) = \frac{p(m,c)}{p(c)} = \frac{q(m)2^{-\ell}}{2^{-\ell}} = q(m)\,,$$

where the first equality is Bayes' rule and the second is the independence between the value of m and the value of the ciphertext c observed in (5.5). Thus, one sees the value of the ciphertext does not help guess the value of m, so secret key encryption is by definition perfectly secure.

[2] Indeed, the resulting state is $\rho_{KK'E} = \sum_{k\in\{0,1\}^\ell} = |k\rangle\langle k|_K\otimes|k\rangle\langle k|_{K'}\otimes|k\rangle\langle k|_E$ which one can see does not satisfy Definition 5.1.

1. Alice and Bob independently with equal probability measure their part of the maximally entangled state in the computational basis, $\{|0\rangle, |1\rangle\}$, or Hadamard basis, $\{|+\rangle, |-\rangle\}$. They store their measurement outcome as X and Y respectively which take values in $\{0, 1, +, -\}$.
2. Alice and Bob announce over the classical channel which basis they measured in, which is a function of their measurement outcome, i.e. they can determine their outcome using

$$g(t) = \begin{cases} 0 & t \in \{0, 1\} \\ 1 & t \in \{+, -\} \end{cases} \tag{5.7}$$

3. Alice and Bob map their value X (resp. Y) to a value K_A (resp. K_B) in $\{0, 1, \perp\}$ according to

   (a) If the bases matched,

$$0 \mapsto 0\,,\ 1 \mapsto 1\,,\ + \mapsto 0\,,\ - \mapsto 1\,. \tag{5.8}$$

   (b) If the bases did not match,

$$0 \mapsto \perp\,,\ 1 \mapsto \perp\,,\ + \mapsto \perp\,,\ - \mapsto \perp\ . \tag{5.9}$$

This scheme is in effect the simplest entanglement-based QKD scheme with a few steps cut out because we assume Alice and Bob share $|\Phi^+\rangle$. Let's show that this scheme obtains a bit of secret key half the time.

Proposition 5.3. *The simple scheme results in a* 1*-bit secret key when Alice and Bob measure in the same basis (i.e. with probability 1/2) and otherwise they both share a trivial variable* $\perp$*.*

Proof. The proof will be to verify that Alice and Bob share a 1-bit secret key as defined in Definition 5.1 conditioned on the basis measurement being the same. We break this into 3 steps. First, we get a simple form for $\rho_{XYB_AB_BE}$ where B_A and B_B are the registers that store Alice and Bob's respective measurement bases. This will show all we need to compute is the joint measurement outcomes $\Pr[[]\, x, y]$, so our second step is doing that. Once we have the joint probabilities, we will apply the mapping in Item 3 of the simple protocol to identify that whenever Alice and Bob measured in the same basis, they get a 1-bit secret key to complete the proof.

Let's first focus on the state after Alice and Bob perform their measurements:

$$\rho_{XYE} = \sum_{x,y} p(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \rho_E^{x,y}\,. \tag{5.10}$$

Note that without loss of generality the purification of $\rho_{AB} = |\Phi^+\rangle\langle\Phi^+|_{AB}$ is $\rho_{ABE} = |\Phi^+\rangle\langle\Phi^+|_{AB} \otimes |\psi\rangle\langle\psi|_E$ for some pure state $|\psi\rangle_E$. This gives Eve

as much information as possible, so we focus on this. Now, letting $\{M_x\}_x$ and $\{N_y\}_y$ denote Alice and Bob's POVMS, we have

$$\rho_{XYE} = \sum_{x,y} \mathrm{Tr}_{AB}[M_x \otimes N_y \otimes \mathbb{1}_E \rho_{ABE}] |x\rangle\langle x| \otimes |y\rangle\langle y| \tag{5.11}$$

$$= \sum_{x,y} \mathrm{Tr}_{AB}[M_x \otimes N_y |\Phi^+\rangle\langle\Phi^+|] |x\rangle\langle x| \otimes |y\rangle\langle y| \otimes |\psi\rangle\langle\psi|_E \,. \tag{5.12}$$

Thus, the state after step 2 of the simple scheme is simply

$$\begin{aligned}\rho_{XYB_AB_BE} = \sum_{x,y} \Pr[x,y] \, &|x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \\ &\otimes |g(x)\rangle\langle g(x)|_{B_A} \otimes |g(y)\rangle\langle g(y)|_{B_B} \otimes |\psi\rangle\langle\psi|_E \,,\end{aligned} \tag{5.13}$$

where $\Pr[x,y]$ is the probability of Alice and Bob obtaining joint outcome $x, y \in \{0,1,+,-\}^{\times 2}$.

We now need to compute the joint probabilities $\Pr[x,y]$. One can simply do the calculation by brute force, but we'll use some common tricks that it's good to see multiple times. Let $b_A, b_B \in \{0,1\}$ be the variable specifying the basis choice of Alice and Bob respectively where 0 denotes the computational basis. Then,

$$\begin{aligned}\Pr[x=0,y=0] =& \Pr[b_A=0,b_B=0] \Pr[x=0,y=0|b_A=0,b_B=0] \\ =& \frac{1}{4}\Pr[x=0,y=0|b_A=0,b_B=0] \\ =& \frac{1}{4}\mathrm{Tr}[|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B \Phi^+_{AB}] \,,\end{aligned} \tag{5.14}$$

where the first equality uses that both parties can only get the outcome 0 if they measure in the computational basis, the second equality uses the independence assumptions on Alice and Bob's basis choices and the third equality is Born's rule. Now, to compute the trace quantity in (5.14), we may use the transpose trick. For $i, j \in \{0,1,+,-\}$

$$\mathrm{Tr}[|i\rangle\langle i|_A \otimes |j\rangle\langle j|_B \Phi^+_{AB}] = \mathrm{Tr}[\mathbb{1} \otimes |i\rangle\langle i|_B |j\rangle\langle j|_B^T \Phi^+_{AB}] \tag{5.15}$$

$$= \mathrm{Tr}[|i\rangle\langle i||j\rangle\langle j|\pi] \tag{5.16}$$

$$= \frac{1}{2}|\langle i|j\rangle|^2 \,, \tag{5.17}$$

where the first equality is the transpose trick and the second is that all the states we consider are invariant under transpose in the computational basis as well as the partial trace on the A system. Direct calculations of the overlaps then give us the joint probability table:

Combining Table 5.1 with (5.13), we have

$$\rho_{XYB_AB_BE} \tag{5.18}$$

$$= \frac{1}{8}(|0\rangle\langle 0|^{\otimes 2} + |1\rangle\langle 1|^{\otimes 2}) \otimes |0\rangle\langle 0|^{\otimes 2} \otimes |\psi\rangle\langle\psi|_E \tag{5.19}$$

$$+ \frac{1}{8}(|+\rangle\langle +|^{\otimes 2} + |-\rangle\langle -|^{\otimes 2}) \otimes |1\rangle\langle 1|^{\otimes 2} \otimes |\psi\rangle\langle\psi|_E \tag{5.20}$$

$$+ \text{terms where bases don't match} \,. \tag{5.21}$$

i \j	0	1	+	-
0	1/8	0	1/16	1/16
1	0	1/8	1/16	1/16
+	1/16	1/16	1/8	0
-	1/16	1/16	0	1/8

Table 5.1: Probabilities of the joint outcomes of the simple scheme

It follows that when Alice and Bob do the mapping given in Step 3 of our simple scheme, the joint state is

$$\begin{aligned}
&\rho_{K_A K_B B_A B_B E} \\
=&\frac{1}{8}(|0\rangle\langle 0|^{\otimes 2} + |1\rangle\langle 1|^{\otimes 2}) \otimes |0\rangle\langle 0|^{\otimes 2} \otimes |\psi\rangle\langle\psi|_E \\
&+ \frac{1}{8}(|0\rangle\langle 0|^{\otimes 2} + |1\rangle\langle 1|^{\otimes 2}) \otimes |1\rangle\langle 1|^{\otimes 2} \otimes |\psi\rangle\langle\psi|_E \\
&+ \frac{1}{4}|\perp\rangle\langle\perp|^{\otimes 2} \otimes (|0\rangle\langle 0| \otimes |1\rangle\langle 1| + |1\rangle\langle 1| \otimes |0\rangle\langle 0|) \otimes \psi_E \,.
\end{aligned} \tag{5.22}$$

By defining a variable J which is 0 when the bases match and 1 when the bases don't match, we have the marginal

$$\begin{aligned}
&\rho_{K_A K_B J E} \\
&= \left[\frac{1}{4}(|0\rangle\langle 0|^{\otimes 2} + |1\rangle\langle 1|^{\otimes 2}) \otimes |0\rangle\langle 0|_J + \frac{1}{2}|\perp\rangle\langle\perp| \otimes |1\rangle\langle 1|_J\right] \otimes |\psi\rangle\langle\psi|_E \\
&=\frac{1}{2}\chi_{Z_A Z_B} \otimes |0\rangle\langle 0|_J \otimes |\psi_E\rangle\langle\psi_E| + \frac{1}{2}|\perp\rangle\langle\perp|^{\otimes 2} \otimes |1\rangle\langle 1|_J \otimes \psi_E \,.
\end{aligned} \tag{5.23}$$

Thus, conditioned on the J variable being 0, by Definition 5.1, Alice and Bob share a 1-bit secret key, and conditioned on the J variable being 1, both Alice and Bob hold $\perp$. □

The above shows that if Alice and Bob shared a maximally entangled state, then they could acquire a 1-bit secret key with probability $1/2$.[3] If Alice and Bob shared n copies of the maximally entangled state, they could run the above protocol on each to get roughly $n/2$-bits of secret key. The problem then is for Alice and Bob to establish the entanglement between them so that they can extract secret key. From this viewpoint, entanglement-based quantum key distribution is a protocol that tries to distribute entanglement, does online testing whether entanglement is being distributed, and extracts a secret key when this distribution has succeeded and otherwise aborts the protocol. The rest of this chapter is introducing a class of such protocols and showing the key steps for proving their security.

[3] Technically, they could always get a secret bit by simply measuring in the same basis as they already share a maximally entangled state. We do this inefficient scheme to make generalizing in the next section clearer.

5.2 General Quantum Key Distribution Protocol

We begin with a rather general structure for an entanglement-based device-dependent quantum key distribution protocol. It is device-dependent because we *assume* we know that Alice and Bob implement specified POVMs $\{M_x\}$ and $\{N_y\}$ in their labs.

1. **State Transmission:** Alice prepares a joint quantum state $\rho_{AA'}^{\otimes n}$ and sends the A' systems over to Bob.

   For $i \in [n]$,

   2. **Measurement:** Alice (resp. Bob) measures the A_i (resp. B_i) system with their POVM obtaining outcome x_i (resp. y_i).
   3. **Test Indicator:** Bob draws an (independent of everything else) binary random variable T_i according to distribution p_T and announces it's value. If $T_i = 1$, the round is a 'test round.'
   4. **General Announcements:** Alice (resp. Bob) computes $p_i^A := f_P(x_i, t_i)$ (resp. $p_i^B = g_P(y_i, t_i)$) and announce the values over the classical channel. Let P_i be the register storing both p_i^A and p_i^B.
   5. **Key Map:** Alice computes a function that updates the value of her measurement outcome according to the public announcements and the test round variable, i.e. she uses a function $f_K : \mathcal{X} \times \mathcal{P} \times \mathcal{T} \to \mathcal{Z}$. The resulting Z_i variables are Alice's 'raw key.'

6. **Parameter Estimation:** Let $f_{PE} : \mathcal{P}^n \times \mathcal{T}^n \to \{\checkmark, \times\}$. Bob computes $f(p_1^n, t_1^n)$ and announces the result stored in register F_{pe}. If $F_{pe} = \times$, the protocol is aborted, Alice and Bob set $Z_i = \widehat{Z}_i = \perp$ for all $i \in [n]$, and $K_A = K_B = \perp$. Otherwise, the protocol continues.
7. **Error Correction & Detection:**

   (a) Alice and Bob implement an error correcting code such that Bob outputs $\hat{Z}^n$, his guess of Alice's raw key, Z^n.

   (b) Alice draws a hash function h_{ec} from a 2-universal function family $(\mathcal{H}_{ec}, p_{ec})$ and sends h_{ec} and $h_{ec}(Z^n)$ to Bob. If $h_{ec}(Z^n) = h_{ec}(\hat{Z}^n)$, Bob announces $F_{ec} = \checkmark$. Otherwise, Bob announces $F_{ec} = \times$, the protocol is aborted, Alice and Bob set $Z_i = \widehat{Z}_i = \perp$ for all $i \in [n]$, and $K_A = K_B = \perp$.

8. **Privacy Amplification:** Alice draws a hash function h_{pa} from a 2-universal function family $(\mathcal{H}_{pa}, p_{pa})$. She sets $K_A = h_{pa}(Z^n)$ and sends h_{pa} to Bob. Bob sets $K_B = h_{pa}(\hat{Z}^n)$.

We make some remarks. First, note that test indicator, parameter estimation, error correction & detection, and privacy amplification are the steps that did not arise in the simple scheme in the previous section. They arise for the following reasons:

- Parameter estimation (which will use the testing rounds) will be what allows us to 'check' if Alice and Bob share an entangled state, so that we can do (something like) the simple scheme.
- Error correction & detection are to fix that Alice and Bob may not be perfectly correlated if they share noisy entanglement, and thus need their measurement outcomes to be re-correlated.

- Privacy amplification is to force Eve's system to be approximately independent of the key as she may be partially correlated if Alice and Bob share noisy entanglement and they announce information about their outcomes.

  Next, we make some remarks on generality of our description.

- In State Transmission, sometimes one can instead assume the joint quantum state is prepared by Eve. We assume Alice prepares the joint state as sometimes that assumption is necessary and that is generally how this is done in practice.

- In Measurement, we assume Alice and Bob each apply a single POVM. This may be confusing if one thinks in terms of choosing a basis and then measuring in that basis, but these views can be unified. For example, in entanglement-based BB84 where Bob with probability $1/2$ measures in the computational basis and with probability $1/2$ measures in the Hadamard basis, then his single POVM is

$$\{N_y\}_{y\in\{0,1,2,3\}} = \left\{\frac{1}{2}|0\rangle\langle 0|, \frac{1}{2}|1\rangle\langle 1|, \frac{1}{2}|+\rangle\langle +|, \frac{1}{2}|-\rangle\langle -|\right\} . \quad (5.24)$$

- In principle, the person who decides T_i could change their POVM dependent upon this, but for simplicity, we don't consider such protocols.

- Parameter Estimation has been defined *extremely* generically. In effect, it should be a decision that determines if the state is sufficiently secure against Eve. To explain what 'sufficiently secure' means in any useful detail, we will have to address the other steps first, which is why we have left it in such generality.

The rest of the chapter is formally analyzing the above steps to understand how a 'key rate' arises. In effect, we will work backwards through the steps of the protocol as this will make clear why we analyze each step of the protocol in the manner we do.

5.3 Security and Completeness

To motivate privacy amplification, we will first need to formally define the security and completeness of a QKD protocol.

We begin with defining security. A QKD protocol should either output a secret key or it should abort. It is impossible for it to output a secret key on arbitrary inputs, so it needs to abort on some inputs (with sufficiently high probability). Thus, we need to be able to formalize if the protocol has aborted or not. To that end, we introduce the notion of conditioning on classical events.

Definition 5.4. *Consider quantum-classical state*

$$\rho_{ABXY} = \sum_{x,y} p(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \rho_{AB}^{(x,y)} .$$

Identify $\emptyset \neq \Omega \subset \mathcal{X} \times \mathcal{Y}$ *as an 'event.' The probability of the event with respect to the state is* $\Pr_\rho[\Omega] := \sum_{(x,y)\in\Omega} p(x,y)$*. The state conditioned on the event* Ω *is*

$$\rho_{ABXY}^{|\Omega} = \frac{1}{\Pr_\rho[\Omega]} \sum_{(x,y)\in\Omega} p(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \rho_{AB}^{(x,y)} . \tag{5.25}$$

We also can define the state joint with the event, which is sub-normalized:

$$\rho_{ABXY\wedge\Omega} = \sum_{(x,y)\in\Omega} p(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \rho_{AB}^{(x,y)} . \tag{5.26}$$

We now use the definition of conditioning on an event to talk about conditioning on the QKD protocol not aborting on an input. Let $\rho_{A^nB^nE}$ be the state *after* **State Transmission**. Let

$$\mathcal{E}^{QKD} : A^n \otimes B^n \to Z^n \otimes X^n \otimes Y^n \otimes T^n \otimes P^{\otimes n} \otimes F_{pe} \otimes L_{ec} \otimes F_{ec} , \tag{5.27}$$

be the CPTP map describing the QKD protocol from Items 2 through 7 of our general protocol, where P_i groups Alice and Bob's respective public announcements of round i and L_{ec} denotes all the information publicly announced during error correction and verification. Then we may denote the state after error correction by

$$\rho_{Z^nX^nY^nT^nP^nF_{pe}L_{\mathrm{EC}}F_{ec}E} := (\mathcal{E}^{QKD} \otimes \mathrm{id}_E)(\rho_{A^nB^nE}) . \tag{5.28}$$

The state has 'passed' parameter estimation and error correction and error verification if $F_{pe} = F_{ec} = \checkmark$. Thus the above formalism allows us to talk about the state conditioned on passing, $\rho^{|\mathrm{Pass}}$. By definition of the protocol, if Alice and Bob have not aborted, then they will perform privacy amplification, and so the state $\rho^{|\mathrm{Pass}}_{K_AK_BE''}$ defined by Alice and Bob applying privacy amplification and letting E'' be Eve's system and all classical information available to her is well-defined. With this, we can define a QKD protocol being secure.

Definition 5.5. *Let* $\varepsilon \in (0,1)$*. A QKD protocol is* ε*-secure on* $\rho_{A^nB^nE}$ *if*

$$\Pr_\rho[Pass] \frac{1}{2} \left\| \rho^{|Pass}_{K_AK_BE''} - \chi_{K_AK_B} \otimes \rho^{|Pass}_{E''} \right\|_1 \leq \varepsilon . \tag{5.29}$$

A QKD protocol is ε*-secure if the QKD protocol is* ε*-secure on all* $\rho_{A^nB^nE}$ *that can arise according to the assumptions on* ***State Transmission****.*

First, using the definition of secret key (Definition 5.1), the above definition shows either the probability of pass has to be small or the key has to

be close to a secret key in trace distance. Thus, a ε-secure QKD protocol is a protocol that approximately makes secret keys so long as it doesn't abort. Second, note that the left hand side of (5.29) is scaled by the probability of not aborting. Thus, a protocol that *always* aborts is perfectly secure. It's also useless. For this reason, we also need completeness, which says that the QKD protocol when implemented as we expect (i.e. there could be noise, but Eve isn't messing with it), does not abort except with probability $\varepsilon_C \in (0,1)$.

Definition 5.6. *Let $\varepsilon_C \in (0,1)$. An n-round QKD protocol is ε_C-complete if for an assumed honest noise model between Alice and Bob, the QKD protocol would abort with probability at most ε_C.*

You would normally want a QKD protocol that is ε-secure and ε_C-complete where both ε is very small and ε_C is acceptably small. For example, it might be natural to choose $\varepsilon = 10^{-10}$ and $\varepsilon_C = 0.25$. This means the key is 10^{-10}-secure and when no one tampers with your devices, you generate a key $3/4$ of the time.

Finally, we state the following lemma, which shows sufficient conditions for security in terms of Alice and Bob's output keys being the same (ε_{ec}-correct) and Alice's key being independent of the information Eve holds (ε_{pa}-secret).

Lemma 5.7. *Let $\varepsilon_{ec}, \varepsilon_{pa} \in [0,1)$. If for every state $\rho_{A^n B^n E}$ that can arise according to the assumptions on **State Transmission** the following conditions hold:*

$$\Pr_\rho[K_A \neq K_B \cap Pass] \leq \varepsilon_{ec} \tag{5.30}$$

$$\Pr_\rho[Pass] \frac{1}{2} \| \rho^{|Pass}_{K_A E''} - \pi_{K_A} \otimes \rho^{|Pass}_{E''} \|_1 \leq \varepsilon_{pa}\,, \tag{5.31}$$

then the QKD protocol is $(\varepsilon_{ec} + \varepsilon_{pa})$-secure.

You will prove this in the homework.

5.4 Privacy Amplification

Privacy amplification identifies what criteria of the raw key guarantees a way to bound ε_{pa} of Lemma 5.7. Note however that the privacy criteria in (5.31) is just the criteria for randomness extraction from last chapter given in (4.90). Moreover, note that

1. we proved randomness extraction works even when the value of the seed S is announced as is captured by (4.92), and

2. as each function f_s in the extractor is deterministic, if $Z^n = \hat{Z}^n$, then $f_s(Z^n) = f_s(\hat{Z}^n)$.

Thus, privacy amplification just involves Alice doing randomness extraction on her raw key using 2-universal hashing and then announcing her choice of the seed s so that Bob can apply f_s to $\hat{Z}$. In other words, privacy amplification follows from the randomness extraction protocol we addressed earlier, and so we obtain it is controlled by the smooth min-entropy in the same manner as the following states.

Lemma 5.8. *For (5.31) to hold for input $\rho_{A^nB^nE}$, it suffices for Alice to use a family of 2-universal hash functions to hash to ℓ-bits where*

$$\ell \leq H_{\min}^{\varepsilon_{pa}/4}(Z^n|E'')_{\rho^{|F_{pe}=\checkmark}_{Z^nE''\wedge F_{ec}=\checkmark}} - 2\log(1/\varepsilon_{pa}) \,. \tag{5.32}$$

Proof. Note $\Pr[\text{Pass}] = \Pr\left[F_{pe} = \checkmark\right] \Pr\left[F_{ec} = \checkmark | F_{pe} = \checkmark\right]$. Thus, using the definition of a state joint with an event (5.26),

$$\Pr_\rho[\text{Pass}] \frac{1}{2}\|\rho^{|\text{Pass}}_{K_AE''} - \pi_{K_A} \otimes \rho^{|\text{Pass}}_{E''}\|_1 \tag{5.33}$$

$$= \Pr_\rho\left[F_{pe} = \checkmark\right] \frac{1}{2}\|\rho^{|F_{pe}=\checkmark}_{K_AE''\wedge F_{ec}=\checkmark} - \pi_{K_A} \otimes \rho^{|F_{pe}=\checkmark}_{E''\wedge F_{ec}=\checkmark}\|_1 \,. \tag{5.34}$$

We then want to apply the leftover hashing lemma to bound the trace term above by ε_{PA} as $\Pr_\rho\left[F_{pe} = \checkmark\right]$ could in principle be one. Note however the state is subnormalized. This is fine as such a variant exists (Proposition 9 of [4]) and doing the same argument recovers (4.114), which we then apply. □

[4] Marco Tomamichel and Anthony Leverrier. A largely self-contained and complete security proof for quantum key distribution. *Quantum*, 1:14, July 2017. ISSN 2521-327X. DOI: 10.22331/q-2017-07-14-14. URL http://dx.doi.org/10.22331/q-2017-07-14-14

5.5 Error Correction and Error Verification

By the previous corollary, we see what we need to bound is the smooth min-entropy of Alice's raw key conditioned on Eve's side-information. By Lemma 5.7, we need to guarantee the error correction scheme guarantees (5.30). Here we first explain this latter point. Then we explain how to remove this joint event of error correction passing in the smooth min-entropy estimate, which will make discussing parameter estimation easier.

Here we show the error verification step allows us to guarantee with high probability we don't pass except if the error correction worked. This is in effect directly from the property of the two-universal hashing.

Lemma 5.9. *Let Z^n be Alice's raw key. Let $\hat{Z}^n$ be Bob's guess of Alice's raw key after error correction. Let $(\mathcal{H}, p_H)$ be the two-universal family of hash functions used for error verification where $h : \mathcal{Z}^n \to \{0,1\}^t$ for all $h \in \mathcal{H}$. Let F^{ec} be the binary random variable where it takes the value ✓ if error verification passes. Then the probability that $Z^n \neq \hat{Z}^n$ and error verification passes is less than 2^{-t}:*

$$\Pr[[] Z^n \neq \hat{Z}^n \wedge F^{ec} = \checkmark] \leq 2^{-t} \tag{5.35}$$

Proof. We have the sequence of (in)equalities:

$$\begin{aligned}
&\Pr[[]\, Z^n \neq \hat{Z}^n \wedge F^{\text{ec}} = \checkmark] && (5.36)\\
=&\Pr[[]\, Z^n \neq \hat{Z}^n \wedge \mathcal{H}_{\text{ec}}(Z^n) = \mathcal{H}_{\text{ec}}(\hat{Z}^n)] && (5.37)\\
=&\Pr[[]\, Z^n \neq \hat{Z}^n] \cdot \Pr[[]\, \mathcal{H}_{\text{ec}}(Z^n) = \mathcal{H}_{\text{ec}}(\hat{Z}^n) | Z^n \neq \hat{Z}^n], && (5.38)\\
\leq&\Pr[[]\, \mathcal{H}_{\text{ec}}(Z^n) = \mathcal{H}_{\text{ec}}(\hat{Z}^n) | Z^n \neq \hat{Z}^n], && (5.39)\\
\leq&|\{0,1\}^t|^{-1} = 2^{-t}\,, && (5.40)
\end{aligned}$$

where the first equality is the definition of when error verification passes, the first inequality is a result of $\Pr[[]\, X^n \neq \hat{X}^n] \leq 1$, and the final inequality is the defining property of a 2-universal hash family. □

We thus know how to guarantee (5.30). What remains are tools for removing the error correction register and conditioning on passing when bounding the smooth min-entropy. This will use three results, two of which can be useful when working with entropic estimates and the last of which is a fundamental point about error correction.

The first result shows that when working with smooth min-entropy of a state with a joint event, one can remove the joint event so long as the smoothing is sufficiently small relative to the probability of the event.

Proposition 5.10. *(Lemma 10 of Ref. [5]) Let ρ_{ABXY} be a quantum state that is classical on X and Y. Let $\emptyset \neq \Omega \subset \mathcal{X} \times \mathcal{Y}$ such that $\Pr_\rho[\Omega] > 0$. For all $\varepsilon \in [0, \sqrt{\Pr_\rho[\Omega]})$,*

$$H_{\min}^{\varepsilon}(AX|BY)_{\rho_{\wedge\Omega}} \geq H_{\min}^{\varepsilon}(AX|BY)_{\rho}\,. \tag{5.41}$$

[5] Marco Tomamichel and Anthony Leverrier. A largely self-contained and complete security proof for quantum key distribution. *Quantum*, 1:14, July 2017. ISSN 2521-327X. DOI: 10.22331/q-2017-07-14-14. URL http://dx.doi.org/10.22331/q-2017-07-14-14

The second is a chain rule that operationally says how much information can be stored in a classical register.

Proposition 5.11. *(Lemma 6.18 of Ref. [6]) Let ρ_{ABY} be a quantum state that is classical on Y. Let $\varepsilon \in [0,1)$. Then,*

$$H_{\min}^{\varepsilon}(A|YB)_{\rho} \geq H_{\min}^{\varepsilon}(A|B)_{\rho} - \log(|Y|)\,. \tag{5.42}$$

[6] Marco Tomamichel. *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016. ISBN 9783319218915. DOI: 10.1007/978-3-319-21891-5. URL http://dx.doi.org/10.1007/978-3-319-21891-5

Combining these two results, we obtain the following.

Proposition 5.12. *(Variant of Eq. 4.39 of [7]) Let $\varepsilon \in (0,1)$, ρ be defined as in (5.28) and satisfy $\Pr_\rho\left[F_{ec} = \checkmark | F_{pe} = \checkmark\right] > \varepsilon^2$. Let E'' be all of the registers available to Eve and define E' such that $E'' = E' L_{ec}$. Then,*

$$H_{\min}^{\varepsilon}(Z|E'')_{\rho^{|F_{pe}=\checkmark}_{ZE''\wedge F_{ec}=\checkmark}} \geq H_{\min}^{\varepsilon}(Z|E')_{\rho^{|F_{pe}=\checkmark}} - \log(|L_{ec}|)\,. \tag{5.43}$$

[7] Ernest Y. Z. Tan. Prospects for device-independent quantum key distribution, 2024. URL https://arxiv.org/abs/2111.11769

Proof. As $\varepsilon^2 < \Pr[[]\, F_{ec} = 1 | F_{pe} = 1] = \operatorname{Tr}[\rho^{|F_{pe}=\checkmark}_{ZE''\wedge F_{ec}=\checkmark}]$, taking the square root tells us we may apply Proposition 5.10,

$$H_{\min}^{\varepsilon}(Z|E'')_{\rho^{|F_{pe}=\checkmark}_{ZE''\wedge F_{ec}=\checkmark}} \geq H_{\min}^{\varepsilon}(Z|E'')_{\rho^{|F_{pe}=\checkmark}}\,. \tag{5.44}$$

By removing the error correction transcript from E', i.e. $E' = E''L_{\mathrm{EC}}$, and using

$$H_{\min}^{\varepsilon}(Z|E'')_{\rho^{|F_{pe}=\checkmark}} = H_{\min}^{\varepsilon}(Z|E'L_{ec})_{\rho^{|F_{pe}=\checkmark}} \tag{5.45}$$

$$\geq H_{\min}^{\varepsilon}(Z|E')_{\rho^{|F_{pe}=\checkmark}} - \log(|L_{ec}|)\,. \tag{5.46}$$

Combining the steps completes the proof. □

A natural concern is the size of $|L_{ec}|$. The following bounds this in terms of smooth max entropy.

Proposition 5.13. *Let p_{ZY} be a joint distribution and $\varepsilon > 0$. There exists an error correction and error verification scheme for guessing the value of Z such that the event $[\hat{Z} \neq Z$ and it goes undetected$]$ happens with probability at most ε satisfying*

$$\log(|L_{ec}|) \lesssim H_{\max}^{\varepsilon}(Z|Y)_p + \log(2/\varepsilon)\,. \tag{5.47}$$

In the above, for intuition, one can identify $H_{\max}^{\varepsilon}(Z|Y)_p$ as the number of bits announced in the error correcting code and the $\log(2/\varepsilon)$ as pertaining to the needed size of the error verification hash.

5.6 Parameter Estimation Part I

We now turn to what arguably is the key idea in quantum key distribution: detecting if Eve is interacting with the quantum state during state transmission by testing random rounds of the signal. This is crucial. If there were an input state $\rho_{A^nB^nE}$ such that $\rho_{Z^nE''}$ had less min-entropy than is sufficient for the length you hash to in privacy amplification, you would generate an insecure key. Thus, on such states you want to abort with high probability. On the other hand, if Alice and Bob actually shared many maximally entangled states, then our motivating simple scheme shows Alice and Bob could extract key. This intuition can be extended via our formalism as we now see. Consider

$$\begin{aligned}\rho_{Z^nT^nP^nEF_{pe}} = \Pr[[]\, F_{pe} = \checkmark]\rho_{Z^nT^nP^nE}^{|F_{pe}=\checkmark} \otimes |\checkmark\rangle\langle\checkmark|_{F_{pe}} \\ + \Pr[[]\, F_{pe} = \times]\rho_{Z^nT^nP^nE}^{|F_{pe}=\times} \otimes |\times\rangle\langle\times|_{F_{pe}}\,.\end{aligned} \tag{5.48}$$

Keeping in mind we need to either abort or have sufficiently high min-entropy, we want to build a function $f_{PE} : P^n \times T^n \to \{\checkmark, \times\}$ such that for all possible inputs $\rho_{A^nB^nE}$ either

1. $H_{\min}^{\varepsilon}(Z^n|E')_{\rho^{|F_{pe}=\checkmark}}$ is high enough to result in ℓ we are satisfied with, or

2. $\Pr_{\rho}\left[F_{pe} = \checkmark\right] = \Pr[\{p^n, t^n : f_{PE}(p^n, t^n) = \checkmark\}]$ is *very* small.

This is a balancing act. The more sequences on which the function f_{pe} outputs ✓, the lower the worst-case min-entropy of a state conditioned on accepting could be. On the other hand, the less sequences you accept, the more likely your protocol is to abort— perhaps even on the honest implementation! We will require an account of how to pick the sequences f_{PE} should accept. To make this tractable, we will address it with the simplifying assumption we make in the next section.

5.7 A Simplifying Assumption: Independent and Identically Distributed States

Note we have now at least introduced all of the steps in our general quantum key distribution protocol that aren't the same on each round. At this point, we make a further assumption, which is the independent and identically distributed (i.i.d.) assumption. Namely, we will assume that that the state from state transmission is actually of the form

$$\rho_{A^n B^n \widetilde{E}} = \rho_{ABE}^{\otimes n} \,. \tag{5.49}$$

Let us explain what this assumption would mean *physically*. First, Alice's device would have to prepare n copies of the same state, $\rho_{AA'}^{\otimes n}$. Then, when the A' systems go to Bob, Eve would have to do the same channel $\mathcal{E}_{A' \to BE}$ each time. This is clearly a strong assumption. This assumption also is **not necessary** thanks to the theory of "i.i.d. reductions," which we do not discuss in this class (you could see [8] for an introduction). We make this simplifying assumption for ease of presentation and conceptual clarity. The reason this simplifying assumption is so powerful is in effect the (weak) law of large numbers.

[8] Rotem Arnon-Friedman. *Device-Independent Quantum Information Processing: A Simplified Analysis.* Springer International Publishing, 2020. ISBN 9783030602314. DOI: 10.1007/978-3-030-60231-4. URL http://dx.doi.org/10.1007/978-3-030-60231-4

Proposition 5.14. *Let $n \in \mathbb{N}$. Let $X_1, X_2, ..., X_n$ be i.i.d. random variables and μ be the mean of X_1. Define the sample mean random variable $\overline{X}_n \coloneqq \frac{1}{n}\sum_i X_i$. Then for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \Pr[[] \, |\overline{X}_n - \mu| > \varepsilon] = 0 \,. \tag{5.50}$$

All of the information-theoretic results in this section are law of large number-type results in some sense.

5.7.1 Learning a Distribution from i.i.d. Sampling

We consider the problem of learning a distribution by sampling from it in an i.i.d. fashion. This will be how we address parameter estimation.

The problem is as follows. Let p_X denote a probability mass function. Imagine there are m time steps. At each time step, you receive a sample $x \in \mathcal{X}$ according to $p_X(x)$. After m times steps, you could construct the

'observed frequency distribution' f_m via

$$f_m(x) \coloneqq \frac{\{\#\text{ of time steps } x \text{ occurred}\}}{n} \,. \tag{5.51}$$

If $\mathcal{X} = \{0,1\}$, then $f_m(0)$ is just the m-round sample mean of X_i being a Bernoulli random variable, so by the LoLN (Proposition 5.14), for all $\varepsilon > 0$, $\lim_{m\to\infty} \Pr[[] \, |f_m(0) - p(0)| > \varepsilon] = 0$. As $f_m(1) = 1 - f_m(0)$, by the same argument, $f_m(1)$ should converge to $1 - p(0) = p(1)$ in probability. That is, the observed frequency distribution should converge to the probability distribution in an entry-wise fashion. The following result generalizes this by bounding the probability that the observed frequency distribution is far from the distribution being sampled from as a function of the size of $\mathcal{X}$ and the number of rounds m. In this sense, it is a quantitative generalization of the law of large numbers.

Proposition 5.15. *Let $X_1, X_2, ..., X_m$ be i.i.d. random variables taking value $x \in \mathcal{X}$ with probability $p_X(x)$. Let $\varepsilon > 0$. Then,*

$$\Pr[[] \, D(f_m \| p_X) > \varepsilon] \leq \exp\left(-m\left[\varepsilon - |\mathcal{X}|\frac{\log(m+1)}{m}\right]\right) \,. \tag{5.52}$$

The above shows that as m goes to infinity, the observed frequency distribution converges to the probability distribution being sampled from with probability 1. What is important for our purposes is that f_m is an estimate of p_X that converges to p_X exponentially fast in the number of samples.

5.7.2 Parameter Estimation Part II

We are now in a good place to address parameter estimation. We first show how this ought to work. Let $\mathcal{M}_{A\to X}$ and $\mathcal{N}_{B\to Y}$ be the quantum channels representing Alice and Bob performing their measurements in Item 2 of the protocol. As Alice and Bob do the same measurements on every round, using our i.i.d. input assumption we have

$$(\mathcal{M}^{\otimes n} \otimes \mathcal{N}^{\otimes n} \otimes \mathrm{id}_{\widetilde{E}})(\rho_{A^n B^n \widetilde{E}}) = [(\mathcal{M} \otimes \mathcal{N})(\rho_{ABE})]^{\otimes n} \,. \tag{5.53}$$

On the 'single-copy' level we have

$$(\mathcal{M} \otimes \mathcal{N})(\rho_{ABE}) = \rho_{XYE} = \sum_{x,y} p(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \rho_E^{x,y} \,. \tag{5.54}$$

The best one can possibly hope for is to learn ρ_{AB} to the detail allowed by the joint distribution $\{p(x,y)\}_{x,y}$ and in fact, we can asymptotically determine this joint distribution by using the previous section.

We decide that when a round i is such that $T_i = 1$, i.e. it is a test round, Alice and Bob announce their values x_i and y_i publicly during general announcements. Then for whatever number of rounds are test rounds, say m, Bob can compute the observed (joint) frequency distribution (5.51), f_m.

Because we let T_i be distributed according to a fixed distribution p_T, as the number of rounds n of the protocol grows, the number of test rounds m will grow. By Proposition 5.15, as m grows, f_m will converge to the joint distribution encoded in ρ_{XY}. Thus, this strategy learns the state better and better and asymptotically determines ρ_{XY}.

For this reason, a tractable and general manner for implementing parameter estimation is to accept distributions around what your QKD protocol should do when not attacked and abort on other distributions. This can be done as follows.

1. Let σ_{AB}^{hon} denote the state we expect to be generated by Alice and the (possibly noisy) quantum channel between Alice and Bob.[9]
2. Define $q_{XY} := (\mathcal{M} \otimes \mathcal{N})(\sigma_{AB}^{\text{hon}})$, i.e. the honest joint distribution, and define the δ-ball of distributions around q_{XY},

$$B_\delta(q_{XY}) := \{\text{prob. dist. } t_{XY} : \frac{1}{2}\|t_{XY} - q_{XY}\|_1 \leq \delta\} . \tag{5.55}$$

3. Bob's decision function is described by the following procedure:
   (a) Convert the the public announcements on testing rounds, which are just joint outcomes (x_i, y_i) to an observed frequency distribution f_m
   (b) If $f_m \in B_\delta(q_{XY})$, then output ✓, otherwise output ×.

[9] Presumably σ_{AB}^{hon} should be such that many copies of it result in reasonable min-entropy. An example of this would be if it were the maximally entangled state from the simple scheme.

Note that parameter estimation is just about a sequence of observations, but we ultimately care about a quantum state. The set of *quantum states* that generate *distributions* that are contained in $B_\delta(q_{XY})$.

$$S_{\text{acc}}^{\text{state}}(\delta) := \{\rho_{ABE} : \frac{1}{2}\|(\mathcal{M} \otimes \mathcal{N})(\rho_{AB}) - q_{XY}\|_1 \leq \delta\} . \tag{5.56}$$

Note that whether or not a state ρ_{AB} is in $S_{\text{acc}}^{\text{state}}(\delta)$, it could result in $f_m \in B_\delta(q_{XY})$. However, Proposition 5.15 tells us that this happens with vanishing probability as n (and thus m grows). This means that asymptotically, we will abort on any state not contained in $S_{\text{acc}}^{\text{state}}$ as the following states.

Proposition 5.16. *A state $\rho_{ABE}^{\otimes n}$ will be accepted asymptotically by parameter estimation if and only if it is in $S_{acc}^{state}(\delta)$. That is,*

$$\lim_{n\to\infty} \Pr_{\rho^{\otimes n}}\left[F_{pe} = \checkmark\right] = \begin{cases} 1 & \rho_{ABE} \in S_{acc}^{state}(\delta) \\ 0 & \textit{otherwise.} \end{cases} \tag{5.57}$$

Proof. Using $D(p\|q) \geq \frac{1}{2\ln(2)}\|p - q\|_1^2$ so that the convergence in Proposi-

tion 5.15 implies convergence in trace distance, we have

$$\lim_{n\to\infty} \Pr_{\rho^{\otimes n}}\left[F_{pe} = \checkmark\right] = \lim_{m\to\infty} \Pr\left[\frac{1}{2}\|f_m - q_{XY}\|_1 \leq \delta\right] \tag{5.58}$$

$$= \begin{cases} 1 & (\mathcal{M}\otimes\mathcal{N})(\rho_{AB}) \in B_\delta(q_{XY}) \\ 0 & \text{otherwise} \end{cases} \tag{5.59}$$

$$= \begin{cases} 1 & \rho_{ABE} \in S_{\text{acc}}^{\text{state}}(\delta) \\ 0 & \text{otherwise.} \end{cases} \tag{5.60}$$

□

Before moving on, we briefly stress the i.i.d. assumption is not necessary. It gives us the intuition. In general, it is natural to define the set of observations you accept as those resulting in an observed frequency distribution contained in $B_\delta(q_{XY})$. Even without the i.i.d. assumption on state transmission, one can prove that the entropy won't be "too bad" once one samples enough. This is in effect what i.i.d. reductions (with testing) are guaranteeing.

5.7.3 *Asymptotic Equipartition Property*

We have already seen how the i.i.d. assumption and the LoLN helps us. Here we see one more case where it helps us. Recall the surprisal $S(p_X) = \log\frac{1}{p_X}$ from Chapter 4. We saw that the entropy is the expectation of the surprisal. The asymptotic equipartition property is an immediate corollary of this observation and the LoLN.

Proposition 5.17. *Let $X_1, X_2, ..., X_n$ be i.i.d. random variables drawn according to distribution p_X. Then*

$$-\frac{1}{n}\log p(X_1, X_2, ..., X_n) \to H(X) \tag{5.61}$$

Proof. As the random variables are independent and the logarithm satisfies $\log(ab) = \log(a) + \log(b)$,

$$-\frac{1}{n}\log p(X_1, X_2, ..., X_n) = -\frac{1}{n}\sum_{i\in[n]} \log(p_{X_i}) \tag{5.62}$$

$$\to -\mathbb{E}_{p_X}\log(p_X) \quad \text{in probability} \tag{5.63}$$

$$= H(X)\,, \tag{5.64}$$

where the convergence in probability is the LoLN. □

The reason this matters is it tells us that, for sufficiently large n, the sequences we see from drawing from p_X in an i.i.d. fashion satisfy $-\frac{1}{n}\log(p(X_1, X_2, ..., X_n)) \approx H(X)_p$. This lets us throw out all of the other sequences of $p_X^{\otimes n}$ without changing much about the total distribution.

We can say something similar for smooth entropies. The smooth min- and max- entropies allow us to deform the state a little. The following says this ability to deform the state as n grows will allow us to recover the von Neumann entropy. This is thus sometimes called the 'fully quantum asymptotic equipartition property.'

Proposition 5.18. *Let ρ_{AB} be a quantum state. Then*

$$\lim_{n\to\infty}\frac{1}{n}H^{\varepsilon}_{\min}(A^n|B^n)_{\rho^{\otimes n}} = H(A|B)_\rho = \lim_{n\to\infty}\frac{1}{n}H^{\varepsilon}_{\max}(A^n|B^n)_{\rho^{\otimes n}}\ . \quad (5.65)$$

5.8 Putting Things Together: Deriving An Asymptotic Key Rate

Finally, we determine an achievable asymptotic rate, i.e. a number of bits of key per signal transmission we can extract as n goes to infinity. This will use the CPTP map

$$\mathcal{G} : A \otimes B \to Z \otimes X \otimes Y \otimes T \otimes P\ ,$$

which is the 'per round' QKD map, i.e. the CPTP map that captures Steps 2-5 of our generic protocol for a single $i \in [n]$.

Theorem 5.19. *For a QKD protocol as explained in this chapter and $\varepsilon \in (0,1)$, the asymptotic rate is given by*

$$\min_{\rho\in S^{state}_{acc}(\delta)}\left[H(Z|E)_{(\mathcal{G}\otimes id_E)(\rho_{ABE})} - H(Z|Y)_{(\mathcal{G}\otimes id_E)(\rho_{ABE})}\right] \quad (5.66)$$

Proof. Fix $\varepsilon_{ec}, \varepsilon_{pa} \in (0,1)$ which we use to decide the $\varepsilon \in (0,1)$ by Lemma 5.7. Next, by Lemma 5.9, we can control (5.30) by 2^{-t} where t is the number of output bits in our error verification hash. Thus, we can pick sufficiently large t so that $2^{-t} \leq \varepsilon_{ec}$ and we guarantee (5.30). Therefore, what we need to focus on is satisfying (5.31).

Next, note that (5.31) is scaled by

$$\Pr_\rho[\text{Pass}] = \Pr_\rho\left[F_{ec} = \checkmark | F_{pe} = \checkmark\right]\Pr_\rho\left[F_{pe} = \checkmark\right]\ ,$$

which we can use this to our advantage. If $\Pr_\rho\left[F_{pe} = \checkmark\right] \leq \varepsilon_{pa}$, then the security condition is trivially satisfied. By Proposition 5.16, there is always sufficiently large n such that $\Pr_{\rho^{\otimes n}}\left[F_{pe} = \checkmark\right] < \varepsilon_{pa}$ unless $\rho \in S^{\text{state}}_{\text{acc}}(\delta)$. So we can focus on states in $S^{\text{state}}_{\text{acc}}(\delta)$. Similarly, if $\Pr_\rho\left[F_{ec} = \checkmark \mid F_{pe} = \checkmark\right] \leq (\varepsilon_{pa}/4)^2$, then (5.31) is trivially satisfied, so we need to only considers states such that this conditional probability is strictly larger, i.e. states such that

$$\Pr_\rho\left[F_{ec} = \checkmark | F_{pe} = \checkmark\right] > (\varepsilon_{PA}/4)^2\ . \quad (5.67)$$

Fix a state that satisfies these criteria. As we wish to appeal to Lemma 5.8, we need to bound the smooth min-entropy of this state conditioned on

parameter estimation passing and joint with error correction passing. So we do this:

$$H_{\min}^{\varepsilon_{pa}/4}(Z^n|E'')_{(\rho^{|F_{pe}=\checkmark})\wedge F_{ec}=\checkmark} \tag{5.68}$$

$$\geq H_{\min}^{\varepsilon_{pa}/4}(Z^n|E')_{\rho^{|F_{pe}=\checkmark}} - \log(|L_{ec}|) \tag{5.69}$$

$$\approx H_{\min}^{\varepsilon_{pa}/4}(Z^n|E^n)_{[(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})]^{\otimes n}} - \log(|L_{ec}|) \tag{5.70}$$

$$\gtrsim H_{\min}^{\varepsilon_{pa}/4}(Z^n|E^n)_{(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})^{\otimes n}} - H_{\max}^{\varepsilon}(Z^n|Y^n)_{(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})^{\otimes n}} \tag{5.71}$$

$$= nH(Z|E)_{(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})} - nH(Z|Y)_{(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})} - o(n)\,, \tag{5.72}$$

where the first inequality is Proposition 5.12 which we may apply as the state satisfies (5.67), the approximation uses that technically $\rho^{|F_{pe}=\checkmark} = [(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})]^{\otimes n}$ only holds asymptotically, the second inequality is assuming we use the error correcting code in Proposition 5.13, and the final equality is Proposition 5.18. Note that the $o(n)$ term depends on the smoothing parameter but not the state.

Now, by Lemma 5.8, to satisfy (5.31) on these states it suffices for

$$\ell(n) \leq n \min_{\rho\in S_{\mathrm{acc}}^{\mathrm{state}}(\delta)} \left[H(Z|E)_{(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})} - H(Z|Y)_{(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})}\right] - o(n)\,,$$

where we used minimizing the term over all states in $S_{\mathrm{acc}}^{\mathrm{state}}(\delta)$ rather than only those could only lead to a lower value. Finally, we divide by n and take the limit.

$$\lim_{n\to\infty} \ell(n)/n \leq \min_{\rho\in S_{\mathrm{acc}}^{\mathrm{state}}(\delta)} \left[H(Z|E)_{(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})} - H(Z|Y)_{(\mathcal{G}\otimes \mathrm{id}_E)(\rho_{ABE})}\right]\,,$$

As this worked for $\varepsilon > 0$, we can let $\varepsilon \to 0$ to obtain vanishing error and the same key length. □

5.9 BONUS: Lifting Prepare-and-Measure to Entanglement-Based

So far we have shown how to prove security for entanglement-based protocols. We can prove security of prepare-and-measure protocols by instead proving security of entanglement-based protocols as we briefly explain. This is known as the "source replacement scheme."

Consider a prepare-and-measure protocol where Alice sends state $|\psi_x\rangle$ with probability p_x. Define the entangled state $|\zeta\rangle_{AA'} = \sum_{x\in\mathcal{X}} \sqrt{p_x}\,|x\rangle\,|\psi_x\rangle$. Then Alice holds the A system and sends the A' system to Bob over the channel resulting in state

$$\rho_{AB} = (\mathrm{id}_A \otimes \mathcal{E}_{A'\to B})(|\zeta\rangle\langle\zeta|)\,. \tag{5.73}$$

If Alice applies the projective measurement $\{|x\rangle\langle x|_A\}_x$, her outcome is x with probability p_x and Bob's conditional state is $\rho_B^x = \mathcal{E}(|\psi_x\rangle\langle\psi_x|)$. This shows that once Alice measures her A register in the computational

basis, she has implemented the prepare and measure QKD protocol. Indeed, proving the security of the QKD protocol where every round Alice prepares $|\zeta\rangle_{AA'}$, sends the A' system to Bob, and measures her system A with $\{|x\rangle\langle x|\}_{x\in\mathcal{X}}$, then one is proving the security of the prepare-and-measure protocol, but are analyzing the security of an entanglement-based QKD protocol.

There is however one nuance. Note the source-replaced entanglement-based QKD protocol implies the only possible states after State Transmission will be $\rho_{A^nB^nE}$ such that $\rho_{A^n} = \zeta_A^{\otimes n}$. This structure generally needs to be enforced such as in Theorem 5.19. This is because the separable state

$$\rho_{AB} = \sum_x |x\rangle\langle x| \otimes |\psi_x\rangle\langle\psi_x| \tag{5.74}$$

can re-produce the same joint statistics between Alice and Bob, but allows Eve to hold a copy of x and thus obtain the key.

6
Source coding and the convex-split lemma

In this lecture, we study different types of *source-coding* protocols, along with a useful tool in quantum information theory known as the *convex-split lemma* which is helpful in designing such protocols. We begin by introducing some classical scenarios.

6.1 Classical setting

Let Alice receive an input random variable $X \in \{0,1\}^n$ which she intends to send to Bob. Alice sends a message M to Bob and Bob outputs $\hat{X}$. The aim of source coding is to send a small message M while keeping the probability of error ($\varepsilon \geq 0$) small, that is,

$$\Pr\left[X \neq \hat{X}\right] \leq \varepsilon.$$

If Alice and Bob share a random variable S before Alice's input arrives, the protocol becomes a *shared randomness-assisted* protocol.

To ensure that Alice is communicating her input (and not something else), we introduce a Referee R, which keeps a copy of X. The correctness of the protocol can then be rewritten as,

$$XX \approx_\varepsilon X\hat{X}.$$

More generally R and X are two correlated random variables and R need not equal X. The correctness of the protocol can then be rewritten as,

$$RX \approx_\varepsilon R\hat{X}.$$

As a further generalization (Figure 6.1), Alice and Bob could have their own random variables Y and Z, respectively, jointly correlated with RX. The correctness of the protocol can then be rewritten as,

$$RYZX \approx_\varepsilon RYZ\hat{X}.$$

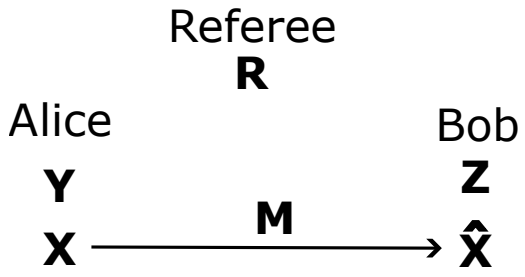

Figure 6.1: Schematics of classical protocol.

6.2 Quantum source coding

Quantum source coding is the quantum analogue of the classical case. The most general scenario is known as *quantum state redistribution.* At the beginning Alice (AC), Bob (B), and Referee (R) share a global pure state

$$\psi_{RABC} = |\psi_{RABC}\rangle \langle\psi_{RABC}| .$$

Alice wants to send the register C to Bob by transmitting message M (after some local operations at her end). If Alice and Bob start with some shared state, then we call this protocol an *entanglement-assisted* protocol. For example, Alice and Bob could share several EPR pairs in their registers $\hat{A}$ and $\hat{B}$. After transmission of the message M and Bob performing some local operations, the global state becomes $\varphi_{RAB\tilde{C}}$[1]. The goal of the protocol is to minimize the size (length) of the message M while keeping the error $\varepsilon \geq 0$ small, that is,

$$\psi_{RABC} \approx_\varepsilon \varphi_{RAB\tilde{C}}.$$

The symbol $\approx_\varepsilon$ between two states represents that the two states have the purified distance at most ε. Recall,

[1] Note that the state $\varphi_{RAB\tilde{C}}$ is not necessarily a pure state.

Definition 6.1 (Fidelity and purified distance)**.** *For quantum states ρ, σ, the fidelity between them is defined as,*

$$\mathrm{F}(\rho,\sigma) \stackrel{def}{=} \|\sqrt{\rho}\sqrt{\sigma}\|_1^2.$$

The purified distance between them is defined as

$$\mathrm{P}(\rho,\sigma) \stackrel{def}{=} \sqrt{1-\mathrm{F}(\rho,\sigma)}.$$

Definition 6.2 (ε-ball)**.** *Let ρ be a state.*

$$\mathcal{B}^\varepsilon(\rho) := \{\rho' \mid \rho' \text{ is a state and } \mathrm{P}(\rho,\rho') \leq \varepsilon\}.$$

We say $\rho' \approx_\varepsilon \rho$ *if* $\rho' \in \mathcal{B}^\varepsilon(\rho)$.

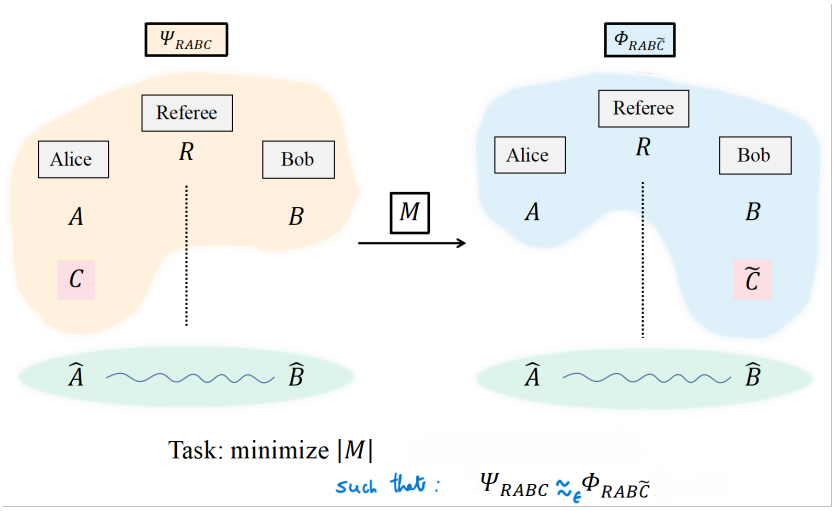

Figure 6.2: Quantum state redistribution.

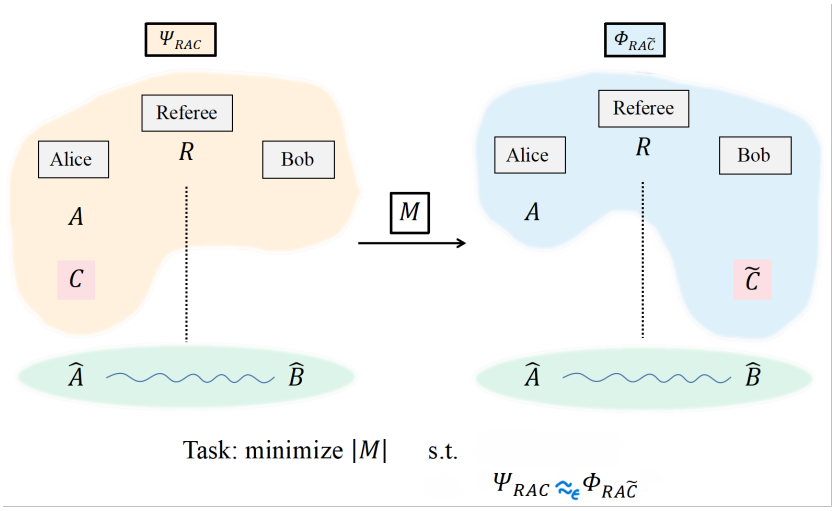

Figure 6.3: Quantum state splitting. A special case of state redistribution when the register B is missing.

An illustrative diagram is shown in Figure 6.2. We consider the following three types of variants.

1. *Quantum state splitting*: No register B at Bob's side (Figure 6.3).
2. *Quantum state merging*: No register A at Alice's side (Figure 6.4).
3. *Quantum state transfer*: Both A and B are absent (Figure 6.5).

Table 6.1 summarizes the presence or absence of the registers and goals.

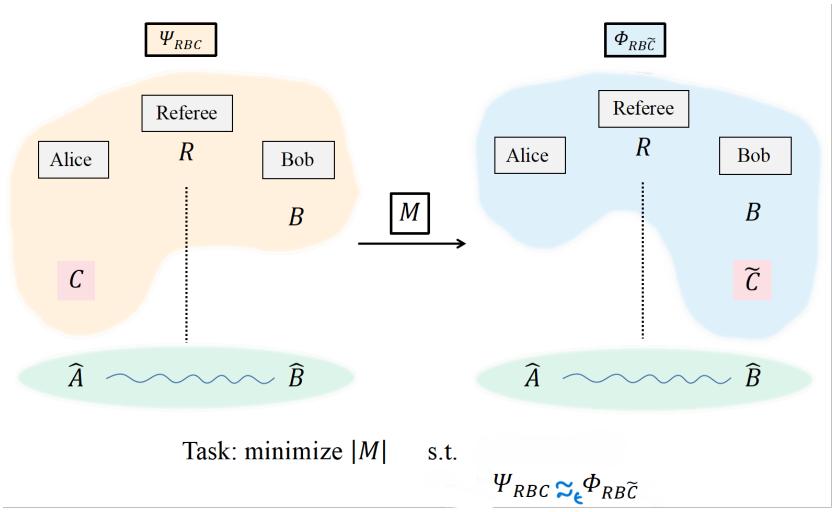

Figure 6.4: Quantum state merging. A special case of state redistribution when the register A is missing.

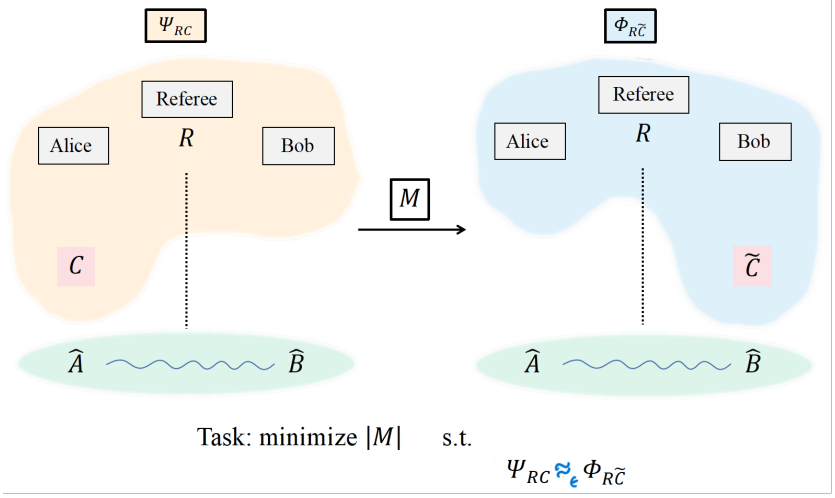

Figure 6.5: Quantum state transfer. A special case of state redistribution when the registers A, B are missing.

Protocol	Alice Has	Bob Has	Goal
State Splitting	A, C	*none*	Split AC and send C to Bob
State Merging	C	B	Merge C with B
State Transfer	C	*none*	Transfer C to Bob

Table 6.1: Comparison of quantum communication protocols depending on the registers available to Alice and Bob.

6.3 *Convex-split lemma (CSL)*

To design some protocols for the above tasks, we will introduce a widely used tool in quantum information theory known as the convex-split lemma [2]. We start with some notation, definitions, and facts.

Quantum information primitives. We denote quantum registers by capital alphabetic A, B, etc. We denote density operators by $\rho_{AB} \in \mathcal{D}(AB), \sigma_A \in$

[2] Anurag Anshu, Vamsi Krishna Devabathini, and Rahul Jain. Quantum communication using coherent rejection sampling. *Physical Review Letters*, 119(12), September 2017. DOI: 10.1103/physrevlett.119.120506. URL http://dx.doi.org/10.1103/PhysRevLett.119.120506

$\mathcal{D}(A), \tau_B \in \mathcal{D}(B)$ and so on.

Definition 6.3 (Relative-entropy)**.** *For quantum states ρ, σ, the relative-entropy between them is defined as,*

$$\mathrm{D}(\rho\|\sigma) \stackrel{def}{=} \mathrm{Tr}\rho\log\rho - \mathrm{Tr}\rho\log\sigma.$$

Definition 6.4 (Max relative-entropy)**.** *Let ρ, σ be quantum states with $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. The max relative-entropy between them is defined as,*

$$\mathrm{D}_{\max}(\rho\|\sigma) := \inf\left\{\lambda \in \mathbb{R} \mid \rho \leq 2^{\lambda}\sigma\right\}.$$

We have the following facts.

Fact 6.5. *For states ρ, σ, τ,*

$$\mathrm{D}(\rho\otimes\tau\|\sigma\otimes\tau) = \mathrm{D}(\rho\|\sigma).$$

Fact 6.6. *Let $A \geq 0, B \geq C$, then $\mathrm{Tr}AB \geq \mathrm{Tr}AC$.*

Fact 6.7 (Pinsker's inequality)**.** *For states ρ, σ,*

$$\mathrm{F}(\rho,\sigma) \geq 2^{-\mathrm{D}(\rho\|\sigma)}.$$

This implies,

$$1 - \mathrm{F}(\rho,\sigma) \leq (\ln 2)\cdot\mathrm{D}(\rho\|\sigma).$$

Fact 6.8 (Data processing inequality (DPI))**.** *For states ρ, σ and a CPTP map $\mathcal{E}$, we have*

$$\mathrm{D}(\rho\|\sigma) \geq \mathrm{D}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)).$$

Consider states $\rho_{AB}, \sigma_{AB} \in \mathcal{D}(AB)$ and let $\mathcal{E}$ be the partial trace of system B. Then,

$$\mathrm{D}(\rho_{AB}\|\sigma_{AB}) \geq \mathrm{D}(\rho_A\|\sigma_A),$$

where $\rho_A = \mathrm{Tr}_B\rho_{AB}$ and $\sigma_A = \mathrm{Tr}_B\sigma_{AB}$.

We begin with the following lemma.

Lemma 6.9. *Let $\mu_1, \mu_2, \ldots, \mu_n, \theta$ be states and $\{p_1, p_2, \ldots p_n\}$ be a prob-*

ability distribution. Let $\mu = \sum_i p_i \mu_i$ be the average state. Then

$$\mathrm{D}(\mu\|\theta) = \sum_i p_i \left(\mathrm{D}(\mu_i\|\theta) - \mathrm{D}(\mu_i\|\mu)\right).$$

Proof. Consider,

$$\begin{aligned}
&\sum_i p_i(\mathrm{D}(\mu_i\|\theta) - \mathrm{D}(\mu_i\|\mu)) \\
&= \sum_i p_i(\mathrm{Tr}\mu_i \log \mu_i - \mathrm{Tr}\mu_i \log \theta - \mathrm{Tr}\mu_i \log \mu_i + \mathrm{Tr}\mu_i \log \mu) \\
&= \sum_i p_i(-\mathrm{Tr}\mu_i \log \theta + \mathrm{Tr}\mu_i \log \mu) \\
&= -\mathrm{Tr}\mu \log \theta + \mathrm{Tr}\mu \log \mu \\
&= \mathrm{D}(\mu\|\theta).
\end{aligned}$$

□

Now we are ready to state and prove the convex-split lemma.

Intuition: The convex-split lemma tells us that if a quantum state φ_{RC} is close (in max relative-entropy) to a product state $\varphi_R \otimes \sigma_C$, then by mixing φ_{RC} among n copies of σ_C, the resulting mixture becomes close to $\varphi_R \otimes \sigma_C^{\otimes n}$. The more copies n, the better the approximation. Please refer to Figure 6.6

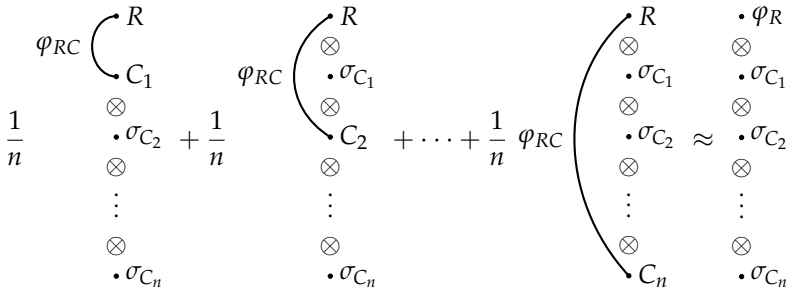

Figure 6.6: State equivalence for the convex-split lemma.

The convex-split lemma guarantees that the state on the LHS (a convex combination where each term has the actual state φ_{RC} in position Referee and C_j and the state σ_C elsewhere) is approximately equal to the product state on the RHS, with fidelity at least $1 - \delta$.

Lemma 6.10 (Convex-split lemma (CSL))**.** *Let $\varphi_{RC} \in \mathcal{D}(RC)$ and $\sigma_C \in \mathcal{D}(C)$ be states such that* $\mathrm{supp}(\varphi_C) \subset \mathrm{supp}(\sigma_C)$*. Let*

$$k := \mathrm{D}_{\max}(\varphi_{RC}\|\varphi_R \otimes \sigma_C).$$

Consider the following $(n+1)$-partite state,

$$\tau_{RC_1C_2\ldots C_n} := \frac{1}{n}\sum_{j=1}^{n} \varphi_{RC_j} \otimes \sigma_{C_1} \otimes \cdots \sigma_{C_{j-1}} \otimes \sigma_{C_{j+1}} \otimes \cdots \otimes \sigma_{C_n},$$

where $\varphi_{RC_j} = \varphi_{RC}$ and $\sigma_{C_j} = \sigma_C$ for all $j \in \{1, 2, \ldots, n\}$. Then,

$$\mathrm{D}(\tau_{RC_1C_2\ldots C_n} \| \varphi_C \otimes \sigma_{C_1} \otimes \cdots \otimes \sigma_{C_n}) \leq \log\left(1 + \frac{2^k}{n}\right).$$

Using Pinsker's inequality (Fact 6.7) we get,

$$\mathrm{F}(\tau_{RC_1C_2\ldots C_n}, \varphi_C \otimes \sigma_{C_1} \otimes \cdots \otimes \sigma_{C_n}) \geq \frac{1}{1+\frac{2^k}{n}}.$$

In particular for $\delta > 0$ and $n = \left\lceil \frac{2^k}{\delta} \right\rceil$,

$$\mathrm{D}(\tau_{RC_1C_2\ldots C_n} \| \varphi_C \otimes \sigma_{C_1} \otimes \cdots \otimes \sigma_{C_n}) \leq \log(1+\delta),$$

and

$$\mathrm{F}(\tau_{RC_1C_2\ldots C_n}, \varphi_C \otimes \sigma_{C_1} \otimes \cdots \otimes \sigma_{C_n}) \geq 1 - \delta.$$

Proof. Denote the n copies of σ_C state by

$$\bar{\sigma} := \sigma_{C_1} \otimes \sigma_{C_2} \otimes \cdots \otimes \sigma_{C_n}.$$

Define a set of quantum states $\left\{\bar{\sigma}^{(-j)}\right\}$ for all $j \in [n]$ as

$$\bar{\sigma}^{(-j)} := \sigma_{C_1} \otimes \sigma_{C_2} \otimes \cdots \otimes \sigma_{C_{j-1}} \otimes \sigma_{C_{j+1}} \otimes \cdots \otimes \sigma_{C_n}.$$

Then we have,

$$\tau_{RC_1C_2\ldots C_n} = \frac{1}{n}\sum_{j=1}^{n} \varphi_{RC_j} \otimes \bar{\sigma}^{(-j)}.$$

Consider

$$\begin{aligned}
\tau_{RC_j} &= \frac{1}{n}\varphi_{RC_j} + \frac{n-1}{n}\left(\varphi_R \otimes \sigma_{C_j}\right) \\
&\leq \frac{2^k}{n}\left(\varphi_R \otimes \sigma_{C_j}\right) + \frac{n-1}{n}\left(\varphi_R \otimes \sigma_{C_j}\right) \qquad (k := \mathrm{D}_{\max}(\varphi_{RC} \| \varphi_R \otimes \sigma_C)) \\
&= \left(1 + \frac{2^k - 1}{n}\right)\left(\varphi_R \otimes \sigma_{C_j}\right).
\end{aligned}$$

Taking logarithm on both the sides (since log is operator monotonic),

$$\log \tau_{RC_j} \leq \log\left(1 + \frac{2^k - 1}{n}\right)\mathbb{I} + \log\left(\varphi_R \otimes \sigma_{C_j}\right). \tag{6.1}$$

Consider

$$
\begin{aligned}
&\mathrm{D}(\tau_{RC_1C_2\ldots C_n}\|\varphi_C\otimes\bar{\sigma})\\
&=\frac{1}{n}\sum_{j=1}^{n}\Big[\mathrm{D}(\varphi_{RC_j}\otimes\bar{\sigma}^{(-j)}\|\varphi_R\otimes\bar{\sigma})-\mathrm{D}(\varphi_{RC_j}\otimes\bar{\sigma}^{(-j)}\|\tau_{RC_1C_2\ldots C_n})\Big] && \text{(Lemma 6.9)}\\
&=\frac{1}{n}\sum_{j=1}^{n}\Big[\mathrm{D}(\varphi_{RC_j}\|\varphi_R\otimes\sigma_{C_j})-\mathrm{D}(\varphi_{RC_j}\otimes\bar{\sigma}^{(-j)}\|\tau_{RC_1C_2\ldots C_n})\Big] && \text{(Fact 6.5)}\\
&\le\frac{1}{n}\sum_{j=1}^{n}\Big[\mathrm{D}(\varphi_{RC_j}\|\varphi_R\otimes\sigma_{C_j})-\mathrm{D}(\varphi_{RC_j}\|\tau_{RC_j})\Big] && \text{(Fact 6.8)}\\
&=\frac{1}{n}\sum_{j=1}^{n}\Big[\mathrm{Tr}\varphi_{RC_j}\log\varphi_{RC_j}-\mathrm{Tr}\varphi_{RC_j}\log(\varphi_R\otimes\sigma_{C_j})\\
&\qquad\qquad-\mathrm{Tr}\varphi_{RC_j}\log\varphi_{RC_j}+\mathrm{Tr}\varphi_{RC_j}\log\tau_{RC_j}\Big]\\
&=\frac{1}{n}\sum_{j=1}^{n}\Big[-\mathrm{Tr}\varphi_{RC_j}\log(\varphi_R\otimes\sigma_{C_j})+\mathrm{Tr}\varphi_{RC_j}\log\tau_{RC_j}\Big]\\
&\le\frac{1}{n}\sum_{j=1}^{n}\Big[-\mathrm{Tr}\varphi_{RC_j}\log(\varphi_R\otimes\sigma_{C_j})\\
&\qquad\qquad+\mathrm{Tr}\varphi_{RC_j}\big(\log\left(1+\tfrac{2^k-1}{n}\right)\mathbb{I}+\log(\varphi_R\otimes\sigma_{C_j})\big)\Big] && \text{(Eq. (6.1), Fact 6.6)}\\
&=\log\left(1+\frac{2^k-1}{n}\right)\quad\le\quad\log\left(1+\frac{2^k}{n}\right).
\end{aligned}
$$

Using Pinsker's inequality (Fact 6.7) we get,

$$
\mathrm{F}^2(\tau_{RC_1C_2\ldots C_n},\varphi_C\otimes\bar{\sigma})\ge 2^{-\mathrm{D}(\tau_{RC_1C_2\ldots C_n}\|\varphi_C\otimes\bar{\sigma})}\ge 2^{-\log\left(1+\frac{2^k}{n}\right)}=\frac{1}{1+\frac{2^k}{n}}.
$$

In particular for $\delta>0$ and $n=\left\lceil\frac{2^k}{\delta}\right\rceil$,

$$
\mathrm{D}(\tau_{RC_1C_2\ldots C_n}\|\varphi_C\otimes\sigma_{C_1}\otimes\cdots\otimes\sigma_{C_n})\le\log(1+\delta),
$$

and

$$
\mathrm{F}(\tau_{RC_1C_2\ldots C_n},\varphi_C\otimes\sigma_{C_1}\otimes\cdots\otimes\sigma_{C_n})\ge(1+\delta)^{-1}\ge 1-\delta.
$$

□

7

Quantum state splitting

7.0.1 *Problem setting*

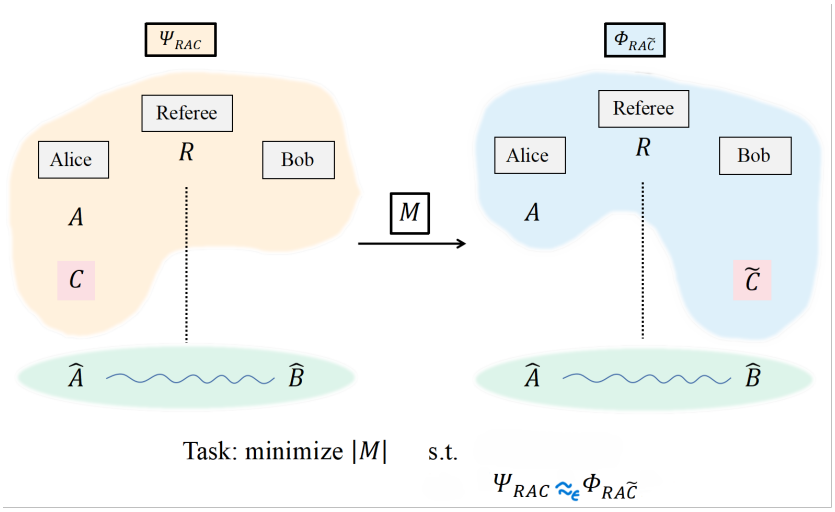

Figure 7.1: Quantum state splitting. The special case of state redistribution when the register B is missing.

7.1 *Protocol for state splitting*

Here we discuss a protocol for state splitting [1]. Let $\varepsilon \geq 0, \delta > 0$ be the error parameters. Let n be the smallest number such that,

$$\log n \geq \mathrm{I}^{\varepsilon}_{\max}(R : C)_{\psi_{RC}} + 2\log\frac{1}{\delta}$$

where,

Definition 7.1 (Max mutual information). *For a quantum state ψ_{RC}, the max mutual information is defined as,*

$$\mathrm{I}_{\max}(R : C)_{\psi_{RC}} \stackrel{def}{=} \inf_{\sigma_C} \mathrm{D}_{\max}(\psi_{RC} \| \psi_R \otimes \sigma_C).$$

[1] Anurag Anshu, Vamsi Krishna Devabathini, and Rahul Jain. Quantum communication using coherent rejection sampling. *Physical Review Letters*, 119(12), September 2017. DOI: 10.1103/physrevlett.119.120506. URL http://dx.doi.org/10.1103/PhysRevLett.119.120506

Definition 7.2 (Smooth max mutual information). *Let $\varepsilon \geq 0$. For a quantum state ψ_{RC}, the ε-smooth max mutual information is defined as,*

$$\begin{aligned} \mathrm{I}^{\varepsilon}_{\max}(R:C)_{\psi_{RC}} &\stackrel{def}{=} \inf_{\psi'_{RC}:\mathrm{P}(\psi_{RC},\psi'_{RC})\leq\varepsilon} \mathrm{I}_{\max}(R:C)_{\psi'_{RC}} \\ &= \inf_{\psi'_{RC}:\mathrm{P}(\psi_{RC},\psi'_{RC})\leq\varepsilon} \inf_{\sigma_C} \mathrm{D}_{\max}(\psi'_{RC}\|\psi'_R\otimes\sigma_C). \end{aligned}$$

The ε-partially smooth max mutual information is defined as,

$$\mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{R}:C)_{\psi_{RC}} \stackrel{def}{=} \inf_{\psi'_{RC}:\mathrm{P}(\psi_{RC},\psi'_{RC})\leq\varepsilon\ ;\ \psi_R=\psi'_R} \inf_{\sigma_C} \mathrm{D}_{\max}(\psi'_{RC}\|\psi_R\otimes\sigma_C).$$

Let $\psi'_{RC} \approx_\varepsilon \psi_{RC}$ (with $\psi_R = \psi'_R$) and σ_C be such that,

$$\mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{R}:C)_{\psi_{RC}} = \mathrm{D}_{\max}(\psi'_{RC}\|\psi_R\otimes\sigma_C).$$

We will need the following fundamental result in quantum information theory.

Fact 7.3 (Uhlmann's Theorem). *Let $|\theta\rangle_{A'B}, |\gamma\rangle_{AB}$ be pure states. There exists an isometry $V : A' \to A$, such that:*

$$\mathrm{F}\big(V|\theta\rangle\langle\theta|V^\dagger, |\gamma\rangle\langle\gamma|\big) = \mathrm{F}(\theta_B, \gamma_B).$$

V is unitary if $A' \equiv A$.

Let $|\psi'\rangle_{RAC}$ be a purification ψ'_{RC} (guaranteed by Uhlmann's theorem) such that

$$\mathrm{F}(|\psi'\rangle\langle\psi'|_{RAC}, |\psi\rangle\langle\psi|_{RAC}) = \mathrm{F}(\psi'_{RC}, \psi_{RC}).$$

By the CSL (Lemma 6.10), we know that the purified distance between the reduced states (with Referee and Bob) between the LHS and RHS of Figure 7.3 is at most $\sqrt{\delta^2} = \delta$. Uhlmann's theorem (in the context of our protocol, $|\theta\rangle$ is the initial state on LHS and $|\gamma\rangle$ is the target state on RHS of Figure 7.2) guarantees the existence of an isometry V that Alice can apply to achieve $\mathrm{P}(V\theta V^\dagger, \gamma) \leq \delta$.

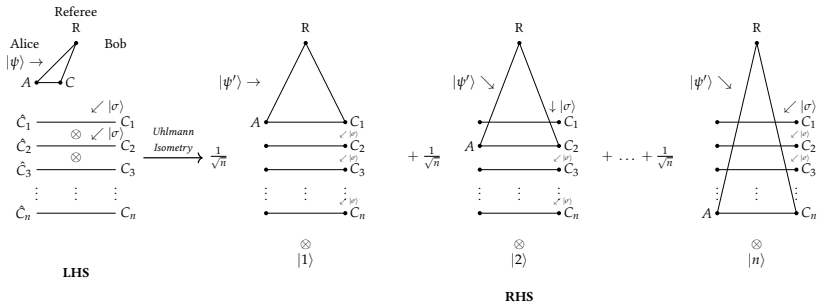

Figure 7.2: Quantum state splitting protocol: The left-hand side (LHS) shows the initial state with Alice holding registers A and C, and Alice and Bob sharing entangled pairs. The right-hand side (RHS) shows the state (that is close to the state after applying Uhlmann's isometry) which is a superposition with coefficients $1/\sqrt{n}$.

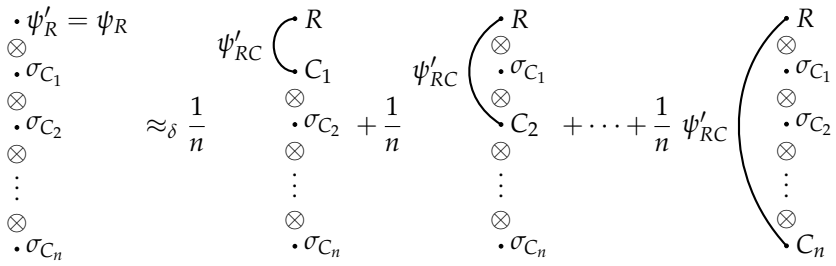

Figure 7.3: State Equivalence: The convex-split lemma guarantees that the state on the RHS (a convex combination where each term has the actual state ψ'_{RC} in position Referee and C_j and the state σ_C elsewhere) is close to the product state on the LHS, with purified distance at most δ.

7.1.1 Protocol steps

1. The protocol starts with the initial state $|\psi\rangle_{RAC}$ shared between Referee (R) and Alice (AC).
2. Alice and Bob share n i.i.d (independent and identically distributed) copies of a purification of σ_C, denoted as $|\sigma\rangle$ in registers $\hat{C}_i$ and C_i for $i \in \{1, 2, ..., n\}$.
3. Alice performs the Uhlmann isometry V on her registers based on the CSL.
4. Alice measures the index register j which collapses the superposition.
5. Alice sends the measurement outcome j to Bob using $\log n$ bits of communication.
6. Alice and Bob swap registers according to the value of j.

7.1.2 Communication cost

The classical communication cost is given by:

$$\log n = \mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{R} : C)_{\psi_{RC}} + 2\log\frac{1}{\delta}.$$

7.1.3 Error analysis

We need the following property about partial traces in quantum systems.

Fact 7.4. *Let*

$$|\gamma\rangle_{A_1A_2B} = \sum_{i=1}^{n} \alpha_i |i\rangle_{A_1} |\gamma^i\rangle_{A_2B}$$

be a state (where $\sum_{i=1}^{n} |\alpha_i|^2 = 1$). The reduced state on system B is given by:

$$\gamma_B = \sum_{i=1}^{n} |\alpha_i|^2 \gamma^i_B.$$

This explains why the convex combination on the right side of Figure 7.3 appears when we trace out Alice's systems. Each term in the superposition contributes a component in the RHS of Figure 7.3 with probability $\frac{1}{n}$. The CSL guarantees that this mixture is (δ close to) the desired product state shown on the LHS of Figure 7.3.

At Step 4. in the protocol, when Alice measures the index register, Referee and Bob's combined state becomes (δ close to) a probabilistic mixture weighted by $\frac{1}{n}$.

When Alice sends her measurement outcome j to Bob, Bob knows which register contains the target state, allowing him to effectively "undo" the mixing effect and recover a state $\varphi_{RA\hat{C}} \approx_\delta \psi'_{RAC}$.

Since $\psi_{RAC} \approx_\varepsilon \psi'_{RAC}$,

$$\psi_{RAC} \approx_\varepsilon \psi'_{RAC} \approx_\delta \varphi_{RA\hat{C}}.$$

Using the triangle inequality for the purified distance, we get

$$\psi_{RAC} \approx_{\varepsilon+\delta} \varphi_{RA\hat{C}}.$$

8

State splitting - alternate protocol

8.1 Alternate protocol for state splitting

Here we discuss an alternate protocol for state splitting [1].

Let $\varepsilon \geq 0, \delta > 0$ be the error parameters. Let n be the smallest number such that,

$$\log n \geq \mathrm{I}^{\varepsilon}_{\max}(R:C)_{\psi_{RC}} + 2\log\frac{1}{\delta}.$$

Let $\psi'_{RC} \approx_{\varepsilon} \psi_{RC}$ and σ_C be such that,

$$\mathrm{I}^{\varepsilon}_{\max}(R:C)_{\psi_{RC}} = \mathrm{D}_{\max}(\psi'_{RC} \| \psi'_R \otimes \sigma_C).$$

Let $|\psi'\rangle_{RAC}$ be a purification ψ'_{RC} (guaranteed by Uhlmann's theorem) such that

$$\mathrm{F}(|\psi'\rangle\langle\psi'|_{RAC}, |\psi\rangle\langle\psi|_{RAC}) = \mathrm{F}(\psi'_{RC}, \psi_{RC}).$$

By the CSL (Lemma 6.10), we know that the purified distance between the reduced states (with Referee and Bob) is at most $\sqrt{\delta^2} = \delta$ (see Figure 8.2). Uhlmann's theorem (in the context of our protocol (Figure 8.1), $|\theta\rangle$ is the initial state on LHS and $|\gamma\rangle$ is the target state on RHS) guarantees the existence of an isometry V that Alice can apply to achieve $\mathrm{P}(V\theta V^{\dagger}, \gamma) \leq \delta$.

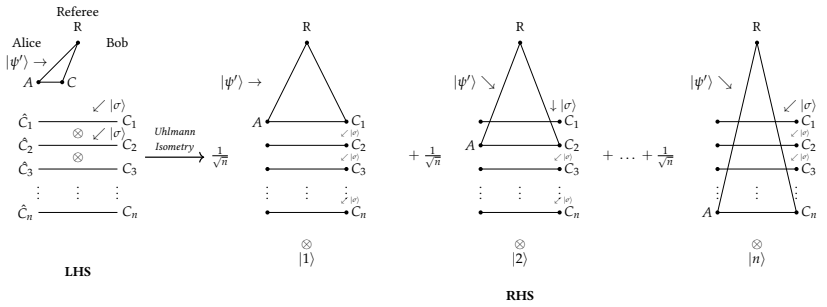

Figure 8.1: Quantum state splitting protocol: The left-hand side (LHS) shows the initial state with Alice holding registers A and C, and Alice and Bob sharing entangled pairs. The right-hand side (RHS) shows the state (that is close to the state after applying Uhlmann's isometry) which is a superposition with coefficients $1/\sqrt{n}$.

[1] Anurag Anshu, Vamsi Krishna Devabathini, and Rahul Jain. Quantum communication using coherent rejection sampling. *Physical Review Letters*, 119(12), September 2017. DOI: 10.1103/physrevlett.119.120506. URL http://dx.doi.org/10.1103/PhysRevLett.119.120506

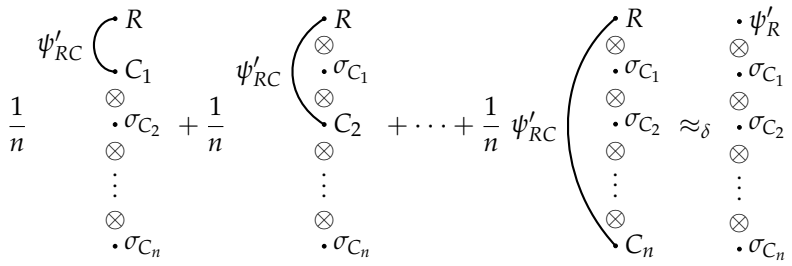

Figure 8.2: State Equivalence: The convex-split lemma guarantees that the state on the LHS (a convex combination where each term has the actual state ψ'_{RC} in position Referee and C_j and the state σ_C elsewhere) is approximately equal to the product state on the RHS, with purified distance at most δ.

8.1.1 Protocol steps

1. The protocol starts with the initial state $|\psi\rangle_{RAC}$ shared between Referee (R) and Alice (AC).
2. Alice and Bob share n i.i.d (independent and identically distributed) copies of a purification of σ_C, denoted as $|\sigma\rangle$ in registers $\hat{C}_i$ and C_i for $i \in \{1, 2, ..., n\}$.
3. Alice performs the Uhlmann isometry V on her registers based on the CSL.
4. Alice measures the index register j which collapses the superposition.
5. Alice sends the measurement outcome j to Bob using $\log n$ bits of communication.
6. Alice and Bob swap registers according to the value of j.

8.1.2 Communication cost

The classical communication cost is given by:

$$\log n = \mathrm{I}^{\varepsilon}_{\max}(R : C)_{\psi_{RC}} + 2 \log \frac{1}{\delta}.$$

8.1.3 Error analysis

Assume first that the protocol starts with the initial state $|\psi'\rangle_{RAC}$ shared between Referee (R) and Alice (AC). Fact 7.4 explains why the convex combination on the right side of Figure 8.2 appears when we trace out Alice's systems. Each term in the superposition contributes a component in the LHS of Figure 8.2 with probability $\frac{1}{n}$. The CSL guarantees that this mixture is (δ close to) the desired product state shown on the RHS of Figure 8.2.

At Step 4. in the protocol, when Alice measures the index register, Referee and Bob's combined state becomes (δ close to) a probabilistic mixture weighted by $\frac{1}{n}$.

When Alice sends her measurement outcome j to Bob, Bob knows which register contains the target state, allowing him to effectively "undo" the mixing effect and recover a state $\varphi'_{RA\hat{C}} \approx_\delta \psi'_{RAC}$.

Now assume that the protocol starts with the initial state ψ_{RAC} instead. Since $\psi_{RAC} \approx_\varepsilon \psi'_{RAC}$,

$$\psi_{RAC} \approx_\varepsilon \psi'_{RAC} \approx_\delta \varphi'_{RA\hat{C}} \approx_\varepsilon \varphi_{RA\hat{C}}.$$

The last approximation above follows using DPI for fidelity (and hence the purified distance) and noting that the entire communication protocol can be thought of as a CPTP map from the input state on the registers RAC to the output state on the registers $RA\hat{C}$.

Using the triangle inequality for the purified distance, we get

$$\psi_{RAC} \approx_{2\varepsilon+\delta} \varphi_{RA\hat{C}}.$$

Open question: Is it possible to get, for all $\delta > 0$, a protocol where

$$\psi_{RAC} \approx_{\varepsilon+\delta} \varphi_{RA\hat{C}}$$

with communication cost

$$\mathrm{I}^{\varepsilon}_{\max}(R:C)_{\psi_{RC}} + f(\delta)?$$

where f is any function of δ.

9

Converse bound for quantum state splitting

9.1 State splitting

Here, we want to prove the converse (lower bound) of state splitting [1]. The diagram for state splitting is shown in Figure 9.1 and the circuit diagram for the protocol for state splitting is shown in Figure 9.2.

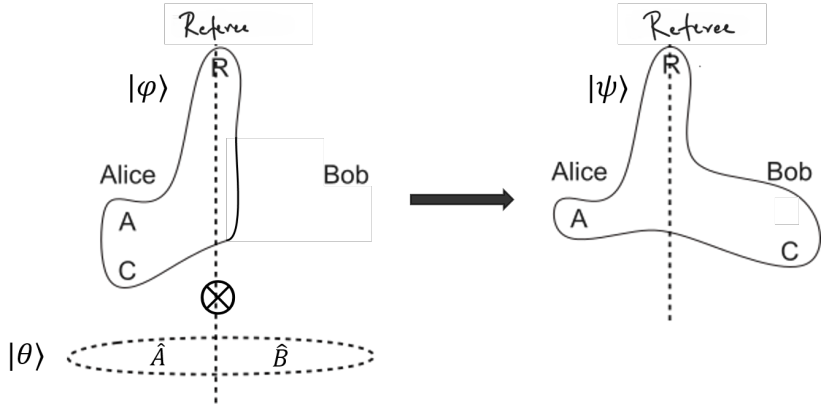

Figure 9.1: Diagram of state splitting with entanglement assistance. Alice splits the states in the system AC and Bob receives the state in the system C.

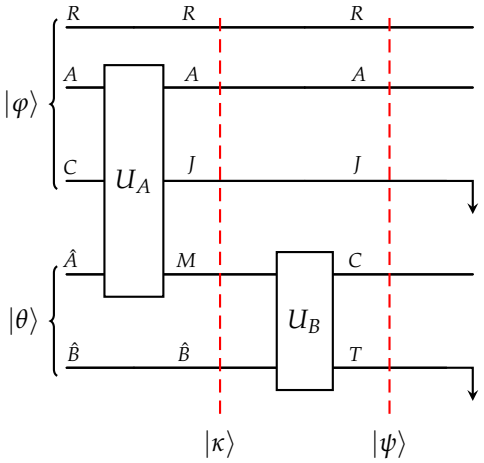

Figure 9.2: Circuit diagram of state splitting with entanglement assistance. Alice and Bob start with the state $|\varphi\rangle_{RAC}$ and share an entangled state $|\theta\rangle_{\hat{A}\hat{B}}$. U_A and U_B are unitary operators. $|\kappa\rangle$ is the state after U_A is applied and $|\psi\rangle$ is the final state after U_B is applied.

[1] Mario Berta, Matthias Christandl, and Dave Touchette. Smooth entropy bounds on one-shot quantum state redistribution. *IEEE Transactions on Information Theory*, 62(3): 1425–1439, 2016

We need the following facts.

Fact 9.1. *For state ρ_{AB},*

$$\mathrm{I}_{\max}(A:B)_\rho \leq 2\min\{\log|A|, \log|B|\}$$

where $|A| = \dim(\mathrm{supp}(\rho_A))$.

Fact 9.2. *For states ρ, σ and unitary U,*

$$\mathrm{D}_{\max}(\rho\|\sigma) = \mathrm{D}_{\max}(U\rho U^\dagger\|U\sigma U^\dagger).$$

9.1.1 Proof of the converse bound

Using Fact 9.1 (with $A \leftarrow R\hat{B}$ and $B \leftarrow M$),

$$\begin{aligned} 2\log|M| &\geq \mathrm{I}_{\max}(R\hat{B}:M)_\kappa \\ &= \min_{\sigma_M} \mathrm{D}_{\max}(\kappa_{R\hat{B}M}\|\kappa_{R\hat{B}}\otimes\sigma_M). \end{aligned}$$

Let ω_M be the state that minimizes the above. Notice that in Figure 9.2, the systems R and $\hat{B}$ remain unchanged after U_A, so $\kappa_{R\hat{B}} = \varphi_R \otimes \theta_{\hat{B}}$. From correctness of the protocol $\varphi_{RAC} \approx_\varepsilon \psi_{RAC}$. This implies (using DPI for purified distance) $\varphi_{RC} \approx_\varepsilon \psi_{RC}$. Hence,

$$\begin{aligned} 2\log|M| &\geq \mathrm{D}_{\max}(\kappa_{R\hat{B}M}\|\kappa_{R\hat{B}}\otimes\omega_M) \\ &= \mathrm{D}_{\max}(\kappa_{R\hat{B}M}\|\varphi_R\otimes\theta_{\hat{B}}\otimes\omega_M) \\ &= \mathrm{D}_{\max}(U_B\kappa_{R\hat{B}M}U_B^\dagger\|\varphi_R\otimes U_B(\theta_{\hat{B}}\otimes\omega_M)U_B^\dagger) && \text{(Fact 9.2)} \\ &= \mathrm{D}_{\max}(\psi_{RCT}\|\varphi_R\otimes\tau_{CT}) && (\tau_{CT} := U_B(\theta_{\hat{B}}\otimes\omega_M)U_B^\dagger) \\ &\geq \mathrm{D}_{\max}(\psi_{RC}\|\varphi_R\otimes\tau_C) && \text{(DPI for } \mathrm{D}_{\max}(\cdot\|\cdot)) \\ &\geq \mathrm{I}^\varepsilon_{\max}(\overset{\bullet}{R}:C)_\varphi. && \text{(Definition 7.2 and } \psi_R = \varphi_R \text{ and } \varphi_{RC} \approx_\varepsilon \psi_{RC}) \end{aligned}$$

9.1.2 Classical communication v/s quantum communication

The protocol for state-splitting presented in the previous lecture used entanglement-assisted classical communication. Recall that using *superdense coding*, classical communication can be converted to quantum communication by reducing the communication cost by a factor of 2 (by sending one qubit per 2 classical bits). This justifies the factor 2 difference between the achievability and the converse bounds.

Using superdense coding, we get that if the communication is classical then,

$$\log|M| \geq \mathrm{I}^\varepsilon_{\max}(\overset{\bullet}{R}:C)_\varphi.$$

9.1.3 *Putting together*

Let cc-ss(ε) denote the optimal entanglement-assisted classical communication cost for state splitting with the purified distance between the input pure state (φ_{RC}) and the output state (ψ_{RC}) being at most ε. Then from the achievability (from the two protocols) and the converse bounds we get $\forall \varepsilon \geq 0, \delta > 0$:

$$\mathrm{I}_{\max}^{\varepsilon+\delta}(\overset{\bullet}{R} : C)_{\varphi} \leq \text{cc-ss}(\varepsilon+\delta) \leq \mathrm{I}_{\max}^{\varepsilon}(\overset{\bullet}{R} : C)_{\varphi} + 2\log\frac{1}{\delta},$$

$$\mathrm{I}_{\max}^{2\varepsilon+\delta}(R : C)_{\varphi} \leq \mathrm{I}_{\max}^{2\varepsilon+\delta}(\overset{\bullet}{R} : C)_{\varphi} \leq \text{cc-ss}(2\varepsilon+\delta) \leq \mathrm{I}_{\max}^{\varepsilon}(R : C)_{\varphi} + 2\log\frac{1}{\delta}.$$

The first inequality in the second expression above follows from the definitions.

9.2 *Quantum state merging*

State merging is basically the time reversal of state splitting. The diagram for state merging is shown in Figure 9.3 and the circuit diagram for the protocol for state merging is shown in Figure 9.4. Due to this, the achievability and the converse bounds are the same.

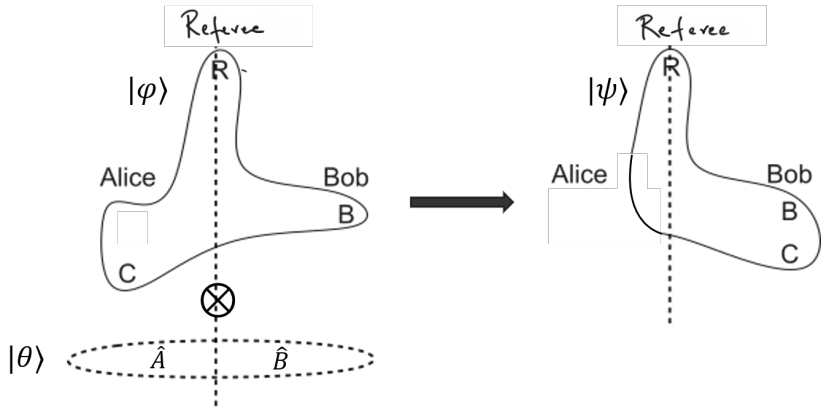

Figure 9.3: Diagram of state merging with entanglement assistance. Alice sends the state in the system C and Bob merges the state with his state in the system B.

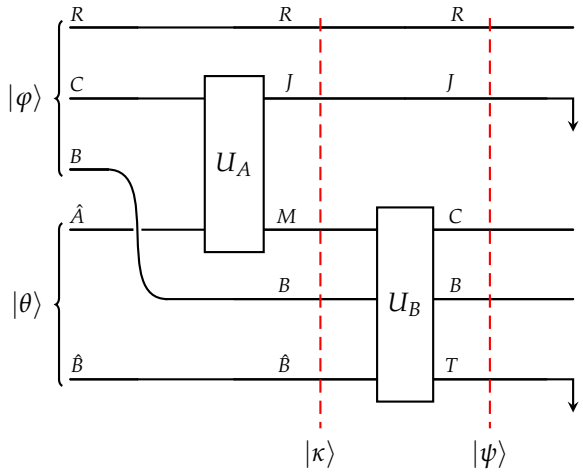

Figure 9.4: Circuit diagram of state merging with entanglement assistance. Alice and Bob start with the state $|\varphi\rangle_{RCB}$ and share an entangled state $|\theta\rangle_{\hat{A}\hat{B}}$. U_A and U_B are unitary gates. $|\kappa\rangle$ is the state after U_A is applied and $|\psi\rangle$ is the final state after U_B is applied.

10
Channel coding and position-based decoding

10.1 Introduction

In this lecture, we introduce the concept of channel coding and explore the *position-based decoding* (PBD) strategy [1].

[1] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. Building blocks for communication over noisy quantum networks. *IEEE Transactions on Information Theory*, 65(2):1287–1306, February 2019b. DOI: 10.1109/TIT.2018.2851297

10.2 Point-to-point quantum communication protocol

A communication protocol using a *point-to-point* quantum channel, involves two parties, Alice (the sender) and Bob (the receiver). Alice is allowed to use the quantum channel $\mathcal{N}_{A\to B}$ once to transmit a message m drawn uniformly from $[2^R]$. Bob decodes $\hat{m}$ and we want $\Pr[m = \hat{m}] \geq 1 - \varepsilon$, where ε is a small error tolerance. Alice and Bob are allowed to use a prior entangled state between them.

This communication strategy is known as an (R, ε) *entanglement-assisted code* for the quantum channel $\mathcal{N}_{A\to B}$. The primary objective is to determine the maximum achievable value of R, representing the number of reliable bits transmitted between Alice and Bob.

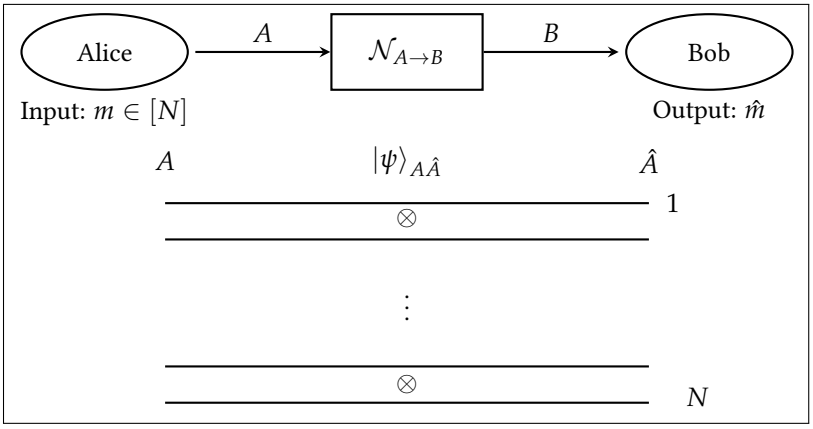

Figure 10.1: Illustration of the position-based decoding protocol.

10.2.1 Position-based decoding (PBD)

We start with the following definition.

Definition 10.1 (Smooth hypothesis testing relative-entropy). *Let ρ, σ be states and $\varepsilon \geq 0$. The* smooth hypothesis testing relative-entropy *is defined as follows:*

$$\mathrm{D}_H^\varepsilon(\rho\|\sigma) \stackrel{def}{=} \max_{\substack{0\leq T\leq \mathbb{I} \\ \mathrm{Tr}(T\rho)\geq 1-\varepsilon}} \log\left(\frac{1}{\mathrm{Tr}\, T\sigma}\right).$$

Let $\varepsilon > 0$. Define,

$$R := \max_{|\varphi\rangle_{A\hat{A}}} \left\{ \mathrm{D}_H^\varepsilon(\theta_{B\hat{A}}\|\theta_B \otimes \varphi_{\hat{A}}) - \log\frac{1}{\varepsilon} \quad \middle| \quad \theta_{B\hat{A}} := \mathcal{N}_{A\to B}(\varphi_{A\hat{A}}) \right\}.$$

Let $|\psi\rangle_{A\hat{A}}$ be the state that achieves the maximum above. Let,

$$k := \mathrm{D}_H^\varepsilon(\theta_{B\hat{A}}\|\theta_B \otimes \psi_{\hat{A}}) \quad ; \quad N := \lfloor 2^R \rfloor = \lfloor \varepsilon \cdot 2^k \rfloor. \tag{10.1}$$

The communication protocol is described as follows (refer to Figure 10.1).

1. Alice and Bob start with $[N]$ i.i.d. copies of $|\psi\rangle_{A\hat{A}}$ shared between them.
2. On receiving input $m \in [N]$, Alice inserts the register A of the mth copy of $|\psi\rangle_{A\hat{A}}$ into the channel.
3. After receiving channel's output, Bob measures the state at his end $\tau^m_{B\hat{A}_1\ldots\hat{A}_N}$ according to the POVM

   $$\{\Omega_j \mid j \in [N+1]\},$$

   as described below.
4. Bob outputs $\hat{m}$, which is the outcome of the measurement above.

10.2.2 Bob's decoding measurement

Let $T_{B\hat{A}}$ be the operator that achieves the maximum in the definition of $\mathrm{D}_H^\varepsilon(\theta_{B\hat{A}}\|\theta_B \otimes \psi_{\hat{A}})$. For all $j \in [N]$, define the following operators,

$$\Lambda(j) := \mathbb{I}_{\hat{A}_1} \otimes \cdots \otimes T_{B\hat{A}_j} \otimes \cdots \otimes \mathbb{I}_{\hat{A}_N} \quad ; \quad \Omega(j) := \Lambda^{-1/2}\Lambda(j)\Lambda^{-1/2},$$

where,

$$\Lambda := \sum_j \Lambda(j).$$

We can note that for all $j \in [N]:\ 0 \leq \Lambda(j) \leq \mathbb{I},\ 0 \leq \Omega(j)$, and

$$\mathrm{Tr}\, \tau^j_{B\hat{A}_1\ldots\hat{A}_N}\Lambda(j) = \mathrm{Tr}\, \theta_{B\hat{A}}T_{B\hat{A}} \geq 1-\varepsilon, \tag{10.2}$$

$$\forall i \neq j : \mathrm{Tr}\, \tau^i_{B\hat{A}_1\ldots\hat{A}_N}\Lambda(j) = \mathrm{Tr}\, (\theta_B \otimes \psi_{\hat{A}})T_{B\hat{A}} \leq 2^{-k}, \tag{10.3}$$

and

$$\begin{aligned}\sum_{j=1}^{N}\Omega(j) &= \sum_{j=1}^{N}\Lambda^{-1/2}\Lambda(j)\Lambda^{-1/2}\\ &= \Lambda^{-1/2}\left(\sum_{j=1}^{N}\Lambda(j)\right)\Lambda^{-1/2}\\ &= \Lambda^{-1/2}\Lambda\Lambda^{-1/2}\\ &= \Pi_{\Lambda},\end{aligned}$$

where Π_Λ is the projection onto the support of Λ. Define,

$$\Omega(N+1) := \mathbb{I} - \Pi_\Lambda.$$

Hence we have a POVM $\{\Omega(j) \mid j \in [N+1]\}$ since

$$\sum_{j=1}^{N+1}\Omega(j) = \mathbb{I}.$$

10.2.3 Error analysis

We have the following useful fact.

Fact 10.2 (Hayashi-Nagaoka)**.** *Let* $0 \leq S \leq \mathbb{I}$ *and* $T \geq 0$*. Then*

$$\mathbb{I} - (S+T)^{-1/2}S(S+T)^{-1/2} \leq 2(\mathbb{I}-S) + 4T.$$

Define $S := \Lambda(m)$ and $T := \Lambda - \Lambda(m)$. Note that

$$\Omega(m) = (S+T)^{-1/2}S(S+T)^{-1/2}.$$

Let $\hat{M}$ denote the output random variable. The probability of incorrect decoding is given by,

$$\begin{aligned}\Pr[\hat{M} \neq m] &= \mathrm{Tr}\left((\mathbb{I}-\Omega(m))\,\tau^{m}_{B\hat{A}_1\ldots\hat{A}_N}\right)\\ &= \mathrm{Tr}\left((\mathbb{I}-(S+T)^{-1/2}S(S+T)^{-1/2})\,\tau^{m}_{B\hat{A}_1\ldots\hat{A}_N}\right)\\ &\leq \mathrm{Tr}\left((2(\mathbb{I}-S)+4T)\,\tau^{m}_{B\hat{A}_1\ldots\hat{A}_N}\right) && \text{(Fact 10.2 and Fact 6.6)}\\ &= 2\mathrm{Tr}\left((\mathbb{I}-\Lambda(m))\,\tau^{m}_{B\hat{A}_1\ldots\hat{A}_N}\right) + 4\sum_{i\neq m}\mathrm{Tr}\left(\Lambda(i)\,\tau^{m}_{B\hat{A}_1\ldots\hat{A}_N}\right)\\ &\leq 2\varepsilon + 4\cdot N\cdot 2^{-k} && \text{(Eq. (10.2) and Eq. (10.3))}\\ &\leq 6\varepsilon. && \text{(Eq. (10.1))}\end{aligned}$$

10.3 Classical-quantum channel

We consider a *shared-randomness* assisted encoding where information is transmitted through a *classical-quantum* (c-q) channel. Let $\varepsilon > 0$ and define,

$$R := \max_{p_{A\hat{A}}}\left\{\mathrm{D}^{\varepsilon}_{H}(\theta_{B\hat{A}}\|\theta_B\otimes p_{\hat{A}}) - \log\frac{1}{\varepsilon} \;\middle|\; \theta_{B\hat{A}} := \mathcal{N}_{A\to B}(p_{A\hat{A}})\right\}.$$

Let $p_{A\hat{A}}$ be the probability distribution that achieves the maximum above. Let,

$$k := \mathrm{D}_H^\varepsilon(\theta_{B\hat{A}} \| \theta_B \otimes p_{\hat{A}}) \quad ; \quad N := \lfloor 2^R \rfloor = \lfloor \varepsilon \cdot 2^k \rfloor.$$

The communication protocol is described as follows.

1. Alice and Bob start with $[N]$ i.i.d. copies of $p_{A\hat{A}}$ shared between them.
2. On receiving input $m \in [N]$, Alice inserts the register A of the mth copy of $p_{A\hat{A}}$ into the channel.
3. After receiving channel's output, Bob measures the state at his end $\tau^m_{B\hat{A}_1 \ldots \hat{A}_N}$ according to the POVM

   $$\{\Omega_j \mid j \in [N+1]\},$$

   as described below.
4. Bob outputs $\hat{m}$, which is the outcome of the measurement above.

10.3.1 Bob's decoding measurement and error analysis

Let $T_{B\hat{A}}$ be the operator that achieves the maximum in the definition of $\mathrm{D}_H^\varepsilon(\theta_{B\hat{A}} \| \theta_B \otimes p_{\hat{A}})$. The rest of the description of the measurement is the same as before. Also the error analysis is the same as before.

10.4 Classical-classical channel

We consider a shared-randomness assisted encoding where information is transmitted through a *classical-classical* (c-c) channel. Let $\varepsilon > 0$ and define,

$$R := \max_{p_{A\hat{A}}} \left\{ \mathrm{D}_H^\varepsilon(q_{B\hat{A}} \| q_B \otimes p_{\hat{A}}) - \log \frac{1}{\varepsilon} \quad \middle| \quad q_{B\hat{A}} := \mathcal{N}_{A\to B}(p_{A\hat{A}}) \right\}.$$

Let $p_{A\hat{A}}$ be the probability distribution that achieves the maximum above. Let,

$$k := \mathrm{D}_H^\varepsilon(q_{B\hat{A}} \| q_B \otimes p_{\hat{A}}) \quad ; \quad N := \lfloor 2^R \rfloor = \lfloor \varepsilon \cdot 2^k \rfloor.$$

The communication protocol is described as follows.

1. Alice and Bob start with $[N]$ i.i.d. copies of $p_{A\hat{A}}$ shared between them.
2. On receiving input $m \in [N]$, Alice inserts the register A of the mth copy of $p_{A\hat{A}}$ into the channel.
3. After receiving channel's output, Bob measures the random variable at his end $\tau^m_{B\hat{A}_1 \ldots \hat{A}_N}$ according to the POVM

   $$\{\Omega_j \mid j \in [N+1]\},$$

   as described below.
4. Bob outputs $\hat{m}$, which is the outcome of the measurement above.

10.4.1 Bob's decoding measurement and error analysis

Let $T_{B\hat{A}}$ be the operator that achieves the maximum in the definition of $\mathrm{D}^{\epsilon}_{H}(q_{B\hat{A}} \| q_B \otimes p_{\hat{A}})$. The rest of the description of the measurement is the same as before. Also the error analysis is the same as before.

10.5 Randomness unassisted coding for c-q and c-c channels for uniformly random input

Let M, drawn uniformly from $[N]$, denote the input random variable. Let S denote the randomness used in the protocol. Let $\mathrm{err}(m, s)$ denote the error of the protocol on input m and randomness s.

From the arguments in previous sections, we have for all m:

$$\mathbb{E}_{s \leftarrow S}[\mathrm{err}(m, s)] \leq 6\varepsilon.$$

This implies,

$$\mathbb{E}_{s \leftarrow S}[\mathbb{E}_{m \leftarrow M}[\mathrm{err}(m, s)]] = \mathbb{E}_{m \leftarrow M}[\mathbb{E}_{s \leftarrow S}[\mathrm{err}(m, s)]] \leq 6\varepsilon.$$

Therefore,

$$\exists s : \quad \mathbb{E}_{m \leftarrow M}[\mathrm{err}(m, s)] \leq 6\varepsilon.$$

On fixing the randomness to s we get a randomness unassisted protocol with average error (for uniform input) at most ε.

11
Channel coding converse bound and asymptotic achievability

In the previous lecture, we looked at the achievability of quantum-quantum (q-q), classical-quantum (c-q), and classical-classical (c-c) channel coding in the one-shot setting. In this lecture, we first look at the converse bound of channel coding in the one-shot setting before moving on to the asymptotic analysis of channel coding [1].

[1] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. On the near-optimality of one-shot classical communication over quantum channels. *Journal of Mathematical Physics*, 60(1):012204, 01 2019a. DOI: 10.1063/1.5039796. URL https://doi.org/10.1063/1.5039796

11.1 Converse of channel coding

First, let's look at a general entanglement-assisted coding protocol as seen below in Figure 11.1 for message M uniformly drawn from $[N]$. Note that the Encode, $\mathcal{N}$ and Decode blocks are CPTP but not necessarily unitary.

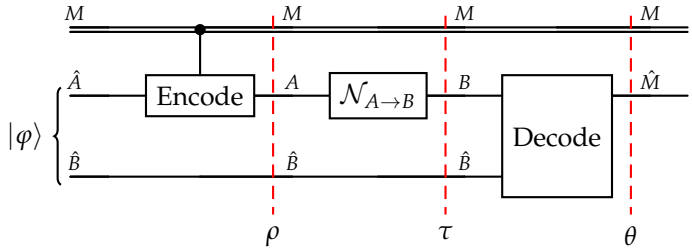

Figure 11.1: A general entanglement-assisted coding protocol.

With figure as a reference, let $|\rho\rangle_{MA\hat{B}T}$ be a purification of $\rho_{MA\hat{B}}$. Define,

$$\Lambda_{M\hat{M}} \stackrel{\text{def}}{=} \sum_{m=1}^{N} |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{\hat{M}}.$$

We can see that this is effectively checking the probability of $M = \hat{M}$. From the correctness of the protocol,

$$\mathrm{Tr}\Lambda_{M\hat{M}}\theta_{M\hat{M}} = \Pr[M = \hat{M}]_{\theta_{M\hat{M}}} \geq 1 - \varepsilon.$$

Define $\gamma_{\hat{M}} := \mathrm{Decode}(\tau_{\hat{B}} \otimes \tau_B)$. Note,

$$\mathrm{Tr}\Lambda_{M\hat{M}}\left(\theta_M \otimes \gamma_{\hat{M}}\right) = \Pr[M = \hat{M}]_{\theta_M \otimes \gamma_{\hat{M}}} = \frac{1}{N}.$$

From the definition of the hypothesis testing relative-entropy (Definition 10.1) we get,

$$\mathrm{D}_H^\varepsilon(\theta_{M\hat{M}}\|\theta_M\otimes\gamma_{\hat{M}})\geq\log N. \tag{11.1}$$

Consider,

$$\begin{aligned}
&\max_{|\psi\rangle_{\tilde{A}A}}\mathrm{D}_H^\varepsilon(\mathcal{N}_{A\to B}\left(\psi_{\tilde{A}A}\right)\|\psi_{\tilde{A}}\otimes\mathcal{N}_{A\to B}\left(\psi_A\right))\\
&\geq\mathrm{D}_H^\varepsilon(\mathcal{N}_{A\to B}(\rho_{M\hat{B}TA})\|\rho_{M\hat{B}T}\otimes\mathcal{N}_{A\to B}\left(\rho_A\right)) && (\text{taking }\tilde{A}=M\hat{B}T)\\
&=\mathrm{D}_H^\varepsilon(\tau_{M\hat{B}TB}\|\rho_{M\hat{B}T}\otimes\tau_B)\\
&\geq\mathrm{D}_H^\varepsilon(\tau_{M\hat{B}B}\|\rho_{M\hat{B}}\otimes\tau_B) && (\text{DPI})\\
&=\mathrm{D}_H^\varepsilon(\tau_{M\hat{B}B}\|\rho_M\otimes\rho_{\hat{B}}\otimes\tau_B) && (\rho_{M\hat{B}}=\rho_M\otimes\rho_{\hat{B}})\\
&=\mathrm{D}_H^\varepsilon(\tau_{M\hat{B}B}\|\rho_M\otimes\tau_{\hat{B}}\otimes\tau_B) && (\rho_{\hat{B}}=\tau_{\hat{B}})\\
&\geq\mathrm{D}_H^\varepsilon(\theta_{M\hat{M}}\|\theta_M\otimes\gamma_{\hat{M}}) && (\text{DPI and }\theta_M=\rho_M))\\
&\geq\log N. && (\textbf{??})
\end{aligned}$$

Let's compare this with the achievability bound that we found in the previous lecture.

$$\log N\geq\max_{|\psi\rangle_{\tilde{A}A}}\mathrm{D}_H^{\varepsilon/6}(\mathcal{N}_{A\to B}\left(\psi_{\hat{A}A}\right)\|\psi_{\tilde{A}}\otimes\mathcal{N}_{A\to B}\left(\psi_A\right))-\log\left(\frac{6}{\varepsilon}\right).$$

11.2 Asymptotic achievability

Suppose we have n uses of the channel available, that is $\mathcal{N}_{A^n\to B^n}^{\otimes n}$, and we wish to transmit message $M\in[N]$ using these channels with error at most ε. We use the following definition and Facts.

Definition 11.1 (Smooth max relative-entropy)**.** *Let ρ,σ be states and $\varepsilon>0$. The ε-smooth max relative-entropy between ρ and σ is defined as,*

$$\mathrm{D}_{\max}^\varepsilon(\rho\|\sigma)\stackrel{def}{=}\inf_{\rho'\approx_\varepsilon\rho}\mathrm{D}_{\max}(\rho'\|\sigma).$$

Fact 11.2 (Asymptotic convergence)**.** *Let ρ,σ be quantum states. Then*

$$\lim_{\varepsilon\to0}\lim_{n\to\infty}\frac{1}{n}\mathrm{D}_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})=\lim_{\varepsilon\to0}\lim_{n\to\infty}\frac{1}{n}\mathrm{D}_{\max}^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})=\mathrm{D}(\rho\|\sigma).$$

Fact 11.3. *Let ρ_{AB} be a state. Then,*

$$\min_{\sigma_A,\sigma_B}\mathrm{D}(\rho_{AB}\|\sigma_A\otimes\sigma_B)=\mathrm{D}(\rho_{AB}\|\rho_A\otimes\rho_B)=\mathrm{I}(A:B)_{\rho_{AB}}.$$

Using the achievability protocol using PBD as seen before we have,

$$\begin{aligned}
&\frac{\log N}{n} \\
&\geq \frac{1}{n}\left[\max_{|\psi\rangle_{\hat{A}^n A^n}} \mathrm{D}_H^{\varepsilon/6}(\mathcal{N}_{A^n\to B^n}^{\otimes n}\left(\psi_{\hat{A}^n A^n}\right) \| \psi_{\hat{A}^n} \otimes \mathcal{N}_{A^n\to B^n}^{\otimes n}\left(\psi_{A^n}\right)) - \log\left(\frac{6}{\varepsilon}\right)\right] \\
&\geq \frac{1}{n}\left[\max_{|\psi\rangle_{\hat{A}A}^{\otimes n}} \mathrm{D}_H^{\varepsilon/6}(\mathcal{N}_{A^n\to B^n}^{\otimes n}\left(\psi_{\hat{A}A}^{\otimes n}\right) \| \psi_{\hat{A}}^{\otimes n} \otimes \mathcal{N}_{A^n\to B^n}^{\otimes n}\left(\psi_A^{\otimes n}\right)) - \log\left(\frac{6}{\varepsilon}\right)\right] \\
&= \max_{|\psi\rangle_{\hat{A}A}} \frac{1}{n}\left[\mathrm{D}_H^{\varepsilon/6}(\left(\mathcal{N}_{A\to B}\left(\psi_{\hat{A}A}\right)\right)^{\otimes n} \| \psi_{\hat{A}}^{\otimes n} \otimes \left(\mathcal{N}_{A\to B}\left(\psi_A\right)\right)^{\otimes n}) - \log\left(\frac{6}{\varepsilon}\right)\right]
\end{aligned}$$

Taking $\lim_{\varepsilon\to 0}$ and $\lim_{n\to\infty}$ above we get (note the second term goes to 0),

$$\begin{aligned}
&\max_{|\psi\rangle_{\hat{A}A}} \mathrm{D}(\mathcal{N}_{A\to B}\left(\psi_{\hat{A}A}\right) \| \psi_{\hat{A}} \otimes \mathcal{N}_{A\to B}\left(\psi_A\right)) && \text{(Fact 11.2)} \\
&= \max_{|\psi\rangle_{A\hat{A}}} \mathrm{I}(\hat{A} : B)_{\mathcal{N}_{A\to B}(\psi_{A\hat{A}})} && \text{(Fact 11.3)} \\
&\stackrel{\text{def}}{=} \mathrm{cap}(\mathcal{N}_{A\to B}).
\end{aligned}$$

Above, $\mathrm{cap}(\mathcal{N}_{A\to B})$ is the *entanglement-assisted classical capacity* of the channel $\mathcal{N}_{A\to B}$.

In the c-q and c-c case, we get the classical channel capacity,

$$\mathrm{cap}(\mathcal{N}_{A\to B}) \stackrel{\text{def}}{=} \max_{p_{A\hat{A}}} \mathrm{I}(\hat{A} : B)_{\mathcal{N}_{A\to B}(p_{A\hat{A}})}.$$

In the c-c case the above becomes (will be discussed in a weekly assignment) the more familiar form of the *Shannon capacity* of classical channels,

$$\mathrm{cap}(\mathcal{N}_{A\to B}) \stackrel{\text{def}}{=} \max_{X\leftarrow p_A \,;\, Y\leftarrow\mathcal{N}_{A\to B}(X)} \mathrm{I}(X : Y).$$

12
Channel-coding converse bound in the asymptotic limit

We study the converse bound for channel-coding in the asymptotic limit [1].

[1] Sumeet Khatri, Ludovico Lami, and Mark M. Wilde. *Principles of Quantum Communication Theory: A Modern Approach*. 2025. URL https://www.markwilde.com/PQCT-khatri-lami-wilde.pdf

12.1 Main theorem and proof

The setting is described in Figure 12.1. The channel $\mathcal{N}_{A^n\to B^n}^{\otimes n}$ is n independent uses of the channel $\mathcal{N}_{A\to B}$. Note that the output of Alice's encoding need not be a tensor product state.

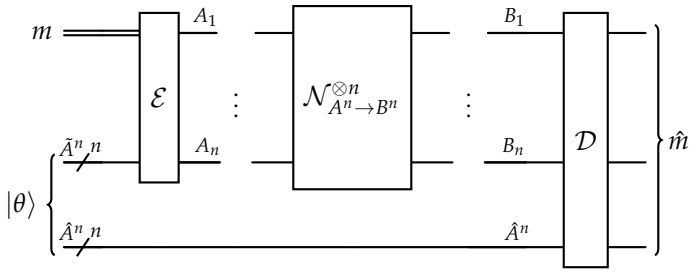

Figure 12.1: An entanglement-assisted protocol over multiple uses of a quantum channel $\mathcal{N}_{A\to B}$. Alice and Bob possess entanglement as a resource in the form of a quantum state $|\theta\rangle$ on the systems $\tilde{A}^n\hat{A}^n$. Alice encodes ($\mathcal{E}$) the message $m \in \{0,1\}^k$ into a quantum state on n quantum systems $A_1 \dots A_n$. Each system A_i is sent through the (same) channel $\mathcal{N}_{A\to B}$ to Bob's side. Bob performs the relevant measurements (decodes, $\mathcal{D}$) and obtains an estimate $\hat{m}$ of the message m sent by Alice.

The following theorem captures the asymptotic limit of the rate of classical information transfer using $\mathcal{N}_{A^n\to B^n}^{\otimes n}$.

Theorem 12.1.

$$\lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \frac{1}{n} \sup_{|\psi\rangle_{\hat{A}^n A^n}} \mathrm{D}_H^{\varepsilon}(\mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{\hat{A}^n A^n}) \| \psi_{\hat{A}^n} \otimes \mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{A^n})) = \mathrm{cap}(\mathcal{N}_{A\to B}),$$

where

$$\mathrm{cap}(\mathcal{N}_{A\to B}) \stackrel{def}{=} \sup_{|\psi\rangle_{A\hat{A}}} \mathrm{I}(B:\hat{A})_{\mathcal{N}_{A\to B}(\psi_{A\hat{A}})}$$

is the entanglement-assisted classical capacity of the channel $\mathcal{N}_{A\to B}$.

Proof. We show the equality by showing the lower and upper bounds.

1. **Lower Bound.**

Consider,

$$\begin{aligned}
&\lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \frac{1}{n} \sup_{|\psi\rangle_{\hat{A}^n A^n}} \mathrm{D}_H^\varepsilon(\mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{\hat{A}^n A^n}) \| \psi_{\hat{A}^n} \otimes \mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{A^n})) \\
&\geq \lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \frac{1}{n} \left[\sup_{|\psi\rangle_{\hat{A}A}} \mathrm{D}_H^\varepsilon(\mathcal{N}_{A^n\to B^n}^{\otimes n}\left(\psi_{\hat{A}A}^{\otimes n}\right) \| \psi_{\hat{A}}^{\otimes n} \otimes \mathcal{N}_{A^n\to B^n}^{\otimes n}\left(\psi_A^{\otimes n}\right)) \right] \\
&= \sup_{|\psi\rangle_{\hat{A}A}} \lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \frac{1}{n} \left[\mathrm{D}_H^\varepsilon(\left(\mathcal{N}_{A\to B}\left(\psi_{\hat{A}A}\right)\right)^{\otimes n} \| \psi_{\hat{A}}^{\otimes n} \otimes \left(\mathcal{N}_{A\to B}\left(\psi_A\right)\right)^{\otimes n}) \right] \\
&= \sup_{|\psi\rangle_{\hat{A}A}} \mathrm{D}(\mathcal{N}_{A\to B}\left(\psi_{\hat{A}A}\right) \| \psi_{\hat{A}} \otimes \mathcal{N}_{A\to B}\left(\psi_A\right)) && \text{(Fact 11.2)} \\
&= \sup_{|\psi\rangle_{A\hat{A}}} \mathrm{I}(\hat{A} : B)_{\mathcal{N}_{A\to B}(\psi_{A\hat{A}})} && \text{(Fact 11.3)} \\
&= \mathrm{cap}(\mathcal{N}_{A\to B}).
\end{aligned}$$

2. **Upper Bound.**

We need the following Fact relating the hypothesis-testing relative entropy and the relative entropy.

Fact 12.2. *Let ρ, σ be quantum states and $\varepsilon > 0$. We have*

$$(1-\varepsilon)\mathrm{D}_H^\varepsilon(\rho\|\sigma) \leq \mathrm{D}(\rho\|\sigma) + \mathrm{H}(\varepsilon).$$

Here

$$\mathrm{H}(x) \stackrel{def}{=} x\log(1/x) + (1-x)\log(1/(1-x))$$

is the binary entropy function.

Consider,

$$\begin{aligned}
&(1-\varepsilon) \sup_{|\psi\rangle_{\hat{A}^n A^n}} \mathrm{D}_H^\varepsilon(\mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{\hat{A}^n A^n}) \| \psi_{\hat{A}^n} \otimes \mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{A^n})) \\
&\leq \sup_{|\psi\rangle_{\hat{A}^n A^n}} \mathrm{D}(\mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{\hat{A}^n A^n}) \| \psi_{\hat{A}^n} \otimes \mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{A^n})) + \mathrm{H}(\varepsilon) && \text{(Fact 12.2)} \\
&= \sup_{|\psi\rangle_{\hat{A}^n A^n}} \mathrm{D}(\tau_{\hat{A}^n B^n} \| \tau_{\hat{A}^n} \otimes \tau_{B^n}) + \mathrm{H}(\varepsilon) && (\tau_{\hat{A}^n B^n} \stackrel{\text{def}}{=} \mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{\hat{A}^n A^n})) \\
&= \sup_{\psi_{\hat{A}^n A^n}} \mathrm{I}(\hat{A}^n : B^n)_\tau + \mathrm{H}(\varepsilon) && \text{(Fact 11.3)} \\
&\leq n \cdot \mathrm{cap}(\mathcal{N}_{A\to B}) + \mathrm{H}(\varepsilon). && \text{(Claim 1)}
\end{aligned}$$

This implies,

$$\begin{aligned}
&\lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \frac{1}{n} \cdot \sup_{|\psi\rangle_{\hat{A}^n A^n}} \mathrm{D}_H^\varepsilon(\mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{\hat{A}^n A^n}) \| \psi_{\hat{A}^n} \otimes \mathcal{N}_{A^n\to B^n}^{\otimes n}(\psi_{A^n})) \\
&\leq \lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \frac{1}{(1-\varepsilon)} \cdot \mathrm{cap}(\mathcal{N}_{A\to B}) + \frac{1}{n(1-\varepsilon)} \cdot \mathrm{H}(\varepsilon) \\
&= \mathrm{cap}(\mathcal{N}_{A\to B}).
\end{aligned}$$

□

Claim 1.

$$\sup_{\psi_{\hat{A}^n A^n}} \mathrm{I}(\hat{A}^n : B^n)_{\mathcal{N}^{\otimes n}_{A^n \to B^n}(\psi_{\hat{A}^n A^n})} \leq n \cdot \mathrm{cap}(\mathcal{N}_{A\to B}).$$

Proof. We show the result for $n = 2$ case. The general case follows similarly. Refer to Figure 12.2.

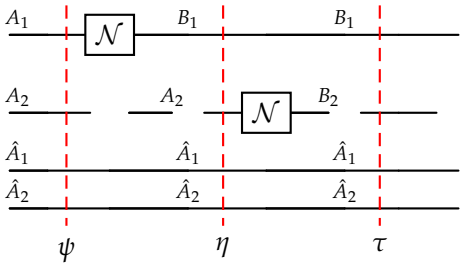

Figure 12.2: $\mathcal{N} \otimes \mathcal{N} = (\mathcal{I} \otimes \mathcal{N})(\mathcal{N} \otimes \mathcal{I})$.

We need the following Facts.

Fact 12.3 (Chain-rule for mutual information)**.** *For a state ρ_{ABC},*

$$\mathrm{I}(A : BC)_\rho = \mathrm{I}(A : B)_\rho + \mathrm{I}(A : C \mid B)_\rho.$$

Fact 12.4. *For a state ρ_{AB},*

$$\mathrm{I}(A : B)_\rho \geq 0.$$

Fact 12.5.

$$\mathrm{cap}(\mathcal{N}_{A\to B}) = \sup_{\rho_{A\hat{A}}} \mathrm{I}(B : \hat{A})_{\mathcal{N}_{A\to B}(\rho_{A\hat{A}})}$$

We have,

$$\begin{aligned}
&\sup_{\psi} \mathrm{I}(\hat{A}_1\hat{A}_2 : B_1B_2)_\tau \\
&= \sup_{\psi} \left(\mathrm{I}(\hat{A}_1\hat{A}_2 : B_1)_\tau + \mathrm{I}(\hat{A}_1\hat{A}_2 : B_2 \mid B_1)_\tau\right) && \text{(Fact 12.3)} \\
&= \sup_{\psi} \left(\mathrm{I}(\hat{A}_1\hat{A}_2 : B_1)_\tau + \mathrm{I}(\hat{A}_1\hat{A}_2B_1 : B_2)_\tau - \mathrm{I}(B_1 : B_2)_\tau\right) && \text{(Fact 12.3)} \\
&\leq \sup_{\psi} \left(\mathrm{I}(\hat{A}_1\hat{A}_2 : B_1)_\tau + \mathrm{I}(\hat{A}_1\hat{A}_2B_1 : B_2)_\tau\right) && \text{(Fact 12.4)} \\
&\leq \sup_{\psi} \mathrm{I}(\hat{A}_1\hat{A}_2 : B_1)_\eta + \sup_{\eta} \mathrm{I}(\hat{A}_1\hat{A}_2B_1 : B_2)_\tau \\
&\leq \sup_{\theta} \mathrm{I}(\tilde{A}_1 : B_1)_{\mathcal{N}_{A_1\to B_1}(\theta)} + \sup_{\theta} \mathrm{I}(\tilde{A}_2 : B_2)_{\mathcal{N}_{A_2\to B_2}(\theta)} \\
&= 2 \cdot \mathrm{cap}(\mathcal{N}_{A\to B}). && \text{(Fact 12.5)}
\end{aligned}$$

□

13
Quantum state redistribution

Quantum state redistribution is the generalization of state splitting and state merging that we have seen so far in the one-shot setting. The protocol starts with pure state $|\varphi\rangle_{RABC}$ shared between three parties where the Referee holds the register R, Alice holds the register A and C, and Bob holds the register B.

The goal of the protocol is to transfer the register C to Bob using the prior entangled state $|\theta\rangle$ in the registers $\hat{A}$ and $\hat{B}$ such that the output state ψ_{RABC} satisfies $\psi_{RABC} \approx_\varepsilon \varphi_{RABC}$ (as shown in Figure 13.1).

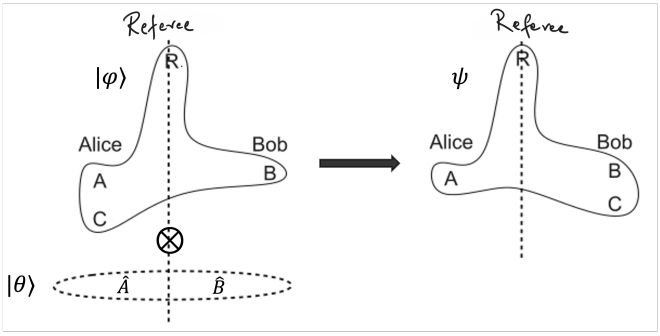

Figure 13.1: State redistribution problem statement where the initial state φ_{RABC} is shared between Alice and Bob. They follow an entanglement-assisted protocol to transfer the register C to Bob. The output state ψ_{RABC} is close to the initial state within purified distance ε.

13.1 Protocol for quantum state redistribution

Here we present a protocol for state redistribution. [1] The initial state and setup are very similar to that of the state splitting protocol. However, now Bob can also make use of the register B which helps Alice compress the communication required from her for Bob to correctly identify the state.

[1] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. A one-shot achievability result for quantum state redistribution. *IEEE Transactions on Information Theory*, 64(3):1425–1435, 2018. DOI: 10.1109/TIT.2017.2776112

13.1.1 Initial state and setup

Let $\varepsilon \geq 0, \delta > 0$. The initial state $|\varphi\rangle_{RABC}$ is shared between Referee (R), Alice (AC) and Bob (B). Let $\varphi'_{RBC} \approx_\varepsilon \varphi_{RBC}$ (with $\varphi_{RB} = \varphi'_{RB}$) and σ_C be states. Let

$$\log n := \mathrm{D}_{\max}(\varphi'_{RBC} \| \varphi_{RB} \otimes \sigma_C) + 2 \log \frac{1}{\delta}.$$

Alice and Bob shares n copies of a purification of the quantum state σ_C, in the registers $\hat{C}_i$ and C_i for $i \in \{1, 2, ..., n\}$.

Let $|\varphi'\rangle_{RABC}$ be a purification φ'_{RBC} (guaranteed by Uhlmann's theorem) such that

$$\mathrm{F}(|\varphi'\rangle\langle\varphi'|_{RABC}, |\varphi\rangle\langle\varphi|_{RABC}) = \mathrm{F}(\varphi'_{RBC}, \varphi_{RBC}).$$

13.1.2 Protocol steps

The protocol consists of the following steps.

1. Alice performs the Uhlmann isometry on her registers based on the convex split lemma (CSL). This creates a state (close to RHS of the figure) where each term in the superposition has amplitude $\frac{1}{\sqrt{n}}$ as shown in Figure 13.2.

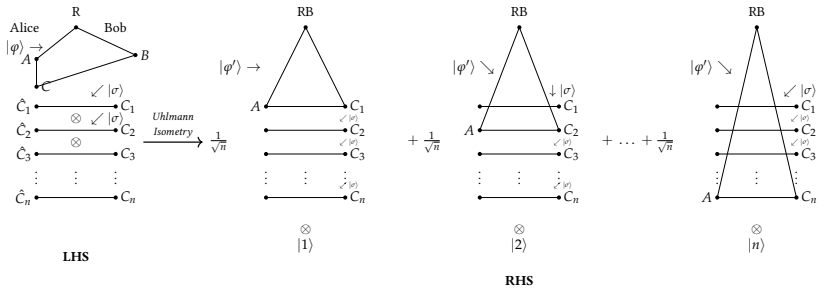

Figure 13.2: State redistribution protocol - Alice's operations: The left-hand side (LHS) shows the initial state with Alice holding registers A and C, and Alice and Bob sharing entangled pairs. The right-hand side (RHS) shows the state (that is close to the state after applying Uhlmann's isometry) which is a superposition with coefficients $1/\sqrt{n}$.

2. Alice measures her index register to find some j.

   If she could communicate j to Bob, Bob would be able to pick up the register C_j obtaining the desired state. The problem is that the number of bits required to communicate j is large (around $\mathrm{D}_{\max}(\varphi'_{RBC} \| \varphi_{RB} \otimes \sigma_C)$).

3. Alice divides $[n]$ into n/k blocks of size k and communicates the block number of j to Bob, where,

$$\log k := \mathrm{D}_H^{2\epsilon}(\varphi'_{BC} \| \varphi_B \otimes \sigma_C) - \log \frac{1}{\delta}.$$

4. Upon receiving the block number from Alice, the quantum state left with Bob is (close to) a convex mixture as shown in Figure 13.3.

   Bob makes use of his quantum side information (B), employing position-based decoding (PBD) strategy to find j inside the block.

13.1.3 Communication cost

We use the following Fact.

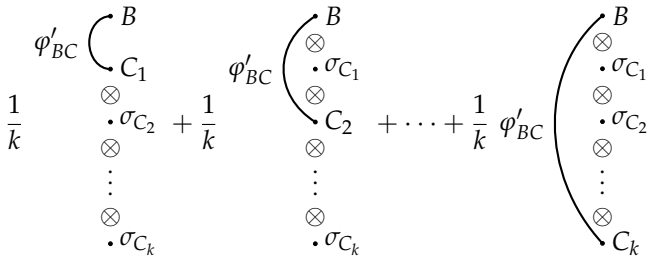

Figure 13.3: State redistribution protocol - Bob's operations: The state with Bob is (close to) a convex combination of states once Alice communicates the block number to Bob. Bob detects j inside the block of size k using PBD.

Fact 13.1. *Let $\varepsilon \geq \delta \geq 0$. Let $\rho \approx_\delta \rho'$ and σ be state. Then,*

$$\mathrm{D}_H^{\varepsilon-\delta}(\rho\|\sigma) \leq \mathrm{D}_H^{\varepsilon}(\rho'\|\sigma).$$

The communication cost is:

$$\begin{aligned}
&\log \frac{n}{k} = \log n - \log k \\
&= \min_{\sigma_C, \varphi'_{RBC} \approx_\varepsilon \varphi_{RBC}, \varphi_{RB} = \varphi'_{RB}} \left[\mathrm{D}_{\max}(\varphi'_{RBC}\|\varphi_{RB}\otimes\sigma_C) - \mathrm{D}_H^{2\varepsilon}(\varphi'_{BC}\|\varphi_B\otimes\sigma_C)\right] + 3\log\frac{1}{\delta} \\
&\leq \min_{\sigma_C, \varphi'_{RBC} \approx_\varepsilon \varphi_{RBC}, \varphi_{RB} = \varphi'_{RB}} \left[\mathrm{D}_{\max}(\varphi'_{RBC}\|\varphi_{RB}\otimes\sigma_C) - \mathrm{D}_H^{\varepsilon}(\varphi_{BC}\|\varphi_B\otimes\sigma_C)\right] + 3\log\frac{1}{\delta}.
\end{aligned}$$

The inequality above follows from Fact 13.1 and $\varphi_{BC} \approx_\varepsilon \varphi'_{BC}$.

13.1.4 Error analysis

Using CSL and Uhlmann's theorem, we can argue as before that at Step 2. in the protocol, when Alice measures the index register, Referee and Bob's combined state becomes (δ close to) a probabilistic mixture weighted by $\frac{1}{n}$ (refer to RHS of Figure 13.2).

When Alice sends the block number of her measurement outcome j to Bob, Bob knows which block contains the target state. The quantum state left with Bob is a convex mixture as shown in Figure 13.3. Bob employs PBD to find out j inside the block.

Let $\hat{\gamma}$ be the overall state produced in the protocol after Alice applies Uhlmann's isometry and measures the index register. Let γ be the state on the RHS of Figure 13.2, on measuring the index register. From CSL and Uhlmann's theorem, we know $\gamma \approx_\delta \hat{\gamma}$.

We use the following Fact [2].

Fact 13.2 (Accurate measurement lemma)**.** *Let*

$$\rho_{AB} = \sum_i p_i |i\rangle\langle i|_A \otimes \rho_B^i$$

[2] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. A one-shot achievability result for quantum state redistribution. *IEEE Transactions on Information Theory*, 64(3):1425–1435, 2018. DOI: 10.1109/TIT.2017.2776112

be a c-q state. Let $\mathcal{M}$ be a measurement performed on the register B to produce an outcome in the register $\hat{A}$. Let $\sigma_{AB\hat{A}}$ be the resulting state after the measurement. Then,

$$\mathrm{F}(\rho_{AB}, \sigma_{AB}) \geq (\Pr[\hat{A} = A]_{\sigma})^2.$$

From the arguments used in analysis of PBD we know that there is measurement that Bob can perform on the state γ in Figure 13.3 to output candidate $\hat{j}$ such that

$$\Pr[J = \hat{J}] \geq 1 - (4\varepsilon + 4\delta).$$

Let γ' be the resulting state after the measurement. From Fact 13.2 we get,

$$\mathrm{F}(\gamma', \gamma) \geq (1 - (4\varepsilon + 4\delta))^2 \geq 1 - (8\varepsilon + 8\delta).$$

This implies $\gamma' \approx_{\sqrt{8\varepsilon+8\delta}} \gamma$.

If Alice and Bob were using the state γ (and index J to output), the output state would have been exactly $|\varphi'\rangle_{RABC}$. If the output state $\hat{\varphi}_{RABC}$ was produced using $\hat{J}$ in γ', then using DPI we would have,

$$\hat{\varphi}_{RABC} \approx_{\sqrt{8\varepsilon+8\delta}} \varphi'_{RABC}$$

In the protocol the actual output ψ_{RABC} is produced using the PBD measurement on $\hat{\gamma}$ instead of γ. Since $\hat{\gamma} \approx_{\delta} \gamma$, using DPI we get,

$$\hat{\varphi}_{RABC} \approx_{\delta} \psi_{RABC}.$$

Using the triangle inequality for the purified distance we get,

$$\psi_{RABC} \approx_{\delta+\sqrt{8\varepsilon+8\delta}} \varphi'_{RABC}.$$

Since $\varphi_{RABC} \approx_{\varepsilon} \varphi'_{RABC}$, again using the triangle inequality we finally get (since $\varepsilon + \delta + \sqrt{8\varepsilon + 8\delta} \leq 4\sqrt{\varepsilon + \delta}$),

$$\psi_{RABC} \approx_{4\sqrt{\varepsilon+\delta}} \varphi_{RABC}.$$

Open question: Is it possible to get a protocol for state redistribution with error ε and communication that is a polynomial in $\mathrm{I}(R \;:\; C \mid B)$ and $\frac{1}{\varepsilon}$?

This is known in the classical case for the analogous task.

14

Converse bound and asymptotics for state redistribution

We will first establish the converse bound, both in the one shot case and the asymptotic limit [1]. Then we will recall the one shot achievability bound and look at its asymptotic limit.

[1] Mario Berta, Matthias Christandl, and Dave Touchette. Smooth entropy bounds on one-shot quantum state redistribution. *IEEE Transactions on Information Theory*, 62(3): 1425–1439, 2016

14.1 Circuit for state redistribution

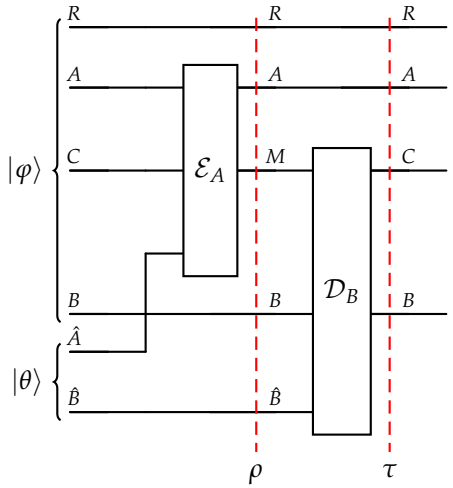

Figure 14.1: State redistribution circuit. Alice's registers are $A, C, \hat{A}$ and Bob's registers are $B, \hat{B}$. Both φ, θ are pure states.

The circuit diagram for a protocol for state redistribution is given above. Alice holds the registers $A, C, \hat{A}$ and Bob holds the registers $B, \hat{B}$ and there is the referee register R. The state $|\varphi\rangle_{RABC}$ is pure. The state $|\theta\rangle$ describing $\hat{A}, \hat{B}$ is also pure and it is taken to be pre-shared entanglement between the two parties. An encoding channel $\mathcal{E}_A$ works on $A, C, \hat{A}$ i.e. on Alice's registers and the outcome are the registers A, M. The quantum register M is then sent to Bob and Bob applies a decoding channel $\mathcal{D}_B$ on $M, B, \hat{B}$. The goal is that the state after the decoding, $\tau_{RACB} \approx_\varepsilon \varphi_{RACB}$, whilst minimizing the size of the register M to be communicated to Bob. We now intend to find a lower bound on $\log |M|$.

We need the following Fact.

Fact 14.1.

$$\mathrm{D}_{\max}(\rho_A\|\sigma_A) + 2\log|B| \geq \mathrm{D}_{\max}(\rho_{AB}\|\sigma_A \otimes \mu_B).$$

where $\mu_B := \mathbb{1}_B/|B|$.

14.2 A first bound

Let σ_B be the state such that,

$$\mathrm{I}_{\max}(R:B)_\varphi = \mathrm{D}_{\max}(\varphi_{RB}\|\varphi_R \otimes \sigma_B).$$

Consider,

$$\begin{aligned}
&\mathrm{I}_{\max}(R:B)_\varphi + 2\log|M| \\
&= \mathrm{D}_{\max}(\varphi_{RB}\|\varphi_R \otimes \sigma_B) + 2\log|M| \\
&= \mathrm{D}_{\max}(\varphi_{RB} \otimes \theta_{\hat{B}}\|\varphi_R \otimes \sigma_B \otimes \theta_{\hat{B}}) + 2\log|M| \\
&= \mathrm{D}_{\max}(\rho_{RB\hat{B}}\|\varphi_R \otimes \sigma_B \otimes \theta_{\hat{B}}) + 2\log|M| && (\rho_{RB\hat{B}} = \varphi_{RB} \otimes \theta_{\hat{B}}) \\
&\geq \mathrm{D}_{\max}(\rho_{RB\hat{B}M}\|\varphi_R \otimes \sigma_B \otimes \theta_{\hat{B}} \otimes \mu_M) && (\text{Fact 14.1 and } \mu_M := \mathbb{I}_M/|M|) \\
&\geq \mathrm{D}_{\max}(\tau_{RBC}\|\varphi_R \otimes \eta_{BC}) && (\text{DPI and } \eta_{BC} := \mathcal{D}_B(\sigma_B \otimes \theta_{\hat{B}} \otimes \mu_M)) \\
&\geq \mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{R}:BC)_\varphi. && (\tau_{RBC} \approx_\varepsilon \varphi_{RBC}\,;\ \tau_R = \varphi_R)
\end{aligned}$$

This implies,

$$\log|M| \geq \frac{\mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{R}:BC)_\varphi - \mathrm{I}_{\max}(R:B)_\varphi}{2}.$$

14.3 Strengthening the bound

One of the terms in the bound is a smoothed max-information and the other a non-smoothed version. It would be good to have only smoothened versions - and this would also mean strengthening the inequality in this case.

Claim 2.

$$2\log|M| \geq \mathrm{I}^{2\varepsilon}_{\max}(\overset{\bullet}{R}:BC)_\varphi - \mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{R}:B)_\varphi.$$

Proof. Let σ_B and $\tilde{\varphi}_{RB} \approx_\varepsilon \varphi_{RB}$ (with $\tilde{\varphi}_R = \varphi_R$) be the states such that,

$$\mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{R}:B)_\varphi = \mathrm{D}_{\max}(\tilde{\varphi}_{RB}\|\varphi_R \otimes \sigma_B).$$

Let $\tilde{\varphi}_{RBAC}$ be a purification of $\tilde{\varphi}_{RB}$ guaranteed by Uhlmann's theorem such that $\tilde{\varphi}_{RBAC} \approx_\varepsilon \varphi_{RBAC}$. Imagine that the circuit in Figure 14.1 starts with $\tilde{\varphi}_{RBAC}$; that is Figure 14.2.

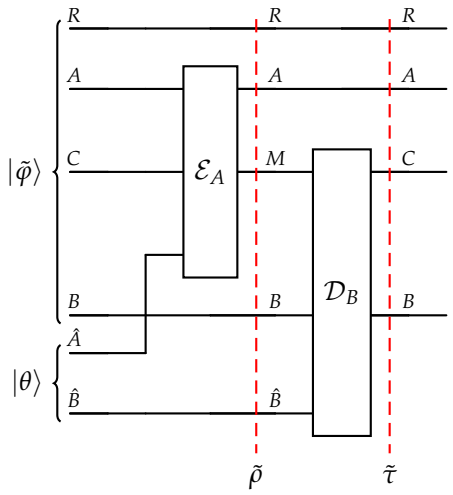

Figure 14.2: Circuit for state redistribution with a close-by starting state.

Consider,

$$
\begin{aligned}
&\mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{R} : B)_{\varphi} + 2\log|M| \\
&= \mathrm{D}_{\max}(\tilde{\varphi}_{RB} \| \varphi_R \otimes \sigma_B) + 2\log|M| \\
&= \mathrm{D}_{\max}(\tilde{\varphi}_{RB} \otimes \theta_{\hat{B}} \| \varphi_R \otimes \sigma_B \otimes \theta_{\hat{B}}) + 2\log|M| \\
&= \mathrm{D}_{\max}(\tilde{\rho}_{RB\hat{B}} \| \varphi_R \otimes \sigma_B \otimes \theta_{\hat{B}}) + 2\log|M| && (\tilde{\rho}_{RB\hat{B}} = \tilde{\varphi}_{RB} \otimes \theta_{\hat{B}}) \\
&\geq \mathrm{D}_{\max}(\tilde{\rho}_{RB\hat{B}M} \| \varphi_R \otimes \sigma_B \otimes \theta_{\hat{B}} \otimes \mu_M) && (\text{Fact 14.1 and } \mu_M := \mathbb{I}_M / |M|) \\
&\geq \mathrm{D}_{\max}(\tilde{\tau}_{RBC} \| \varphi_R \otimes \eta_{BC}) && (\text{DPI and } \eta_{BC} := \mathcal{D}_B(\sigma_B \otimes \theta_{\hat{B}} \otimes \mu_M)) \\
&\geq \mathrm{I}^{2\varepsilon}_{\max}(\overset{\bullet}{R} : BC)_{\varphi}.
\end{aligned}
$$

The last inequality follows since $\tilde{\varphi}_{RBC} \approx_{\varepsilon} \varphi_{RBC}$, we have using DPI, $\tilde{\tau}_{RBC} \approx_{\varepsilon} \tau_{RBC}$. Since $\varphi_{RBC} \approx_{\varepsilon} \tau_{RBC}$, from the triangle inequality we have $\tilde{\tau}_{RBC} \approx_{2\varepsilon} \varphi_{RBC}$. Also $\tilde{\tau}_R = \tilde{\varphi}_R = \varphi_R$. □

Open question: Can we get tight achievability and converse bounds for state redistribution in the one-shot setting?

14.4 *Asymptotic limit of the converse bound*

Here we show the asymptotic limit of the one shot converse bound derived above [2]. This means we want to find a lower bound on the rate

$$R \overset{\text{def}}{=} \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\log |M_n|}{n}$$

of communication, where $\log |M_n|$ is the number of qubits that are needed to be communicated for a protocol with input $|\varphi\rangle^{\otimes n}_{RABC}$ and overall error ε.

We need the following definition and Facts.

[2] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. A one-shot achievability result for quantum state redistribution. *IEEE Transactions on Information Theory*, 64(3):1425–1435, 2018. DOI: 10.1109/TIT.2017.2776112

Fact 14.2.

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{A}^n : B^n)_{\rho^{\otimes n}} = \mathrm{I}(A:B)_\rho.$$

Consider,

$$\begin{aligned}
R &\overset{\text{def}}{=} \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{\log|M_n|}{n}\\
&\geq \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{2n}\left(\mathrm{I}^{2\varepsilon}_{\max}(\overset{\bullet}{R}^n : B^nC^n)_{\varphi^{\otimes n}} - \mathrm{I}^{\varepsilon}_{\max}(\overset{\bullet}{R}^n : B^n)_{\varphi^{\otimes n}}\right) && \text{(Claim 2)}\\
&= \frac{1}{2}\left(\mathrm{I}(R:BC)_\varphi - \mathrm{I}(R:B)_\varphi\right) && \text{(Fact 14.2)}\\
&\geq \frac{\mathrm{I}(R:C\mid B)_\varphi}{2}.
\end{aligned}$$

14.5 Asymptotic limit of the achievability bound

We need the following definition and Fact.

Definition 14.3 (Partially smooth max relative-entropy)**.** *Let ρ_{AB}, σ_{AB} be states and $\varepsilon > 0$. The ε-partially smooth max relative-entropy between ρ_{AB} and σ_{AB} is defined as,*

$$\mathrm{D}^{\varepsilon}_{\max}(\overset{\bullet}{\rho}_{AB}\|\sigma_{AB}) \overset{def}{=} \inf_{\substack{\rho'_{AB}\approx_\varepsilon\rho_{AB},\\ \rho_A=\rho'_A}} \mathrm{D}_{\max}(\rho'_{AB}\|\sigma_{AB}).$$

Fact 14.4 (Asymptotic convergence)**.** *Let ρ_{AB}, σ_{AB} be quantum states. Then*

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\mathrm{D}^{\varepsilon}_{\max}(\overset{\bullet}{\rho}^{\otimes n}_{AB}\|\sigma^{\otimes n}_{AB}) = \mathrm{D}(\rho_{AB}\|\sigma_{AB}).$$

Recall the one shot achievebility bound (with error $4\sqrt{2\varepsilon}$ by setting $\delta=\varepsilon$),

$$\begin{aligned}
&\log|M|\\
&\leq \min_{\sigma_C,\tilde\varphi_{RBC}\approx_\varepsilon\varphi_{RBC},\varphi_{RB}=\tilde\varphi_{RB}}\frac{1}{2}\left[\mathrm{D}_{\max}(\tilde\varphi_{RBC}\|\tilde\varphi_{RB}\otimes\sigma_C) - \mathrm{D}^{\varepsilon}_H(\varphi_{BC}\|\varphi_B\otimes\sigma_C) + 3\log\frac{1}{\varepsilon}\right].
\end{aligned}$$

In the asymptotic case with input $|\varphi\rangle^{\otimes n}_{RABC}$ and overall error $4\sqrt{2\varepsilon}$ we get,

$$\begin{aligned}
&\log|M_n|\\
&\leq \min_{\sigma_{C^n},\tilde\varphi_{R^nB^nC^n}\approx_\varepsilon\varphi^{\otimes n}_{RBC},\varphi^{\otimes n}_{RB}=\tilde\varphi_{R^nB^n}}\frac{1}{2}\left[\mathrm{D}_{\max}(\tilde\varphi_{R^nB^nC^n}\|\varphi^{\otimes n}_{RB}\otimes\sigma_{C^n}) - \mathrm{D}^{\varepsilon}_H(\varphi^{\otimes n}_{BC}\|\varphi^{\otimes n}_B\otimes\sigma_{C^n}) + 3\log\frac{1}{\varepsilon}\right]\\
&\leq \min_{\tilde\varphi_{R^nB^nC^n}\approx_\varepsilon\varphi^{\otimes n}_{RBC},\varphi^{\otimes n}_{RB}=\tilde\varphi_{R^nB^n}}\frac{1}{2}\left[\mathrm{D}_{\max}(\tilde\varphi_{R^nB^nC^n}\|\varphi^{\otimes n}_{RB}\otimes\varphi^{\otimes n}_C) - \mathrm{D}^{\varepsilon}_H(\varphi^{\otimes n}_{BC}\|\varphi^{\otimes n}_B\otimes\varphi^{\otimes n}_C) + 3\log\frac{1}{\varepsilon}\right]\\
&= \frac{1}{2}\left[\mathrm{D}^{\varepsilon}_{\max}(\overset{\bullet}{\varphi}^{\otimes n}_{RBC}\|\varphi^{\otimes n}_{RB}\otimes\varphi^{\otimes n}_C) - \mathrm{D}^{\varepsilon}_H(\varphi^{\otimes n}_{BC}\|\varphi^{\otimes n}_B\otimes\varphi^{\otimes n}_C) + 3\log\frac{1}{\varepsilon}\right].
\end{aligned}$$

Using Fact 14.4 and Fact 11.2, we get (note that the last term above goes to 0 in the limit),

$$\begin{aligned} R &\overset{\text{def}}{=} \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{\log|M_n|}{n} \\ &\leq \frac{1}{2}\left(\mathrm{D}(\varphi_{RBC}\|\varphi_{RB}\otimes\varphi_C) - \mathrm{D}(\varphi_{BC}\|\varphi_B\otimes\varphi_C)\right) \\ &= \frac{1}{2}\left(\mathrm{I}(RB:C)_\varphi - \mathrm{I}(B:C)_\varphi\right) \\ &= \frac{\mathrm{I}(R:C\mid B)_\varphi}{2}. \end{aligned}$$

Remark: From here we also get the asymptotic achievability and converse bounds for state splitting (and hence also state merging) by removing B, that is $\frac{\mathrm{I}(R:C)_\varphi}{2}$.

15
A compilation of achievability and converse bounds

15.1 Source coding

Here we list the quantum communication cost. Let $\varepsilon \geq 0, \delta > 0$.

15.1.1 State splitting, state merging

- Achievability
  - One-shot:

$$\text{with error } \varepsilon+\delta: \ \frac{1}{2}\left(\mathrm{I}_{\max}^{\varepsilon}(\overset{\bullet}{R}:C)_{\varphi}+2\log\left(\frac{1}{\delta}\right)\right),$$

$$\text{with error } 2\varepsilon+\delta: \ \frac{1}{2}\left(\mathrm{I}_{\max}^{\varepsilon}(R:C)_{\varphi}+2\log\left(\frac{1}{\delta}\right)\right).$$

  - Asymptotic:

$$\frac{1}{2}\mathrm{I}(R:C)_{\varphi}.$$

- Converse
  - One-shot (with error ε):

$$\frac{1}{2}\mathrm{I}_{\max}^{\varepsilon}(\overset{\bullet}{R}:C)_{\varphi}.$$

  - Asymptotic:

$$\frac{1}{2}\mathrm{I}(R:C)_{\varphi}.$$

15.1.2 State redistribution

- Achievability
  - One-shot (with error $4\sqrt{\varepsilon+\delta}$):

$$\frac{1}{2}\min_{\sigma_C,\varphi'_{RBC}\approx_{\varepsilon}\varphi_{RBC},\varphi'_{RB}=\varphi_{RB}}\left(\mathrm{D}_{\max}(\varphi'_{RBC}\|\varphi_{RB}\otimes\sigma_C)-\mathrm{D}_{H}^{\varepsilon}(\varphi_{BC}\|\varphi_B\otimes\sigma_C)+3\log\left(\frac{1}{\delta}\right)\right).$$

  - Asymptotic:
$$\frac{1}{2}\mathrm{I}(R:C\mid B)_{\varphi}.$$

- Converse
  - One-shot (with error ε):
$$\frac{1}{2}\left(\mathrm{I}_{\max}^{2\varepsilon}(\overset{\bullet}{R}:BC)_{\varphi}-\mathrm{I}_{\max}^{\varepsilon}(\overset{\bullet}{R}:B)_{\varphi}\right).$$
  - Asymptotic:
$$\frac{1}{2}\mathrm{I}(R:C\mid B)_{\varphi}.$$

15.2 Channel-coding

Let $\varepsilon>0$. Let $\mathcal{N}_{A\to B}$ be a q-q channel.

- Achievability
  - One-shot (with error 6ε):
$$\sup_{|\psi\rangle_{A\hat{A}}}\mathrm{D}_{H}^{\varepsilon}(\mathcal{N}_{A\to B}(\psi_{A\hat{A}})\|\mathcal{N}_{A\to B}(\psi_{A})\otimes\psi_{\hat{A}})-\log\left(\frac{1}{\varepsilon}\right).$$
  - Asymptotic:
$$\mathrm{cap}(\mathcal{N}_{A\to B}):=\sup_{|\psi\rangle_{A\hat{A}}}\mathrm{I}(B:\hat{A})_{\mathcal{N}_{A\to B}(\psi_{A\hat{A}})},$$
    $\mathrm{cap}(\mathcal{N}_{A\to B})$ is called the entanglement-assisted classical capacity.

- Converse
  - One-shot (with error ε):
$$\sup_{|\psi\rangle_{A\hat{A}}}\mathrm{D}_{H}^{\varepsilon}(\mathcal{N}_{A\to B}(\psi_{A\hat{A}})\|\mathcal{N}_{A\to B}(\psi_{A})\otimes\psi_{\hat{A}}).$$
  - Asymptotic:
$$\mathrm{cap}(\mathcal{N}_{A\to B}).$$

Similar for the c-q and c-c channels where quantum state $|\psi\rangle_{A\hat{A}}$ is replaced with probability distribution $p_{A\hat{A}}$ in sup.

16
The quantum substate theorem

16.1 Theorem statement

The *quantum substate theorem* [1] [2] upper bounds the smooth max relative-entropy in terms of the relative-entropy between two quantum states.

Theorem 16.1 (Substate theorem)**.** *Let ρ, σ be states such that* $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ *and* $\varepsilon > 0$*. Then*

$$\mathrm{D}_{\max}^{\sqrt{\varepsilon}}(\rho\|\sigma) \leq \frac{1}{\varepsilon}\left(\mathrm{D}(\rho\|\sigma) + 1\right) + \log\left(\frac{1}{1-\varepsilon}\right).$$

[1] Rahul Jain and Ashwin Nayak. Short proofs of the quantum substate theorem. *IEEE Transactions on Information Theory*, 58(6):3664–3669, 2012. DOI: 10.1109/TIT.2012.2184522

[2] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A property of quantum relative entropy with an application to privacy in quantum communication. *J. ACM*, 56(6), September 2009. ISSN 0004-5411. DOI: 10.1145/1568318.1568323. URL https://doi.org/10.1145/1568318.1568323

16.2 Observational divergence

Before getting into the proof, let us introduce a new information theoretic quantity, the *observational divergence* [3].

Definition 16.2 (Observational divergence)**.** *Let ρ, σ be states, such that* $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$*. The observational divergence between ρ and σ is defined as*

$$\mathrm{D}^{\mathrm{o}}(\rho\|\sigma) \stackrel{def}{=} \max\left\{ \mathrm{Tr}(T\rho)\log\frac{\mathrm{Tr}(T\rho)}{\mathrm{Tr}(T\sigma)} \;\middle|\; \begin{array}{l} 0 \leq T \leq \mathbb{1}, \\ \mathrm{Tr}(T\sigma) \neq 0 \end{array} \right\}.$$

[3] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A property of quantum relative entropy with an application to privacy in quantum communication. *J. ACM*, 56(6), September 2009. ISSN 0004-5411. DOI: 10.1145/1568318.1568323. URL https://doi.org/10.1145/1568318.1568323

The following claim shows that the observational divergence is upper bounded by the relative-entropy plus one.

Claim 3. *Let ρ, σ be states such that* $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$*. Then,*

$$\mathrm{D}^{\mathrm{o}}(\rho\|\sigma) \leq \mathrm{D}(\rho\|\sigma) + 1.$$

Proof. Denote T^* as the operator achieving the max in $\mathrm{D}^{\mathrm{o}}(\rho\|\sigma)$. Define the random variable P, Q to be the outcomes on measuring ρ and σ by the

POVM $\{\Pi_0 = T^*, \Pi_1 = \mathbb{1} - T^*\}$, respectively. Let $p \stackrel{\text{def}}{=} \mathrm{Tr}(T^*\rho)$ and $q \stackrel{\text{def}}{=} \mathrm{Tr}(T^*\sigma)$. Then,

$$\Pr[P=0] = p, \quad \Pr[P=1] = 1-p,$$
$$\Pr[Q=0] = q, \quad \Pr[Q=1] = 1-q,$$

and

$$\mathrm{D}^{\mathrm{o}}(\rho\|\sigma) = p\log\frac{p}{q}.$$

By the DPI for relative-entropy we have,

$$p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} = \mathrm{D}(P\|Q) \stackrel{\text{DPI}}{\leq} \mathrm{D}(\rho\|\sigma). \tag{16.1}$$

We know that the binary entropy $\mathrm{H}(p) \leq 1$, and hence,

$$(1-p)\log\frac{1}{1-p} \leq p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p} = \mathrm{H}(p) \leq 1.$$

Using above it follows,

$$(1-p)\log\frac{1-p}{1-q} \geq (1-p)\log(1-p) \geq -1. \tag{16.2}$$

Consider,

$$\begin{aligned}
\mathrm{D}^{\mathrm{o}}(\rho\|\sigma) - 1 &= p\log\frac{p}{q} - 1 \\
&\leq p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} && \text{(Eq. (16.2))} \\
&\leq \mathrm{D}(\rho\|\sigma). && \text{(Eq. (16.1))}
\end{aligned}$$

Rearranging the above inequality gives us the desired result. □

16.3 *Proof of the substate theorem*

We need the following Facts.

Fact 16.3. *Let τ, σ be states. Then,*

$$2^{\mathrm{D}_{\max}(\tau\|\sigma)} := \min_{\kappa:\tau\leq\kappa\sigma} \kappa = \max_{M\geq 0, \mathrm{Tr}(M\sigma)\leq 1} \mathrm{Tr}(M\tau).$$

Fact 16.4 (Minimax theorem)**.** *Let A_1, A_2 be non-empty convex, compact subsets of $\mathbb{R}^n$ for some $n \geq 1$. Let $u : A_1 \times A_2 \to \mathbb{R}$ be a continuous function such that,*

1. *$\forall a_2 \in A_2, u(\cdot, a_2)$ is quasi-concave, that is,*
   $\{a_1 \in A_1 : \forall a_1' \in A_1, u(a_1, a_2) \geq u(a_1', a_2)\}$ is convex.
2. *$\forall a_1 \in A_1, u(a_1, \cdot)$ is quasi-convex, that is,*
   $\{a_2 \in A_2 : \forall a_2' \in A_2, u(a_1, a_2) \leq u(a_1, a_2')\}$ is convex.

There exists $(a_1^*, a_2^*) \in A_1 \times A_2$ such that,

$$\max_{a_1 \in A_1} \min_{a_2 \in A_2} u(a_1, a_2) = \min_{a_2 \in A_2} \max_{a_1 \in A_1} u(a_1, a_2) = u(a_1^*, a_2^*).$$

Fact 16.5 (Gentle measurement lemma)**.** *Let ρ be a state and $0 \leq A^\dagger A \leq \mathbb{1}$. Then,*

$$\mathrm{F}\Big(\rho, \frac{A\rho A^\dagger}{\mathrm{Tr}(A\rho A^\dagger)}\Big) \geq \mathrm{Tr}(A\rho A^\dagger).$$

Informally: The fidelity between the initial state ρ and the state conditioned on the success of POVM element $(A^\dagger A)$ is large if the success is close to 1.

Fact 16.6 (Joint concavity of square-root fidelity)**.** *Let $\rho_0, \rho_1, \sigma_0, \sigma_1$ be states and $p \in [0,1]$. Then,*

$$\begin{aligned}&\mathrm{F}^{1/2}(p\rho_0 + (1-p)\rho_1, p\sigma_0 + (1-p)\sigma_1)\\ &\geq p\mathrm{F}^{1/2}(\rho_0, \sigma_0) + (1-p)\mathrm{F}^{1/2}(\rho_1, \sigma_1).\end{aligned}$$

Fact 16.7. *Let $\rho_0, \rho_1, \sigma_0, \sigma_1$ be states, $p \in [0,1]$ and*

$$\rho = p\,|0\rangle\langle 0| \otimes \rho_0 + (1-p)\,|1\rangle\langle 1| \otimes \rho_1,$$

$$\sigma = p\,|0\rangle\langle 0| \otimes \sigma_0 + (1-p)\,|1\rangle\langle 1| \otimes \sigma_1.$$

Then,

$$\mathrm{F}^{1/2}(\rho, \sigma) = p\mathrm{F}^{1/2}(\rho_0, \sigma_0) + (1-p)\mathrm{F}^{1/2}(\rho_1, \sigma_1).$$

Proof of Theorem 16.1. Since $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, we can assume without loss of generality that the support of σ is the full support. Denote the dimension of the full support by n. We start with the following claim.

Claim 4.

$$2^{\mathrm{D}_{\max}^{\sqrt{\varepsilon}}(\rho\|\sigma)} = \max_{\substack{M \geq 0,\\ \mathrm{Tr}(M\sigma) \leq 1}} \min_{\substack{\tilde{\rho} \geq 0,\\ \mathrm{Tr}(\tilde{\rho}) = 1,\\ \mathrm{F}(\tilde{\rho}, \rho) \geq 1-\varepsilon}} \mathrm{Tr}(M\tilde{\rho}).$$

Proof. By the definition of the ε-smooth max mutual information, we have

$$\begin{aligned}2^{\mathrm{D}_{\max}^{\sqrt{\varepsilon}}(\rho\|\sigma)} &\stackrel{\text{def}}{=} \min_{\substack{\tilde{\rho} \geq 0,\\ \mathrm{Tr}(\tilde{\rho}) = 1,\\ \mathrm{F}(\tilde{\rho}, \rho) \geq 1-\varepsilon}} \min_{\kappa : \tilde{\rho} \leq \kappa\sigma} \kappa\\ &= \min_{\substack{\tilde{\rho} \geq 0,\\ \mathrm{Tr}(\tilde{\rho}) = 1,\\ \mathrm{F}(\tilde{\rho}, \rho) \geq 1-\varepsilon}} \max_{\substack{M \geq 0,\\ \mathrm{Tr}(M\sigma) \leq 1}} \mathrm{Tr}(M\tilde{\rho}). \qquad \text{(Fact 16.3)}\end{aligned}$$

Now we want to swap the order of the minimization and maximization using the minimax theorem (Fact 16.4). To apply the minimax theorem, we first need to show that all conditions for the theorem are satisfied. Let

$$A_\rho \stackrel{\text{def}}{=} \{\tilde{\rho} \geq 0 \mid \mathrm{Tr}(\tilde{\rho}) = 1, \mathrm{F}(\tilde{\rho}, \rho) \geq 1 - \varepsilon\}$$

and

$$A_\sigma \stackrel{\text{def}}{=} \{M \geq 0 \mid \mathrm{Tr}(M\sigma) \leq 1\}.$$

Recall that if a set is closed and bounded, it is compact, so it suffices to prove that A_ρ and A_σ are closed and bounded.

1. The conditions $\mathrm{Tr}(\tilde{\rho}) = 1$ and $\mathrm{F}(\tilde{\rho}, \rho) \geq 1 - \varepsilon$ imply that A_ρ is closed and bounded. We show that A_ρ is convex by the concavity of square-root fidelity. Let $\tilde{\rho}_1, \tilde{\rho}_2 \in A_\rho$, and $p \in [0, 1]$. Then we have

$$\begin{aligned}
&\mathrm{F}^{1/2}(p\tilde{\rho}_1 + (1-p)\tilde{\rho}_2, \rho) \\
&\geq p\mathrm{F}^{1/2}(\tilde{\rho}_1, \rho) + (1-p)\mathrm{F}^{1/2}(\tilde{\rho}_2, \rho) && \text{(Fact 16.6)} \\
&\geq \sqrt{1-\varepsilon}.
\end{aligned}$$

   The other conditions

$$\mathrm{Tr}(p\tilde{\rho}_1 + (1-p)\tilde{\rho}_2) = 1$$

   and $p\tilde{\rho}_1 + (1-p)\tilde{\rho}_2 \geq 0$ are naturally satisfied, and hence, A_ρ is convex.

2. For A_σ, verifying that A_σ is closed and convex is straightforward. Let $\lambda_{\min}(\sigma)$ be the minimum eigenvalue of σ. Since we assume σ has the full support, $\sigma - \lambda_{\min}(\sigma)\mathbb{1} \geq 0$ follows. It implies that

$$\mathrm{Tr}(M(\sigma - \lambda_{\min}(\sigma)\mathbb{1})) \geq 0,$$

   and so

$$\mathrm{Tr}(M) \leq \frac{\mathrm{Tr}M\sigma}{\lambda_{\min}(\sigma)} \leq \frac{1}{\lambda_{\min}(\sigma)}.$$

   Thus, A_σ is also bounded.

3. Notice that our target function $\mathrm{Tr}(M\tilde{\rho})$ is linear in both $\tilde{\rho}$ and M, so it's also convex and concave on both of them.

Now we can apply the minimax theorem to swap the order of the minimization and maximization, and get the desired. □

To bound the RHS in Claim 4, it is sufficient to show that, for any $M \in A_\sigma$, we can construct a density operator $\rho'_M \in A_\rho$ such that $\mathrm{Tr}(M\rho'_M)$ is bounded. Let the spectral decomposition of M be

$$M = \sum_{i=1}^{n} p_i |v_i\rangle \langle v_i|.$$

For $i \in [n]$ we let

$$\lambda_i \stackrel{\text{def}}{=} \langle v_i| \rho |v_i\rangle \text{ and } \gamma_i \stackrel{\text{def}}{=} \langle v_i| \sigma |v_i\rangle .$$

To better understand this, we can view M, ρ and σ in the eigen basis of M, i.e., $|v_i\rangle$'s, then λ_i's and γ_i's are diagonal entries of ρ and σ in this basis, respectively, i.e.,

$$M = \begin{bmatrix} p_1 & & & \\ & p_2 & 0 & \\ & 0 & \ddots & \\ & & & p_n \end{bmatrix}$$

$$\rho = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & * & \\ & * & \ddots & \\ & & & \lambda_n \end{bmatrix} \quad \sigma = \begin{bmatrix} \gamma_1 & & & \\ & \gamma_2 & * & \\ & * & \ddots & \\ & & & \gamma_n \end{bmatrix}$$

Let $d \stackrel{\text{def}}{=} \mathrm{D}^{\mathrm{o}}(\rho\|\sigma)$. We define a set of 'bad' indices,

$$B \stackrel{\text{def}}{=} \{i \in [n] \mid \lambda_i > 2^{\frac{d}{\varepsilon}} \gamma_i\}.$$

Let $\Pi_B = \sum_{i\in B} |v_i\rangle \langle v_i|$ be the projector onto the subspace spanned by $\{|v_i\rangle\}_{i\in B}$. It follows that

$$\mathrm{Tr}(\Pi_B \rho) > 2^{\frac{d}{\varepsilon}} \mathrm{Tr}(\Pi_B \sigma). \tag{16.3}$$

By the definition of the observational divergence, we have

$$\begin{aligned} d &\geq \mathrm{Tr}(\Pi_B \rho) \log \frac{\mathrm{Tr}(\Pi_B \rho)}{\mathrm{Tr}(\Pi_B \sigma)} \\ &> \frac{d}{\varepsilon} \mathrm{Tr}(\Pi_B \rho). && \text{(Eq. (16.3))} \end{aligned}$$

Rearranging gives us

$$\mathrm{Tr}(\Pi_B \rho) < \varepsilon. \tag{16.4}$$

We can construct a density operator ρ'_M by projecting ρ onto the orthogonal subspace of the subspace specified by B.

$$\begin{aligned} \rho''_M &\stackrel{\text{def}}{=} (\mathbb{I} - \Pi_B)\rho(\mathbb{I} - \Pi_B), \\ \rho'_M &\stackrel{\text{def}}{=} \frac{\rho''_M}{\mathrm{Tr}(\rho''_M)}. \end{aligned}$$

From Eq. (16.4), we have

$$\begin{aligned} \mathrm{Tr}(\rho''_M) &= 1 - \mathrm{Tr}(\Pi_B \rho) \\ &> 1 - \varepsilon. \end{aligned} \tag{16.5}$$

Together with the gentle measurement lemma (Fact 16.5), we have

$$\mathrm{F}(\rho'_M, \rho) \geq \mathrm{Tr}(\rho''_M) > 1 - \varepsilon. \tag{16.6}$$

It implies that the constructed density operator ρ'_M is in A_ρ. Consider,

$$\begin{aligned}
(1-\varepsilon)\mathrm{Tr}(M\rho'_M) &\leq \mathrm{Tr}(\rho''_M)\mathrm{Tr}(M\rho'_M) && \text{(Eq. (16.5))} \\
&= \mathrm{Tr}(M\rho''_M) \\
&= \sum_{i \notin B} p_i \lambda_i \\
&\leq 2^{\frac{d}{\varepsilon}} \sum_{i \notin B} p_i \gamma_i && \text{(definition of } B\text{)} \\
&\leq 2^{\frac{d}{\varepsilon}} \sum_{i=1}^{n} p_i \gamma_i \\
&= 2^{\frac{d}{\varepsilon}} \mathrm{Tr}(M\sigma) \\
&\leq 2^{\frac{d}{\varepsilon}}. && \text{(definition of } A_\sigma\text{)}
\end{aligned}$$

Therefore, for any $M \in A_\sigma$, we can always find some density operator $\rho'_M \in A_\rho$ such that $\mathrm{Tr}(M\rho'_M)$ is upper bounded by $\frac{2^{d/\varepsilon}}{1-\varepsilon}$. Therefore,

$$\begin{aligned}
2^{\mathrm{D}^{\sqrt{\varepsilon}}_{\max}(\rho\|\sigma)} &= \max_{\substack{M \geq 0, \\ \mathrm{Tr}(M\sigma) \leq 1}} \min_{\substack{\tilde{\rho} \geq 0, \\ \mathrm{Tr}(\tilde{\rho}) = 1, \\ \mathrm{F}(\tilde{\rho}, \rho) \geq 1-\varepsilon}} \mathrm{Tr}(M\tilde{\rho}) \\
&\leq \frac{2^{\frac{d}{\varepsilon}}}{1-\varepsilon} \\
&\leq \frac{2^{\frac{\mathrm{D}(\rho\|\sigma)+1}{\varepsilon}}}{1-\varepsilon}. && \text{(Claim 3)}
\end{aligned}$$

Taking log on both sides gives us the desired.

$$\mathrm{D}^{\sqrt{\varepsilon}}_{\max}(\rho\|\sigma) \leq \frac{1}{\varepsilon}(\mathrm{D}(\rho\|\sigma) + 1) + \log \frac{1}{1-\varepsilon}. \qquad \square$$

Proof idea: By the minimax theorem, we can swap the order of the minimization and maximization, and then the proof of the substate theorem becomes simpler. Instead of picking a suitable state and optimizing over all POVM operators, the minimax theorem allows us to first fix a POVM operator and construct a corresponding state to bound the ε-smooth maximum relative-entropy.

17
The reverse Shannon theorem

We have seen what it looks like to send information reliably through a noisy channel, that is, simulating a noiseless channel using a noisy channel. Today we see how to do the reverse: try to simulate a noisy channel using a noiseless (ideal) channel. The reason for performing this reverse is to minimize communication cost needed when simulating a noisy channel, using shared resources like randomness or entanglement.

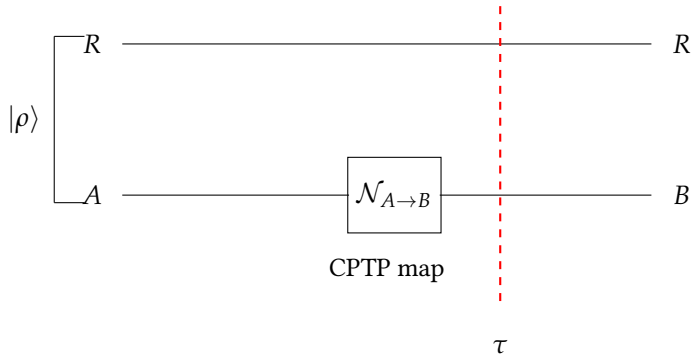

Figure 17.1: Channel $\mathcal{N}_{A\to B}$.

As shown in Figure 17.1, we start with the state $|\rho\rangle_{RA}$ sent through the channel $\mathcal{N}_{A\to B}$ to obtain the state τ.

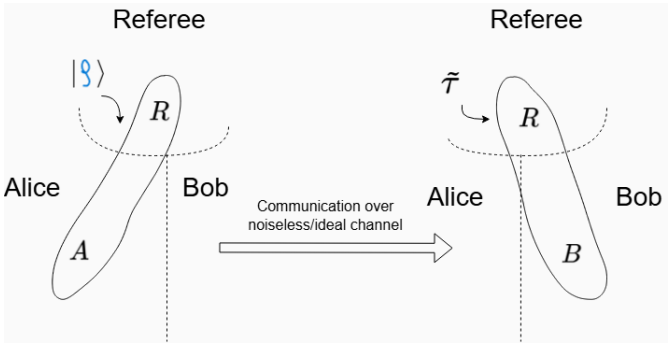

Figure 17.2: Simulating channel $\mathcal{N}_{A\to B}$ using an ideal channel.

As shown in Figure 17.2, Alice and Bob simulate the action of channel

$\mathcal{N}_{A\to B}$ using a communication protocol with a noiseless channel. They both know the description of $\mathcal{N}_{A\to B}$ but they don't know the starting state $|\rho\rangle_{RA}$. The requirement is,

$$\forall\, |\rho\rangle_{RA} : \tilde{\tau}_{RB} \approx_\varepsilon \tau_{RB}.$$

The goal is to minimize communication cost while faithfully simulating the channel for every input state $|\rho\rangle_{RA}$.

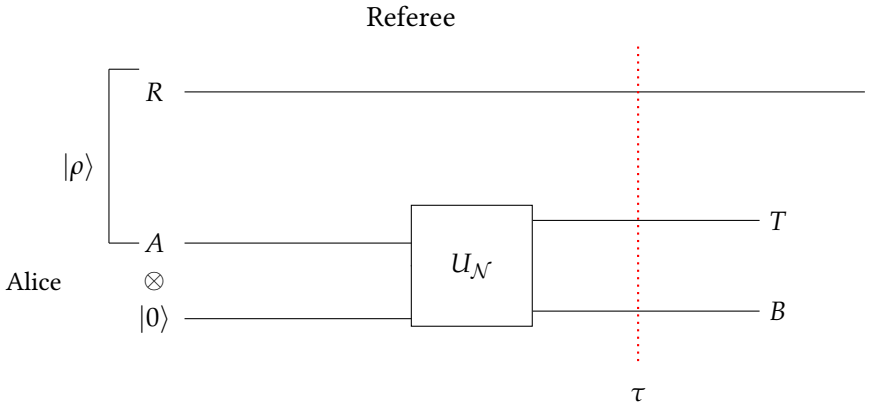

Figure 17.3: Alice's initial operation in communication protocol. After this Alice and Bob would want to implement state splitting protocol to transfer the register B to Bob.

Let $\varepsilon > 0$. We present a protocol (see Figure 17.3) with communication cost [1]

$$c \stackrel{\text{def}}{=} \max_{|\rho\rangle_{RA}} \mathrm{I}^{\varepsilon/2}_{\max}(\overset{\bullet}{R} : B)_{\mathcal{N}_{A\to B}(\rho)} + 2\log\left(\frac{2}{\varepsilon}\right).$$

We assume $|\rho\rangle_{RA}$ is the canonical purification of ρ_A. Since we can always perform local isometries on the system R, which commute with the operations of the communication protocol, so the particular purification that exists is not important.

Alice first performs a unitary $U_{\mathcal{N}}$ that effectively applies the channel $\mathcal{N}_{A\to B}$ and generates τ at her end. Alice and Bob then would want to run the state-splitting protocol to transfer the register B across. In contrast to the state-splitting protocol where Alice and Bob need to know τ, here they don't know the starting state $|\rho\rangle_{RA}$ and therefore they don't know τ. Because of this issue, we will again use the powerful minimax theorem.

For a starting state ρ_A and a protocol $\mathcal{P}$, define

$$f(\rho_A, \mathcal{P}) \stackrel{\text{def}}{=} \sqrt{\mathrm{F}(\tilde{\tau}_{RB}, \tau_{RB})},$$

where $\tilde{\tau}_{RB}$ is the output state of $\mathcal{P}$. Let $\varepsilon > 0$. We know that for every state ρ_A, there exists some protocol $\mathcal{P}_\rho$ with communication c bits (and bounded entanglement), such that

$$f(\rho_A, \mathcal{P}_\rho) \geq \sqrt{1-\varepsilon^2}.$$

Let $\mathcal{S}$ be the closure of the convex hull of $\{\mathcal{P}_\rho\}_\rho$. Then,

$$\sqrt{1-\varepsilon^2} \leq \min_{\rho_A}\max_{\mathcal{P}\in\mathcal{S}} f(\rho_A, \mathcal{P}) = \max_{\mathcal{P}\in\mathcal{S}}\min_{\rho_A} f(\rho_A, \mathcal{P}).$$

[1] Michael X. Cao, Rahul Jain, and Marco Tomamichel. Quantum channel simulation in fidelity is no more difficult than state splitting. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 1421–1425, 2024. DOI: 10.1109/ISIT57864.2024.10619461

We will show that the equality above follows from the minimax theorem (Fact 16.4). For this we need to verify that the conditions needed for the minimax theorem hold. This would imply that there exists a protocol, with communication at most c bits, such that for every input state the purified distance of the output state with τ_{RB} is at most ε.

Below we verify all the conditions needed for the minimax theorem.

1. The set of states $\{\rho_A\}$ is convex and compact.

2. The set $\mathcal{S}$ is convex and closed by definition. Since, for each ρ, the protocol $\mathcal{P}_\rho$ is a bounded-communication, bounded-entanglement protocol, the set $\mathcal{S}$ is also bounded (since the closure of a convex hull of a bounded set is bounded).

3. Fix ρ_A. We want to show that the set (call it $\mathcal{S}(\rho_A)$) of protocols in $\mathcal{S}$ maximizing $f(\rho_A, \cdot)$ is convex.

   Let $p \in [0,1]$ and $\tilde{\tau}^0_{RB}, \tilde{\tau}^1_{RB}$ be the output states of protocols $\mathcal{P}_0, \mathcal{P}_1 \in \mathcal{S}(\rho_A)$, respectively, on input $|\rho\rangle_{RA}$ (the canonical purification of ρ_A). Let $\mathcal{P} \stackrel{\text{def}}{=} p\mathcal{P}_0 + (1-p)\mathcal{P}_1$. The protocol $\mathcal{P}$ can be implemented as follows: Alice and Bob choose (using public coins) to implement $\mathcal{P}_0$ with probability p and $\mathcal{P}_1$ with probability $1-p$. Then,

   $$\tilde{\tau}_{RB} \stackrel{\text{def}}{=} p\tilde{\tau}^0_{RB} + (1-p)\tilde{\tau}^1_{RB}$$

   would be the output state of the protocol $\mathcal{P}$. Let $\tau_{RB} = \mathcal{N}_{A\to B}(\rho_{RA})$. From Fact 16.6 we have,

   $$\begin{aligned} f(\rho_A, \mathcal{P}) &= \mathrm{F}^{1/2}(\tilde{\tau}_{RB}, \tau_{RB}) \\ &\geq p \cdot \mathrm{F}^{1/2}(\tilde{\tau}^0_{RB}, \tau_{RB}) + (1-p) \cdot \mathrm{F}^{1/2}(\tilde{\tau}^1_{RB}, \tau_{RB}) \\ &= \max f(\rho_A, \cdot). \end{aligned}$$

   This implies $\mathcal{P} \in \mathcal{S}(\rho_A)$.

4. Fix $\mathcal{P}$. We want to show that the set (call it $\mathcal{S}(\mathcal{P})$) of minimizers of the function $f(\cdot, \mathcal{P})$ is convex.

   Let $p \in [0,1]$ and $\rho^0_A, \rho^1_A \in \mathcal{S}(\mathcal{P})$. Their convex combination is

   $$\rho_A \stackrel{\text{def}}{=} p\rho^0_A + (1-p)\rho^1_A.$$

   Define

   $$|\psi\rangle_{DRA} \stackrel{\text{def}}{=} \sqrt{p} \cdot |0\rangle_D |\rho^0\rangle_{RA} + \sqrt{1-p} \cdot |1\rangle_D |\rho^1\rangle_{RA}.$$

   Note that $|\psi\rangle_{DRA}$ is a purification of ρ_A.

   (a) Let $\tilde{\tau}_{DRB}, \tilde{\tau}^0_{RB}, \tilde{\tau}^1_{RB}, \tilde{\gamma}_{RB}$ be the output states of $\mathcal{P}$ on the input states $|\psi\rangle_{DRA}$ (the registers DR are with the Referee, see Figure 17.4), $|\rho^0\rangle_{RA}, |\rho^1\rangle_{RA}$ and $|\rho\rangle_{RA}$ respectively.

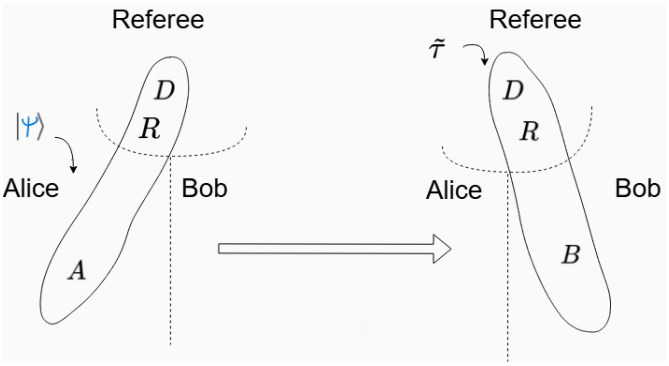

Figure 17.4: Protocol $\mathcal{P}$ with input state $|\psi\rangle_{RDA}$.

(b) Let $\tau_{DRB} \stackrel{\text{def}}{=} \mathcal{N}_{A\to B}(\psi_{DRA})$ and $\gamma_{RB} \stackrel{\text{def}}{=} \mathcal{N}_{A\to B}(\rho_{RA})$.

(c) Let $\theta_{DRB}, \tilde{\theta}_{DRB}$ be the states obtained by measuring the register D (in the computational basis) in the states $\tau_{DRB}, \tilde{\tau}_{DRB}$ respectively.

Note that the Uhlmann isometry V that takes $|\psi\rangle_{DRA}$ to $|\rho\rangle_{RA}$, commutes with all the operations in $\mathcal{P}$ and also with $\mathcal{N}_{A\to B}$. Consider,

$$\begin{aligned}
f(\rho_A, \mathcal{P}) &= \mathrm{F}^{1/2}(\tilde{\gamma}_{RB}, \gamma_{RB}) \\
&= \mathrm{F}^{1/2}(\mathcal{P}(\rho_{RA}), \mathcal{N}_{A\to B}(\rho_{RA})) \\
&= \mathrm{F}^{1/2}(\mathcal{P}(V\psi_{DRB}V^\dagger), \mathcal{N}_{A\to B}(V\psi_{DRB}V^\dagger)) \\
&= \mathrm{F}^{1/2}(V\mathcal{P}(\psi_{DRB})V^\dagger, V\mathcal{N}_{A\to B}(\psi_{DRB})V^\dagger) \\
&= \mathrm{F}^{1/2}(\mathcal{P}(\psi_{DRB}), \mathcal{N}_{A\to B}(\psi_{DRB})) \\
&= \mathrm{F}^{1/2}(\tilde{\tau}_{DRB}, \tau_{DRB}) \\
&\leq \mathrm{F}^{1/2}(\tilde{\theta}_{DRB}, \theta_{DRB}) && \text{(DPI)} \\
&= p \cdot \mathrm{F}^{1/2}(\tilde{\tau}^0_{RB}, \tau^0_{RB}) + (1-p) \cdot \mathrm{F}^{1/2}(\tilde{\tau}^1_{RB}, \tau^1_{RB}) && \text{(Fact 16.7)} \\
&= \min f(\cdot, \mathcal{P}).
\end{aligned}$$

Hence $\rho_A \in \mathcal{S}(\mathcal{P})$.

Bibliography

Anurag Anshu, Vamsi Krishna Devabathini, and Rahul Jain. Quantum communication using coherent rejection sampling. *Physical Review Letters*, 119(12), September 2017. DOI: 10.1103/physrevlett.119.120506. URL http://dx.doi.org/10.1103/PhysRevLett.119.120506.

Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. A one-shot achievability result for quantum state redistribution. *IEEE Transactions on Information Theory*, 64(3):1425–1435, 2018. DOI: 10.1109/TIT.2017.2776112.

Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. On the near-optimality of one-shot classical communication over quantum channels. *Journal of Mathematical Physics*, 60(1):012204, 01 2019a. DOI: 10.1063/1.5039796. URL https://doi.org/10.1063/1.5039796.

Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. Building blocks for communication over noisy quantum networks. *IEEE Transactions on Information Theory*, 65(2):1287–1306, February 2019b. DOI: 10.1109/TIT.2018.2851297.

Rotem Arnon-Friedman. *Device-Independent Quantum Information Processing: A Simplified Analysis*. Springer International Publishing, 2020. ISBN 9783030602314. DOI: 10.1007/978-3-030-60231-4. URL http://dx.doi.org/10.1007/978-3-030-60231-4.

Mario Berta, Matthias Christandl, and Dave Touchette. Smooth entropy bounds on one-shot quantum state redistribution. *IEEE Transactions on Information Theory*, 62(3):1425–1439, 2016.

Michael X. Cao, Rahul Jain, and Marco Tomamichel. Quantum channel simulation in fidelity is no more difficult than state splitting. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 1421–1425, 2024. DOI: 10.1109/ISIT57864.2024.10619461.

Rahul Jain and Ashwin Nayak. Short proofs of the quantum substate theorem. *IEEE Transactions on Information Theory*, 58(6):3664–3669, 2012. DOI: 10.1109/TIT.2012.2184522.

Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A property of quantum relative entropy with an application to privacy in quantum communication. *J. ACM*, 56(6), September 2009. ISSN 0004-5411. DOI: 10.1145/1568318.1568323. URL https://doi.org/10.1145/1568318.1568323.

Sumeet Khatri, Ludovico Lami, and Mark M. Wilde. *Principles of Quantum Communication Theory: A Modern Approach*. 2025. URL https://www.markwilde.com/PQCT-khatri-lami-wilde.pdf.

Ernest Y. Z. Tan. Prospects for device-independent quantum key distribution, 2024. URL https://arxiv.org/abs/2111.11769.

Marco Tomamichel. *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016. ISBN 9783319218915. DOI: 10.1007/978-3-319-21891-5. URL http://dx.doi.org/10.1007/978-3-319-21891-5.

Marco Tomamichel and Anthony Leverrier. A largely self-contained and complete security proof for quantum key distribution. *Quantum*, 1:14, July 2017. ISSN 2521-327X. DOI: 10.22331/q-2017-07-14-14. URL http://dx.doi.org/10.22331/q-2017-07-14-14.