Contents

1 Course Info

2 Background

2.1 Hilbert spaces

A complex Hilbert space $\mathcal{H}$ is a complex vector space equipped with an inner product $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ and which is complete in the norm

$$\|v\| = \langle v, v \rangle^{1/2}$$

induced by the inner product. Recall that an inner product is

1. Linear in 2nd argument[1]: $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$, $\langle x, \alpha y \rangle = \alpha \langle x, y \rangle$.

2. Conjugate symmetric: $\langle x, y \rangle = \langle y, x \rangle^*$.

3. Positive definite: $\langle x, x \rangle \geq 0$ with equality if and only if $x = 0$.

Note that (1) and (2) imply conjugate linearity in the first argument: $\langle y + z, x \rangle = \langle y, x \rangle + \langle z, x \rangle$, $\langle \alpha y, x \rangle = \alpha^* \langle y, x \rangle$.

In this course we will only deal with Hilbert spaces of finite dimension, so the completeness condition is automatically satisfied. From now on we will nearly always write vectors in $\mathcal{H}$ as **'kets'** e.g. $|\psi\rangle$. If a Hilbert space is called $\mathcal{H}_{\mathsf{Q}}$, for some Q, then we will sometimes label vectors in $\mathcal{H}_{\mathsf{Q}}$ with the same subscript (e.g. $|\psi\rangle_{\mathsf{Q}}$) to indicate where they live, and write d_{Q} for the dimension $\dim(\mathcal{H}_{\mathsf{Q}})$ of $\mathcal{H}_{\mathsf{Q}}$.

2.1.1 Computational basis

We assume that each Hilbert space comes equipped with an orthonormal basis $\{|i\rangle : i = 0, \ldots, \dim(\mathcal{H}) - 1\}$ which we declare to be *real* vectors, i.e. $|i\rangle^* = |i\rangle$ for all $0 \leq i < \dim(\mathcal{H})$, and which we call the **computational basis**. If we write $|\psi\rangle \in \mathcal{H}$ as a column vector, this consists of the components of $|\psi\rangle$ in the computational basis (unless otherwise specified) e.g. if $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ then $|\psi\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$. The orthonormality of the computational basis means that if $|\psi\rangle = \sum_{0 \leq j < \dim(\mathcal{H})} \alpha_j |j\rangle$ and $|\phi\rangle = \sum_{0 \leq j < \dim(\mathcal{H})} \beta_j |j\rangle$ then $\langle |\psi\rangle, |\phi\rangle \rangle = \sum_{0 \leq j < \dim(\mathcal{H})} \alpha_j^* \beta_j$.

2.2 Linear maps between Hilbert spaces

Given two spaces $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$ we denote the complex vector space of linear maps from $\mathcal{H}_{\mathsf{A}}$ to $\mathcal{H}_{\mathsf{B}}$ by $\mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$. We call elements of $\mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{A}})$ *operators* on $\mathcal{H}_{\mathsf{A}}$ and use the abbreviation $\mathcal{L}(\mathcal{H}_{\mathsf{A}}) := \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{A}})$.

[1] This is the usual convention in physics texts, and matches the bra-ket notation.

2.2.1 Bras; Bra-kets

The **dual space** to $\mathcal{H}_\mathsf{A}$ is the vector space $\mathcal{L}(\mathcal{H}_\mathsf{A}, \mathbb{C})$ of complex linear functionals on $\mathcal{H}_\mathsf{A}$. The inner product on $\mathcal{H}_\mathsf{A}$ provides a natural way to associate to every $|\psi\rangle$ in $\mathcal{H}_\mathsf{A}$ a unique linear functional $|\psi\rangle^\dagger$, which is defined by $|\psi\rangle^\dagger : |\phi\rangle \mapsto \langle |\psi\rangle, |\phi\rangle\rangle$. We write $|\psi\rangle^\dagger$ as a **'bra'** $\langle\psi| := |\psi\rangle^\dagger$. Conversely, for every linear functional $f \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathbb{C})$ there is a unique vector $f^\dagger$ in $\mathcal{H}_\mathsf{A}$ such that $f(|\phi\rangle) = \langle f^\dagger, |\phi\rangle\rangle$. Therefore, $\mathcal{L}(\mathcal{H}_\mathsf{A}, \mathbb{C}) = \{\langle\psi| : |\psi\rangle \in \mathcal{H}_\mathsf{A}\}$ and $\langle\psi|^\dagger = |\psi\rangle$. We will abbreviate $\langle\psi||\phi\rangle$ to $\langle\psi|\phi\rangle$ (this is called a **'bra-ket'**), and will normally make use of the identity $\langle\psi|\phi\rangle = \langle|\psi\rangle, |\phi\rangle\rangle$ to express inner products.

2.2.2 Computational basis for linear maps

The computational basis for the dual space to $\mathcal{H}_\mathsf{A}$ is $\{\langle j| : 0 \leq j < d_\mathsf{A}\}$ and we will write elements of the dual space $\langle\chi|_\mathsf{A} = \sum_{0\leq j<d_\mathsf{A}} c_j\langle j|_\mathsf{A}$ as row vectors $\langle\chi|_\mathsf{A} = (c_0\ c_1\ \cdots\ c_{d-1})$. If $|\psi\rangle_\mathsf{A} = \sum_{0\leq j<d_\mathsf{A}} \alpha_j|j\rangle$ then $\langle\psi|_\mathsf{A} = \sum_{0\leq j<d_\mathsf{A}} \alpha_j^*\langle j|_\mathsf{A} = (\alpha_0^*\ \alpha_1^*\ \cdots\ \alpha_{d-1}^*)$, which is the conjugate transpose of the column vector associated to $|\psi\rangle_\mathsf{A}$ (we will denote conjugate transpose of a matrix with a †, also). The set of maps $\{|j\rangle_\mathsf{B}\langle k|_\mathsf{A} : 0 \leq j < d_\mathsf{B}, 0 \leq k < d_\mathsf{A}\}$ comprise the *computational basis* for $\mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$. If $X = \sum_{0\leq j<d_\mathsf{B}} \sum_{0\leq k<d_\mathsf{A}} X_{jk}|j\rangle_\mathsf{B}\langle k|_\mathsf{A}$ then we will write it as a $d_\mathsf{B} \times d_\mathsf{A}$ matrix with entries X_{jk}.

2.2.3 Hermitian adjoint

Given $X \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$, the **hermitian adjoint** of X is the unique operator $X^\dagger$ in $\mathcal{L}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{A})$ such that $\langle|\phi\rangle_\mathsf{B}, X|\psi\rangle_\mathsf{A}\rangle = \langle X^\dagger|\phi\rangle_\mathsf{B}, |\psi\rangle_\mathsf{A}\rangle$ for all $|\phi\rangle_\mathsf{B} \in \mathcal{H}_\mathsf{B}$ and $|\psi\rangle_\mathsf{A} \in \mathcal{H}_\mathsf{A}$. If $X = \sum_{j,k} X_{jk}|j\rangle_\mathsf{B}\langle k|_\mathsf{A}$ then $X^\dagger = \sum_{j,k} X_{jk}^*|k\rangle_\mathsf{A}\langle j|_\mathsf{B}$, and the matrix representation for $X^\dagger$ is the conjugate transpose of X. Note that the hermitian adjoint of a complex number (considered as an operator on a one-dimensional Hilbert space) is just its complex conjugate. We have the following identities: (1) $(X|\psi\rangle)^\dagger = \langle\psi|X^\dagger$, (2) $(\langle\psi|X)^\dagger = X^\dagger|\psi\rangle$, (3) $(XY)^\dagger = Y^\dagger X^\dagger$.

2.2.4 Hermitian operators

We say an operator $X \in \mathcal{L}(\mathcal{H}_\mathsf{Q})$ is **hermitian** if $X^\dagger = X$. The set $\mathrm{Herm}(\mathcal{H}_\mathsf{Q})$ of hermitian operators on $\mathcal{H}_\mathsf{Q}$, is a *real* vector space of dimension $\dim(\mathsf{Q})^2$.

We call $E \in \mathcal{L}(\mathcal{H}_\mathsf{Q})$ an **orthogonal projection operator** (or just **'projector'**) if it satisfies $E^\dagger E = E$ (equivalently, E is hermitian and $E^2 = E$).

Theorem 2.1 (Eigendecomposition)**.** Any hermitian operator $X \in \mathrm{Herm}(\mathcal{H}_\mathsf{Q})$ has a unique decomposition

$$X = \sum_{\lambda\in\mathrm{spec}(X)} \lambda\Pi_\lambda$$

where $\mathrm{spec}(X)$ is the set of eigenvalues of X and Π_λ is the orthogonal projector onto the eigenspace corresponding to eigenvalue λ. $\mathrm{spec}(X) \subset \mathbb{R}$ and

$$\Pi_\lambda\Pi_\mu = \begin{cases} \Pi_\lambda \text{ if } \lambda = \mu, \\ 0 \text{ if } \lambda \neq \mu. \end{cases}$$

Equivalently, any hermitian operator $X \in \mathrm{Herm}(\mathcal{H}_\mathsf{Q})$ can be written as $X = \sum_{0\leq j<d_\mathsf{Q}} \lambda_j|a_j\rangle\langle a_j|$ where $|a_j\rangle$ is an eigenvector of X corresponding to eigenvalue λ_j. The eigenvalues are all real and the eigenvectors $\{|a_j\rangle : 0 \leq j < d_\mathsf{Q}\}$ form an orthonormal basis for $\mathcal{H}_\mathsf{Q}$.

2.2.5 Support of an operator

The **support** of an operator $X \in \mathcal{L}(\mathcal{H})$ is the subspace $\mathrm{supp}(X)$ of $\mathcal{H}$ that is orthogonal to the kernel of X, $\ker(X)$.

For hermitian X with an eigendecomposition $X = \sum_{0 \le i < d} \lambda_i |\alpha_j\rangle\langle\alpha_j|$, $\mathrm{supp}(X) = \sum_{i:\lambda_i \neq 0} |\alpha_j\rangle\langle\alpha_j|$.

2.2.6 Transpose; Complex conjugate

Given a linear map $X \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$, its computational basis representation is $\sum_{j,k} \langle b|_\mathsf{B} X |a\rangle_\mathsf{A} |b\rangle_\mathsf{B} \langle a|_\mathsf{A}$. Its **transpose** (with respect to the computational basis) is the map $X^\mathrm{T} \in \mathcal{L}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{A})$ defined by $\sum_{a,b} \langle b|_\mathsf{B} X |a\rangle_\mathsf{A} |a\rangle_\mathsf{A} \langle b|_\mathsf{B}$. The **complex conjugate** of X is the map $X^* \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ defined by $\sum_{a,b} \langle b|_\mathsf{B} X |a\rangle^*_\mathsf{A} |b\rangle_\mathsf{B} \langle a|_\mathsf{A}$. Note that this definition is also basis dependent, in that it reflects the fact that we have declared the computational basis vectors are real. We have the following equations: (1) $(XY)^\mathrm{T} = Y^\mathrm{T} X^\mathrm{T}$; (2) $(XY)^* = X^* Y^*$ (3) $X^\dagger = (X^*)^\mathrm{T}$.

2.2.7 Trace

Definition 2.2. For any operator $X \in \mathcal{L}(\mathcal{H}_\mathsf{Q})$, the trace of X is $\mathrm{Tr} X := \sum_{0 \le j < d_Q} \langle j|X|j\rangle$.

The trace is a linear function and it has the **cyclic property**:

$$\forall X \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B}), Y \in \mathcal{L}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{A}) : \ \mathrm{Tr} XY = \mathrm{Tr} YX.$$

From the cyclic property it follows that $\mathrm{Tr} M^{-1} X M = \mathrm{Tr} X$ for any invertible M, so although we used the computational basis in the definition above, the trace is basis independent. Note that $\mathrm{Tr} X^\mathrm{T} = \mathrm{Tr} X$, and $\mathrm{Tr} X^\dagger = (\mathrm{Tr} X)^*$.

Proposition 2.3. $\forall\ |\psi\rangle, |\phi\rangle \in \mathcal{H}$, $\mathrm{Tr} |\psi\rangle\langle\phi| = \langle\phi|\psi\rangle$. (Proof is an easy exercise.)

2.2.8 The Hilbert-Schmidt inner product

We can use the trace to define an inner product for spaces of linear maps:

Definition 2.4 (Hilbert-Schmidt inner product)**.** For X, Y in $\mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ let $\langle X, Y\rangle := \mathrm{Tr} X^\dagger Y$.

This is also an inner product on pairs of hermitian operators from the *real* vector space $\mathrm{Herm}(\mathcal{H}_\mathsf{Q})$.

2.2.9 Positivity of operators

We say an operator $X \in \mathcal{L}(\mathcal{H}_\mathsf{Q})$ is **positive semi-definite** (or simply "positive") and write $X \ge 0$, if $\langle\psi|X|\psi\rangle \ge 0$ for all $|\psi\rangle \in \mathcal{H}_\mathsf{Q}$. Given $X, Y \in \mathcal{L}(\mathcal{H}_\mathsf{Q})$ we write $X \ge Y$ iff $X - Y \ge 0$.

2.2.10 Isometries and unitaries

Definition 2.5 (Isometries)**.** A map V in $\mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ which satisfies $\langle V|\psi\rangle, V|\phi\rangle\rangle = \langle|\psi\rangle, |\phi\rangle\rangle$ (equivalently $\langle\psi|V^\dagger V|\phi\rangle = \langle\psi|\phi\rangle$) for all $|\psi\rangle, |\phi\rangle \in \mathcal{H}_\mathsf{A}$, is called an *isometry*. V is an isometry iff $V^\dagger V = \mathbb{1}_\mathsf{A}$.

Remark 2.6. $\mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ contains isometries iff $d_\mathsf{B} \ge d_\mathsf{A}$.

Definition 2.7 (Unitaries)**.** If $d_{\mathsf{B}} = d_{\mathsf{A}}$ then an isometry $U \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$ satisfies $U^\dagger U = UU^\dagger = \mathbb{1}_{\mathsf{A}}$, so $U^\dagger = U^{-1}$. Such an isometry is called *unitary*.

Proposition 2.8 (Unitary diagonalisation of unitary operators)**.** Any unitary operator $U \in \mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ can be written

$$U = WDW^\dagger.$$

where $W = \sum_{0 \le j < d_{\mathsf{Q}}} |w_j\rangle_{\mathsf{Q}}\langle j|_{\mathsf{Q}}$, $D = \sum_{0 \le j < d_{\mathsf{Q}}} \lambda_j |j\rangle_{\mathsf{Q}}\langle j|_{\mathsf{Q}}$ where, for each j, $|w_j\rangle_{\mathsf{Q}}$ is a normalised eigenvector of U corresponding to eigenvalue λ_j. The set $\{|w_j\rangle_{\mathsf{Q}} : 0 \le j < d_{\mathsf{Q}}\}$ is an orthonormal basis for $\mathcal{H}_{\mathsf{Q}}$ and the eigenvalues all have modulus one $|\lambda_j| = 1$, so W and D are unitary.

2.3 Tensor products

Given spaces $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$, the tensor product space $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ consists of finite linear combinations of the "product vectors" (or elementary tensors) $\{|\psi\rangle_{\mathsf{A}} \otimes |\phi\rangle_{\mathsf{B}} : |\psi\rangle_{\mathsf{A}} \in \mathcal{H}_{\mathsf{A}}, |\phi\rangle_{\mathsf{B}} \in \mathcal{H}_{\mathsf{B}}\}$ where the tensor product $\otimes$ of two vectors is bilinear:

1. $(|u\rangle_{\mathsf{A}} + |v\rangle_{\mathsf{A}}) \otimes |y\rangle_{\mathsf{B}} = |u\rangle_{\mathsf{A}} \otimes |y\rangle_{\mathsf{B}} + |v\rangle_{\mathsf{A}} \otimes |y\rangle_{\mathsf{B}}$,
2. $|u\rangle_{\mathsf{A}} \otimes (|x\rangle_{\mathsf{B}} + |y\rangle_{\mathsf{B}}) = |u\rangle_{\mathsf{A}} \otimes |x\rangle_{\mathsf{B}} + |w\rangle_{\mathsf{A}} \otimes |y\rangle_{\mathsf{B}}$,
3. $(\alpha|u\rangle_{\mathsf{A}}) \otimes |x\rangle_{\mathsf{B}} = \alpha(|u\rangle_{\mathsf{A}} \otimes |x\rangle_{\mathsf{B}})$,
4. $|u\rangle_{\mathsf{A}} \otimes (\alpha|x\rangle_{\mathsf{B}}) = \alpha(|u\rangle_{\mathsf{A}} \otimes |x\rangle_{\mathsf{B}})$.

for all $|u\rangle_{\mathsf{A}}, |v\rangle_{\mathsf{A}} \in \mathcal{H}_{\mathsf{A}}$, $|x\rangle_{\mathsf{B}}, |y\rangle_{\mathsf{B}} \in \mathcal{H}_{\mathsf{B}}$, $\alpha \in \mathbb{C}$. It has an inner product defined on product vectors by

$$\langle |u\rangle_{\mathsf{A}} \otimes |x\rangle_{\mathsf{B}}, |v\rangle_{\mathsf{A}} \otimes |y\rangle_{\mathsf{B}} \rangle = \langle |u\rangle_{\mathsf{A}}, |v\rangle_{\mathsf{A}} \rangle \langle |x\rangle_{\mathsf{B}}, |y\rangle_{\mathsf{B}} \rangle$$

and extended by linearity to the whole of $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$. If $\{|\alpha_j\rangle_{\mathsf{A}} : 0 \le j < d_{\mathsf{A}}\}$ is a basis for $\mathcal{H}_{\mathsf{A}}$, and $\{|\beta_k\rangle_{\mathsf{B}} : 0 \le k < d_{\mathsf{B}}\}$ is a basis for $\mathcal{H}_{\mathsf{B}}$, then the set $\{|\alpha_j\rangle_{\mathsf{A}} \otimes |\beta_k\rangle_{\mathsf{B}} : 0 \le j < d_{\mathsf{A}}, 0 \le k < d_{\mathsf{B}}\}$ is a basis for $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$, called the **product basis** of the two given bases. The **computational basis** for $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ is the product basis of the computational bases for $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$, and it is orthonormal. The tensor product of $\mathcal{H}_{\mathsf{A}}$ with a one-dimensional space $\mathcal{H}_{\mathsf{W}}$ is isomorphic to $\mathcal{H}_{\mathsf{A}}$ and we identify $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{W}}$ with $\mathcal{H}_{\mathsf{A}}$ in the obvious way: $\forall z \in \mathbb{C}, |\psi\rangle \in \mathcal{H}_{\mathsf{A}}$, $|\psi\rangle_{\mathsf{A}} \otimes (z|0\rangle_{\mathsf{W}}) = z|\psi\rangle_{\mathsf{A}}$.

The tensor product of n vectors is multilinear in its n arguments, and $\mathcal{H}_{\mathsf{Q}_1} \otimes \cdots \otimes \mathcal{H}_{\mathsf{Q}_n}$ consists of finite linear combinations of product vectors $|\psi_1\rangle_{\mathsf{Q}_1} \otimes \cdots \otimes |\psi_n\rangle_{\mathsf{Q}_n}$. The tensor product is associative, so $(|a\rangle_{\mathsf{A}} \otimes |b\rangle_{\mathsf{B}}) \otimes |c\rangle_{\mathsf{C}} = |a\rangle_{\mathsf{A}} \otimes (|b\rangle_{\mathsf{B}} \otimes |c\rangle_{\mathsf{C}}) = |a\rangle_{\mathsf{A}} \otimes |b\rangle_{\mathsf{B}} \otimes |c\rangle_{\mathsf{C}} \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}} \otimes \mathcal{H}_{\mathsf{C}}$.

2.3.1 Tensor products of bras

The tensor product of the dual spaces for $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$ consists of finite linear combinations of product bras $\langle\psi| \otimes \langle\phi|$, where $\otimes$ is again bilinear. This space is isomorphic to the dual of $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$, and we make the identification $(|\psi\rangle_{\mathsf{A}} \otimes |\phi\rangle_{\mathsf{B}})^\dagger = \langle\psi|_{\mathsf{A}} \otimes \langle\phi|_{\mathsf{B}}$. This extends to the whole of $|\psi\rangle_{\mathsf{A}} \otimes |\phi\rangle_{\mathsf{B}}$ by conjugate linearity of the adjoint †, e.g. if $|\chi\rangle_{\mathsf{AB}} = \sum_{j,k} c_{jk}|j\rangle_{\mathsf{A}} \otimes |k\rangle_{\mathsf{B}}$ then $\langle\chi|_{\mathsf{AB}} = \sum_{j,k} c^*_{jk}\langle j|_{\mathsf{A}} \otimes \langle k|_{\mathsf{B}}$

2.3.2 Tensor products of linear maps

Given $X \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{C}), Y \in \mathcal{L}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{D})$ we identify $X \otimes Y \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{C}) \otimes \mathcal{L}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{D})$ with the element of $\mathcal{L}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{C} \otimes \mathcal{H}_\mathsf{D})$ such that, for all $|\alpha\rangle_\mathsf{A}$, $|\beta\rangle_\mathsf{B}$

$$X \otimes Y |\alpha\rangle_\mathsf{A} \otimes |\beta\rangle_\mathsf{B} = (X|\alpha\rangle_\mathsf{A})_\mathsf{C} \otimes (Y|\beta\rangle_\mathsf{B})_\mathsf{D}.$$

This extends by linearity

$$\left(\sum_i \lambda_i X_i \otimes Y_i\right)\left(\sum_j \mu_j |\alpha_j\rangle_\mathsf{A} \otimes |\beta_j\rangle_\mathsf{B}\right) = \sum_i \sum_j \lambda_i \mu_j X_i \otimes Y_i |\alpha_j\rangle_\mathsf{A} \otimes |\beta_j\rangle_\mathsf{B}$$

to an isomorphism between $\mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{C}) \otimes \mathcal{L}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{D})$ and $\mathcal{L}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{C} \otimes \mathcal{H}_\mathsf{D})$.

Just like for vectors, given bases $\{X_i : i \in \{1, \ldots, d_\mathsf{A} \times d_\mathsf{C}\}\} \subset \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{C}), \{Y_j : j \in \{1, \ldots, d_\mathsf{B} \times d_\mathsf{D}\}\} \subset \mathcal{L}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{D})$, there is a product basis $\{X_i \otimes Y_j\} \subset \mathcal{L}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}) \otimes \mathcal{L}(\mathcal{H}_\mathsf{C} \otimes \mathcal{H}_\mathsf{D})$. In terms of computational bases, the isomorphism mentioned amounts to the identification

$$|c\rangle_\mathsf{C}\langle a|_\mathsf{A} \otimes |d\rangle_\mathsf{D}\langle b|_\mathsf{B} = |c\rangle_\mathsf{C} \otimes |d\rangle_\mathsf{D}\langle a|_\mathsf{A} \otimes \langle b|_\mathsf{B}.$$

Since a bra $\langle\psi|_\mathsf{Q}$ is a linear map from $\mathcal{H}_\mathsf{Q}$ to the one-dimensional space $\mathbb{C}$, $\langle\psi|_\mathsf{Q} \otimes X$ belongs to $\mathcal{L}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{Q}, \mathcal{H}_\mathsf{C})$. In a tensor product between X and a vector $|\psi\rangle_\mathsf{Q}$ in $\mathcal{H}_\mathsf{Q}$, we regard the vector as the linear map $|\psi\rangle_\mathsf{Q} : z \mapsto |\psi\rangle_\mathsf{Q}$ in $\mathcal{L}(\mathbb{C}, \mathcal{H}_\mathsf{Q})$. So $|\psi\rangle_\mathsf{Q} \otimes X \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{C} \otimes \mathcal{H}_\mathsf{Q})$.

2.3.3 Matrix representation of tensor products

We choose the matrix representation of the computational basis of a composite system so that the tensor product of two or more objects corresponds to the "Kronecker product" of the corresponding matrices: For example, if

$$\begin{aligned}
|a\rangle_\mathsf{A} =& \alpha_0|0\rangle_\mathsf{A} + \alpha_1|1\rangle_\mathsf{A} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}, \ |b\rangle_\mathsf{B} = \beta_0|0\rangle_\mathsf{B} + \beta_1|1\rangle_\mathsf{B} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \\
|c\rangle_\mathsf{C} =& \gamma_0|0\rangle_\mathsf{C} + \gamma_1|1\rangle_\mathsf{C} = \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix}, \ L \in \mathcal{L}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{C}) = \sum_{0 \le j,k < 2} \lambda_{jk} |j\rangle_\mathsf{C}\langle k|_\mathsf{B} = \begin{pmatrix} \lambda_{00} & \lambda_{01} \\ \lambda_{10} & \lambda_{11} \end{pmatrix}, \\
M \in& \mathcal{L}(\mathcal{H}_\mathsf{C}) = \sum_{0 \le j,k < 2} \mu_{jk} |j\rangle\langle k|_\mathsf{C} = \begin{pmatrix} \mu_{00} & \mu_{01} \\ \mu_{10} & \mu_{11} \end{pmatrix}
\end{aligned}$$

then

$$|a\rangle_\mathsf{A} \otimes |b\rangle_\mathsf{B} = \begin{pmatrix} \alpha_0\beta_0 \\ \alpha_0\beta_1 \\ \alpha_1\beta_0 \\ \alpha_1\beta_1 \end{pmatrix}, \ |a\rangle_\mathsf{A} \otimes |b\rangle_\mathsf{B} \otimes |c\rangle_\mathsf{C} = \begin{pmatrix} \alpha_0\beta_0\gamma_0 \\ \alpha_0\beta_0\gamma_1 \\ \alpha_0\beta_1\gamma_0 \\ \alpha_0\beta_1\gamma_1 \\ \alpha_1\beta_0\gamma_0 \\ \alpha_1\beta_0\gamma_1 \\ \alpha_1\beta_1\gamma_0 \\ \alpha_1\beta_1\gamma_1 \end{pmatrix}, \ |a\rangle \otimes L = \begin{pmatrix} \lambda_{00}\alpha_0 & \lambda_{01}\alpha_0 \\ \lambda_{10}\alpha_0 & \lambda_{11}\alpha_0 \\ \lambda_{00}\alpha_1 & \lambda_{01}\alpha_1 \\ \lambda_{10}\alpha_1 & \lambda_{11}\alpha_1 \end{pmatrix},$$

$$L \otimes M = \begin{pmatrix} \lambda_{00}\mu_{00} & \lambda_{00}\mu_{01} & \lambda_{01}\mu_{00} & \lambda_{01}\mu_{01} \\ \lambda_{00}\mu_{10} & \lambda_{00}\mu_{11} & \lambda_{01}\mu_{10} & \lambda_{01}\mu_{11} \\ \lambda_{10}\mu_{00} & \lambda_{10}\mu_{01} & \lambda_{11}\mu_{00} & \lambda_{11}\mu_{01} \\ \lambda_{10}\mu_{10} & \lambda_{10}\mu_{11} & \lambda_{11}\mu_{10} & \lambda_{11}\mu_{11} \end{pmatrix}, \ \langle a| \otimes L = \begin{pmatrix} \alpha_0^*\lambda_{00} & \alpha_0^*\lambda_{01} & \alpha_1^*\lambda_{00} & \alpha_1^*\lambda_{01} \\ \alpha_0^*\lambda_{10} & \alpha_0^*\lambda_{11} & \alpha_1^*\lambda_{10} & \alpha_1^*\lambda_{11} \end{pmatrix}.$$

2.4 Probability notation

1. If X is a random variable (RV) which takes values in $\mathcal{A}_X$, then we will use P_X, Q_X etc. for probability distributions on $\mathcal{A}_X$, which for our (finite/discrete) purposes are really *probability mass functions* $P_X : \mathcal{A}_X \to [0, 1]$ which satisfy $\sum_{x \in \mathcal{A}_X} P_X(x) = 1$. I will call them 'distributions'.

2. To say "X is distributed according to P_X" is to say that $\Pr(X = x) = P_X(x)$. But note that this equation is not the *definition* of P_X: For example, $P_X(X) \neq \Pr(X = X)$. $P_X(X)$ is the function P_X applied to the RV X. Therefore, $P_X(X)$ is itself an RV which takes the value $P_X(x)$ with probability $P_X(x)$. $\Pr(X = X)$ is just the number one! If X is distributed according to P_X, then $Q_X(X)$ is a RV which takes the value $Q_X(x)$ with probability $P_X(x)$.

3. A joint distribution P_{XY} for two random variables X, Y is a function $P_{XY} : \mathcal{A}_X \times \mathcal{A}_Y \to [0, 1]$ satisfying $\sum_{x,y} P_{XY}(x, y) = 1$. If we say "the (joint) distribution of X, Y is P_{XY}" this means $\Pr(X = x, Y = y) = P_{XY}(x, y)$. A conditional distribution $P_{X|Y}$ is a function $P_{X|Y} : \mathcal{A}_X \times \mathcal{A}_Y \to [0, 1]$ satisfying $\sum_x P_{X|Y}(x|y) = 1$ for all $y \in \mathcal{A}_Y$. $P_{Y|X=x} : \mathcal{A}_Y \to [0, 1]$ denotes a distribution for Y conditioned on $X = x$.

4. Distributions and conditional distributions with the same label but different subscripts are assumed to be compatible according to the rules of probability. So, if we have a joint distribution Q_{XY} then Q_Y must be the marginal distribution of Y when the joint distribution of X, Y is Q_{XY} (i.e. $Q_Y(y) = \sum_x Q_{XY}(x, y)$) and similarly for Q_X. Likewise, the product rule must hold $Q_{XY}(x, y) = Q_{X|Y}(x|y)Q_Y(y) = Q_{Y|X}(y|x)Q_X(x) = Q_{Y|X=x}(y)Q_X(x)$.

5. The 'default' distribution of everything is called P: That is, unless otherwise stated, $\Pr(X = x, Y = y, \ldots, W = w) = P_{XY\cdots W}(x, y, \ldots, w)$.

3 Postulates of Quantum Mechanics

This section reviews the postulates of quantum mechanics as given in many introductory quantum mechanics courses, but in a slightly unusual form. In the course we will build on these to obtain more general notions of state, measurement and time evolution.

3.1 States

Postulate 3.1 (State postulate)**.** To any system Q there is associated a complex Hilbert space $\mathcal{H}_\mathsf{Q}$ (the 'state space' of Q) and the state of the system is described by a **state vector** $|\psi\rangle$ which is simply a unit vector in $\mathcal{H}_\mathsf{Q}$ i.e. $\langle\psi|\psi\rangle = 1$. State vectors $|\psi\rangle$ and $|\psi'\rangle$ which differ only by an global phase, i.e. $|\psi'\rangle = e^{i\alpha}|\psi\rangle$ for some $\alpha \in \mathbb{R}$, represent the same state of the system.

It is important to note that *relative* phase in a superposition (a linear combination) does have physical significance, e.g. the state vectors $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle + e^{i\pi}|1\rangle)/\sqrt{2}$ are certainly not equivalent up to a global phase and so do not represent the same physical state (in fact, they are orthogonal).

The smallest interesting systems are two-dimensional, and we call these **qubits**. An arbitrary state vector of a qubit can be written thus

$$|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}.$$

where the normalisation condition is $\langle\psi|\psi\rangle = |\alpha_0|^2 + |\alpha_1|^2 = 1$. Since,

$$|\psi\rangle = e^{i\arg(\alpha_0)}\left(|\alpha_0||0\rangle + |\alpha_1|e^{i\phi}|1\rangle\right) = e^{i\arg(\alpha_0)}\left(\cos(\theta/2)|0\rangle + \sin(\theta/2)e^{i\phi}|1\rangle\right)$$

where $\phi = (\arg(\alpha_1) - \arg(\alpha_0))$ and $\theta/2 = \arccos|\alpha_0|$, we can parameterise the physically distinct state vectors by $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$. Interpreting these as spherical polar coordinates lets us identify physically distinct state vectors with points on the surface of a three-dimensional sphere which, in this context, is called the **Bloch sphere**. Examples of physical qubits are photon polarization and electron spin.

3.2 Measurements

Postulate 3.2 (Measurement postulate)**.** A measurement on a system Q whose result X (a random variable) takes values in $\mathcal{A}_X$ is represented by a PVM E which assigns to each $x \in \mathcal{A}_X$ a projector $E(x)$ on $\mathcal{H}_\mathsf{Q}$ such that

1. Projectors for different values project onto orthogonal subspaces:

$$E(x)E(y) = \begin{cases} E(x) \text{ if } x = y, \\ 0 \text{ if } x \neq y. \end{cases}$$

2. The projectors sum to the identity operator $\sum_{x \in \mathcal{A}_X} E(x) = \mathbb{1}_{\mathsf{Q}}$.

If the state vector of Q is $|\psi\rangle$, then for a measurement on Q with PVM E:

1. The probability that the measurement result is x is

$$\Pr(X = x) = \langle\psi|E(x)|\psi\rangle.$$

2. Immediately after the measurement, conditioned $X = x$, the state vector of Q becomes

$$\frac{E(x)|\psi\rangle}{\|E(x)|\psi\rangle\|}.$$

In the case where $\mathcal{A}_X$ is a set of real numbers, the eigendecomposition of hermitian operators places the PVMs on Q in one-to-one correspondence with hermitian operators on $\mathcal{H}_{\mathsf{Q}}$, which in this context are called *observables*.

3.3 Time evolution

The dynamics of a closed quantum system Q are encoded by a *Hamiltonian* H which is a an hermitian operator on $\mathcal{H}_{\mathsf{Q}}$. Its physical meaning as an observable is that it measures the total energy of the system.

Postulate 3.3 (Time evolution)**.** The time evolution of the state vector of a closed quantum system is given by a linear differential equation called the **Schrödinger equation**:

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle.$$

Here, $\hbar$ is a physical constant called the *reduced Planck constant*. We will choose our units so that it's equal to one. Solving the Schrödinger equation we obtain, for any $|\psi(t_1)\rangle$,

$$|\psi(t_2)\rangle = U(\Delta t)|\psi(t_1)\rangle$$

where $U(\Delta t) := \exp(-iH\Delta t)$ is a *unitary* operator on $\mathcal{H}_{\mathsf{Q}}$, and $\Delta t := t_2 - t_1$.

We won't talk about Hamiltonians again in this course. For us the important point is that the time evolution of a closed system is unitary and, furthermore, for *any* unitary transformation U we can find a Hamiltonian H and time interval Δt such that

$$U = \exp(-iH\Delta t).$$

3.4 Composite systems

Postulate 3.4 (Composite systems)**.** If systems Q_j, $1 \leq j \leq n$, have state spaces $\mathcal{H}_{\mathsf{Q}_j}$ then the state space of the composite system $\mathsf{Q}_1\mathsf{Q}_2 \cdots \mathsf{Q}_n$ is the tensor product $\mathcal{H}_{\mathsf{Q}_1} \otimes \mathcal{H}_{\mathsf{Q}_2} \otimes \cdots \otimes \mathcal{H}_{\mathsf{Q}_n}$. If the state vector of Q_j is $|\psi_j\rangle_{\mathsf{Q}_j}$ for each j then the state vector of the composite system is the product vector $|\psi_1\rangle_{\mathsf{Q}_1} \otimes |\psi_2\rangle_{\mathsf{Q}_2} \otimes \cdots \otimes |\psi_n\rangle_{\mathsf{Q}_n}$.

Any state vector of the composite system which is *not* a product vector is called *entangled*.

Measurements on composite systems

Proposition 3.5. Given a measurement on system A represented by a PVM E_{A}, the corresponding PVM on AB is the one which associates outcome x to $E(x)_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}}$.

♣♣Use the composite systems postulate and the measurement postulate to derive this proposition.

4 CHSH game

4.1 The rules of the game

There are two players, Alice and Bob, and a referee. We imagine that Alice and Bob each have their own laboratory. Before the game begins, Alice and Bob can communicate freely. They can discuss their strategy and send each other physical systems. It is just as if they were together in the same laboratory.

During the game all communication between the players is forbidden: There can be no communication whatsoever between Alice's lab and Bob's lab. In fact, they only communicate with the referee, and this communication has the following specific form:

- The referee sends Alice a bit S and Bob a bit T, chosen independently and uniformly at random. That is, for all $s \in \{0,1\}$, $t \in \{0,1\}$, $\Pr(S = s, T = t) = P_{ST}(s,t) = 1/4$.
- Alice must reply to referee with a bit X and Bob with a bit Y.
- They win the game if $X + Y \equiv ST \pmod 2$. We call this event **WIN**.

Note that, for $n, m \in \mathbb{Z}$, $k \in \mathbb{N}$, $k > 0$, $n \equiv m \pmod k$ means $n - m = kj$ for some $j \in \mathbb{Z}$, while $n \bmod k$ means the remainder of n divided by k. (I mixed up these notations in the first lecture, which was confusing.)

4.2 Classical strategies

We make the following assumptions:

1. After the pre-game communication has ceased, but before the game begins, the state of any systems in Alice's lab which she will use to play the game can be described by some random variable A. Similarly the state of Bob's systems can be described by some random variable B.

2. We assume that Alice's answer X is determined by A and the question S she received from the referee, while Bob's answer is determined by T and B. That is,

$$X = f_A(S) \text{ and } Y = g_B(T) \tag{4.1}$$

   where

$$\forall a \in \mathcal{A}_A,\ f_a : \{0,1\} \to \{0,1\} \text{ and} \tag{4.2}$$
$$\forall b \in \mathcal{A}_B,\ g_b : \{0,1\} \to \{0,1\}. \tag{4.3}$$

3. (A, B) is independent of (S, T), i.e.

$$\forall a, b, s, t :\ P_{ST|AB}(s,t|a,b) = P_{ST}(s,t). \tag{4.4}$$

That's it. $\mathcal{A}_A$ and $\mathcal{A}_B$ can be anything you like. In a classical physical model we could take A and B to be points in phase space which represent a snap-shot of the the physical state of the entire contents of the two laboratories before the game begins. Since the players have been communicating freely, A and B could be correlated in any way: we make no assumption on their joint distribution P_{AB}. A model for how X and Y depend on S and T which satisfies these assumptions is called a "local hidden variables model", and A and B are "local hidden variables".

Given these assumptions, what is the maximum possible value of $\Pr(\mathbf{WIN})$? We claim the answer is 3/4. First note that

$$\begin{aligned}
\Pr(\mathbf{WIN}) &\leq \max_{a,b} \Pr(\mathbf{WIN}|A=a, B=b) = \Pr(\mathbf{WIN}|A=a^*, B=b^*) && (4.5)\\
&= \sum_{(s,t)\in\{0,1\}^2} \Pr(\mathbf{WIN}, S=s, T=t|A=a^*, B=b^*) && (4.6)\\
&= \sum_{(s,t)\in\{0,1\}^2} \Pr(\mathbf{WIN}|S=s, T=t, A=a^*, B=b^*) P_{ST|AB}(s,t|a^*,b^*) && (4.7)\\
&= \sum_{(s,t)\in\{0,1\}^2} [f(s)+g(t) \equiv st \pmod 2] \frac{1}{4}. && (4.8)
\end{aligned}$$

where $f = f_{a^*}$ and $g = g_{b^*}$ and in (4.8) we used the notation

$$[q] := \begin{cases} 1 \text{ if } q \text{ is true,} \\ 0 \text{ if } q \text{ is false.} \end{cases}$$

Here we are just using basic probability and, in the final equality, the independence of (S,T) and (A,B), (4.4), and $P_{ST}(s,t) = 1/4$, and the rules of the game. The expression (4.8) is the probability that the *deterministic* strategy

$$X = f(S), Y = g(Y)$$

wins the game, so we have shown that there must be an optimal deterministic strategy. Evidently, a deterministic strategy must have $\Pr(\mathbf{WIN}) \in \{0, 1/4, 1/2, 3/4, 1\}$. $\Pr(\mathbf{WIN}) = 3/4$ can be achieved by the strategy both players always answer with zero, since this fails iff $(S,T) = (1,1)$. However, to win with probability one would require that $f(s) + g(t) \equiv 1 \pmod 2$ only for the single pair of questions $(1,1)$. But

$$\sum_{s=0}^{1}\sum_{t=0}^{1} (f(s)+g(t)) = 2\left(\sum_{x=0}^{1} f(s) + \sum_{y=0}^{1} g(t)\right) \equiv 0 \pmod 2,$$

so we must have $f(s) + g(t) \equiv 1 \pmod 2$ for an *even* number m of question pairs $(x,y) \in \{0,1\}^2$. Therefore, the best deterministic strategy has probability 3/4 of winning, and this is maximal among *all* strategies.

4.3 Some quantum mechanics

Here we will review the quantum mechanics that we'll need to describe and analyse a quantum strategy for the CHSH game.

4.3.1 Projectors

Definition 4.1 (Projectors)**.** $E \in \mathcal{L}(\mathcal{H})$ is a **projector** is $E^\dagger E$. The following statements are equivalent

1. $E^\dagger E = E$.
2. $E^\dagger = E$ and $E^2 = E$.
3. $E^\dagger = E$ and $\text{spec}(E) \subseteq \{0, 1\}$.
4. There is a subspace $S \subseteq \mathcal{H}$ such that, $\forall |\psi\rangle \in \mathcal{H}$, $E|\psi\rangle = |\psi_S\rangle$ where
$$|\psi\rangle = |\psi_S\rangle + |\psi_S^\perp\rangle,$$
is the unique decomposition of $|\psi\rangle$ with $|\psi_S\rangle \in S$ and $|\psi_{S^\perp}\rangle \in S^\perp$ ($S^\perp$ being the orthogonal complement of S in $\mathcal{H}$). We say that E projects onto S.

For finite dimensional $\mathcal{H}$, for any S there is a unique projector E_S which projects onto S. S is the $+1$ eigenspace of E_S and $\dim(S) = \text{rank}(E_S)$.

Given a unit vector $|\psi\rangle \in \mathcal{H}$, the projector onto $\text{span}(\{\psi\})$ is $|\psi\rangle\langle\psi|$. We have the following equivalent ways to represent a state[1] of Q

1. An equivalence class of unit vectors in $\mathcal{H}_\mathsf{Q}$ up to phase: $\{e^{i\phi}|\psi\rangle : \phi \in \mathbb{R}\}$, $\langle\psi|\psi\rangle = 1$.
2. A one-dimensional subspace of $\mathcal{H}_\mathsf{Q}$: $\{ae^{i\phi}|\psi\rangle : \phi, a \in \mathbb{R}\}$.
3. A rank-one projector in $\mathcal{L}(\mathcal{H}_\mathsf{Q})$: $|\psi\rangle\langle\psi|$.

4.3.2 Qubits

If $d_\mathsf{Q} := \dim(\mathcal{H}_\mathsf{Q}) = 2$, then we call Q a **qubit**. (For example, Q is the spin of an electron (or other spin-1/2 particle) or Q is the polarisation of a photon.) In this case, a general state vector $|\psi\rangle$ of Q can be written

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

where α and β are any complex numbers satisfying such that $\langle\psi|\psi\rangle = |\alpha|^2 + |\beta|^2 = 1$. The corresponding projector is

$$|\psi\rangle\langle\psi| = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{pmatrix} \alpha^* & \beta^* \end{pmatrix} = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}.$$

Note that $\text{Tr}|\psi\rangle\langle\psi| = \langle\psi|\psi\rangle = 1$.

The state vector $\cos(\theta/2)|0\rangle + \sin(\theta/2)e^{i\phi}|1\rangle$, where $\theta = 2\arccos(|\alpha|) \in [0, \pi]$ and $\phi = \arg(\alpha^*\beta) \in [0, 2\pi)$ is equivalent to $|\psi\rangle$ up to global phase. So, the states of a qubit (according to the state postulate) can be identified with points on the unit sphere by treating (θ, ϕ) as spherical polar coordinates. We call this representation the **Bloch sphere**. Note that antipodal points on the Bloch sphere correspond to pairs of orthogonal states.

[1] According to the state postulate. Soon, we will introduce a more general notion of state.

4.3.3 Composite systems

The composite systems postulate says that, given systems A and B the Hilbert space $\mathcal{H}_{\mathsf{AB}}$ of the composite system AB is $\mathcal{H}_{\mathsf{AB}} = \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ and if the state of A is $|\psi\rangle_{\mathsf{A}}$ and the state of B is $|\phi\rangle_{\mathsf{B}}$ then the state of AB is $|\psi\rangle_{\mathsf{A}} \otimes |\phi\rangle_{\mathsf{B}}$.

Note that not all state vectors in $\mathcal{H}_{\mathsf{AB}}$ can be written as elementary tensor products, for example

$$|\phi^+\rangle_{\mathsf{AB}} = \frac{1}{\sqrt{2}}(|0\rangle_{\mathsf{A}} \otimes |0\rangle_{\mathsf{B}} + |1\rangle_{\mathsf{A}} \otimes |1\rangle_{\mathsf{B}}).$$

4.3.4 Measurement

Measurement postulate

A measurement on a system Q whose result takes values in $\mathcal{A}$ is represented by a PVM, which is a map $E : \mathcal{A} \to \mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ satisfying

1. $\forall x \in S$, $E(x)$ is a projector;
2. If $x \neq y$ then $E(x)E(y) = 0$;
3. $\sum_{x\in\mathcal{A}} E(x) = \mathbb{1}$ where $\mathbb{1}$ is the identity operator on $\mathcal{H}$.

If X (an RV) is the result of measuring E (so $\mathcal{A}_X = \mathcal{A}$) when the state of Q is $|\psi\rangle$ then

- $\Pr(X = x) = \langle\psi|E(x)|\psi\rangle$;
- Immediately after the measurement, if $X = x$ then the state of Q is $E(x)|\psi\rangle/\|E(x)|\psi\rangle\|$.

Note that $\Pr(X = x) = \mathrm{Tr}|\psi\rangle\langle\psi|E(x) = \langle\psi|E(x)^\dagger E(x)|\psi\rangle = \|E(x)|\psi\rangle\|^2$.

Examples

1. The measurement of an *observable* $M \in \mathrm{Herm}(\mathcal{H}_{\mathsf{Q}})$ with eigenvalues $\mathrm{spec}(M)$, is represented by the PVM E with $\mathcal{A} = \mathrm{spec}(M)$ and $E(x)$ the projector onto the eigenspace $\{|\psi\rangle : M|\psi\rangle = x|\psi\rangle\}$.
2. Given any orthonormal basis $B = \{|e_i\rangle : 1, \ldots, d\} \subset \mathcal{H}_{\mathsf{Q}}$, $E : \{1, \ldots, d\} \to \mathcal{L}(\mathcal{H}_{\mathsf{Q}}) : i \mapsto |e_i\rangle\langle e_i|$ is a PVM, which we say "measures in the basis B".

4.3.5 Sequential measurements

Note: Unless otherwise stated we are assuming that between measurements there is no time evolution of the systems under consideration i.e. their states do not change. This is the same as assuming a time independent Hamiltonian of the form $H = h\mathbb{1}$ for some $h \in \mathbb{R}$.

1. Suppose that initially the state of Q is $|\psi\rangle$.
2. A PVM $E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ is measured with result X.
3. The measurement postulate says that $P_X(x) = \|E(x)|\psi\rangle\|^2$ and if $X = x$ then the state of Q is $E(x)|\psi\rangle/\sqrt{P_X(x)}$.

4. Now a PVM $F : \mathcal{A}_Y \to \mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ is measured with result Y.

5. By (3) and the measurement postulate, $P_{Y|X}(y|x) = \langle\psi|E(x)F(y)E(x)|\psi\rangle/P_X(x)$. $P_{YX}(y,x) = P_{Y|X}(y|x)P_X(x) = \langle\psi|E(x)F(y)E(x)|\psi\rangle$.

6. If $X = x$ and $Y = y$, the state of Q is $F(y)E(x)|\psi\rangle/\sqrt{P_{YX}(y,x)}$.

In general $E(x)$ and $F(x)$ may not commute, so the order of the measurements matters.

4.3.6 Measurements on composite systems

Suppose we have a measurement on a system A with PVM $E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_{\mathsf{A}})$. On a composite system AB, this measurement is represented by the PVM $E_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}} : x \mapsto E(x)_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}} \in \mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$. That this must be the case follows from the measurement and composite system postulates. ♣♣Why?

Suppose that this measurement is performed yielding result X, followed by a measurement on B with result Y and PVM $F : \mathcal{A}_Y \to \mathcal{L}(\mathcal{H}_{\mathsf{B}})$. If the state of AB prior to these measurements is $|\psi\rangle_{\mathsf{AB}}$ then (see previous section) the joint distribution of the results is

$$\begin{aligned} P_{XY}(x,y) =& \langle\psi|(E(x)_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}})(\mathbb{1}_{\mathsf{A}} \otimes F(y)_{\mathsf{B}})(E(x)_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}})|\psi\rangle && (4.9)\\ =& \langle\psi|(E(x)_{\mathsf{A}}\mathbb{1}_{\mathsf{A}}E(x)_{\mathsf{A}}) \otimes (\mathbb{1}_{\mathsf{B}}F(y)_{\mathsf{B}}\mathbb{1}_{\mathsf{B}})|\psi\rangle && (4.10)\\ =& \langle\psi|E(x)_{\mathsf{A}} \otimes F(y)_{\mathsf{B}}|\psi\rangle. && (4.11) \end{aligned}$$

Note that here we would get the same expression for $P_{XY}(x,y)$ if Bob measured before Alice because, for any operators J_{A} and K_{B}, $J_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}}$ commutes with $\mathbb{1}_{\mathsf{A}} \otimes K_{\mathsf{B}}$.

4.4 A quantum strategy for the CHSH game

How might Alice and Bob use quantum systems to play the CHSH game? A simple form of strategy is as follows: Alice has a system A and Bob has a system B. Before the game begins they prepare the composite system AB so that its state is $|\psi\rangle_{\mathsf{AB}} \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$.

Let $E_0 : \{0,1\} \to \mathcal{L}(\mathcal{H}_{\mathsf{A}})$ and $E_1 : \{0,1\} \to \mathcal{L}(\mathcal{H}_{\mathsf{A}})$ be PVMs representing measurements of A whose results take values in $\{0,1\}$. In playing the game Alice will use her question bit S to determine which of these measurements she performs, and then use the result of the measurement as her answer, X. That is, Alice measures E_S obtaining a result X. Likewise, Bob measures a PVM G_T on B obtaining a result Y, where $G_0 : \{0,1\} \to \mathcal{L}(\mathcal{H}_{\mathsf{B}})$ and $G_1 : \{0,1\} \to \mathcal{L}(\mathcal{H}_{\mathsf{B}})$. From the previous section we know that

$$P_{XY|ST}(x,y|s,t) = \langle\psi|_{\mathsf{AB}}E_s(x)_{\mathsf{A}} \otimes G_t(y)_{\mathsf{B}}|\psi\rangle_{\mathsf{AB}}. \tag{4.12}$$

The probability that the game is won is

$$\Pr(\mathbf{WIN}) = \sum_{s,t,x,y} [x + y \equiv st \pmod 2] P_{XY|ST}(x,y|s,t)P_{ST}(s,t). \tag{4.13}$$

4.4.1 A particular quantum strategy

Let us take A and B to be qubits (i.e. $d_{\mathsf{A}} = d_{\mathsf{B}} = 2$) and let $|\psi\rangle_{\mathsf{AB}} = |\phi^+\rangle_{\mathsf{AB}}$ where

$$|\phi^+\rangle_{\mathsf{AB}} = (|0\rangle_{\mathsf{A}} \otimes |0\rangle_{\mathsf{B}} + |1\rangle_{\mathsf{A}} \otimes |1\rangle_{\mathsf{B}})/\sqrt{2}. \tag{4.14}$$

Given a qubit Q lets define the state vector $|\eta[\phi]\rangle_{\mathsf{Q}}$ to be

$$|\eta[\phi]\rangle_{\mathsf{Q}} := \frac{1}{\sqrt{2}}(|0\rangle + e^{i\phi}|1\rangle)_{\mathsf{Q}} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}}e^{i\phi} \end{pmatrix}.$$

This lies on the 'equator' of the Bloch sphere, with azimuthal angle ϕ. Let

$$\eta[\phi]_{\mathsf{Q}} := |\eta(\phi)\rangle\langle\eta(\phi)|_{\mathsf{Q}} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}}e^{i\phi} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}e^{-i\phi} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & e^{-i\phi} \\ e^{i\phi} & 1 \end{pmatrix}$$

be the corresponding projector. Note that for all ϕ, $\{|\eta[\phi]\rangle_{\mathsf{Q}}, |\eta[\phi+\pi]\rangle_{\mathsf{Q}}\}$ is an orthonormal basis for $\mathcal{H}_{\mathsf{Q}}$. Therefore, setting

$$\begin{aligned} E_s(x)_{\mathsf{A}} =& \eta[\pi(s/2+x)]_{\mathsf{A}} \text{ and} \\ G_t(y)_{\mathsf{B}} =& \eta[\pi(t/2-y-1/4)]_{\mathsf{B}} \end{aligned} \tag{4.15}$$

defines valid PVMs. For example, G_0 is a measurement on system B in the basis

$$\{|\eta[-\pi/4]\rangle, |\eta[-3\pi/4]\rangle\}.$$

Now, let's compute $P_{XY|ST}(x,y|s,t)$ for this choice of state and measurements, and then compute the probability of winning the game. First, we note that

$$\begin{aligned} \langle\phi^+|_{\mathsf{AB}}\eta[\alpha]_{\mathsf{A}} \otimes \eta[\beta]_{\mathsf{B}}|\phi^+\rangle_{\mathsf{AB}} =& \frac{1}{8}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}^\dagger \begin{pmatrix} 1 & e^{-i\beta} & e^{-i\alpha} & e^{-i(\alpha+\beta)} \\ e^{i\beta} & 1 & e^{i(\beta-\alpha)} & e^{-i\alpha} \\ e^{i\alpha} & e^{i(\alpha-\beta)} & 1 & e^{-i\beta} \\ e^{i(\alpha+\beta)} & e^{i\alpha} & e^{i\beta} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \\ =& \frac{1}{4}\left(1 + \mathrm{Re}\left(e^{i(\alpha+\beta)}\right)\right). \end{aligned} \tag{4.16}$$

Using (4.12), (4.15) and (4.16), we find

$$\begin{aligned} P_{XY|ST}(x,y|s,t) =& \langle\phi^+|_{\mathsf{AB}}\eta[\pi(s/2+x)]_{\mathsf{A}} \otimes \eta[\pi(t/2-y-1/4)]_{\mathsf{B}}|\phi^+\rangle_{\mathsf{AB}} \\ =& \frac{1}{4}\left(1 + \mathrm{Re}\left(e^{i\pi((x-y)+(s+t)/2-1/4)}\right)\right) \end{aligned}$$

Now, since $x,y,s,t \in \{0,1\}$, we have $e^{i\pi(x-y)} = (-1)^{(x-y)} = (-1)^{(x+y) \bmod 2}$ (which is real) and

$$\mathrm{Re}(e^{i\pi((s+t)/2-1/4)}) = \frac{1}{\sqrt{2}}(-1)^{st}$$

so

$$P_{XY|ST}(x,y|s,t) = \frac{1}{4}\left(1 + \frac{1}{\sqrt{2}}(-1)^{(x+y+st) \bmod 2}\right).$$

The probability that they win the CHSH game, when $(S,T)=(s,t)$ is

$$\Pr(\mathbf{WIN}|S=s,T=t)=\sum_{x,y}[x+y\equiv st\ (\mathrm{mod}\,2)]P_{XY|ST}(x,y|s,t), \tag{4.17}$$

$$=\frac{1}{2}\left(1+\frac{1}{\sqrt{2}}\right) \tag{4.18}$$

where we used the fact that $x+y\equiv st\ (\mathrm{mod}\,2)\iff (x+y+st)\bmod 2=0$, and that there are always exactly two answer pairs (x,y) s.t. $x+y\equiv st\ (\mathrm{mod}\,2)$. Therefore,

$$\Pr(\mathbf{WIN})=\frac{1}{4}\sum_{0\leq s,t\leq 1}\Pr(\mathbf{WIN}|S=s,T=t)=\frac{1}{2}\left(1+\frac{1}{\sqrt{2}}\right)\approx 0.854>3/4.$$

4.5 Conclusions

The CHSH game can be interpreted as a kind of distributed information processing task. We have shown that using quantum effects allows them perform this simple task with a lower probability of error than would be possible classically. We have also shown that it isn't possible to explain the behaviour of quantum systems by any *local hidden variables* model. That the existence of a local hidden variables model places constraints on the statistics of repeated experiments, which are violated by the predictions of quantum mechanics (and by real experiments! See, for example, `http://fqxi.org/community/forum/topic/2581`), was the insight of John Stewart Bell. The CHSH game is named for the CHSH (Clauser, Horne, Shimony, Holt) inequality, a particular form of constraint which applies to local hidden variable models.

4.6 A closer look at the quantum strategy

Let's analyse the quantum strategy for the CHSH game described in section 4.4.1 in a sightly different way. Basic probability tells us that

$$P_{XY|ST}(x,y|s,t)=P_{Y|XST}(y|x,s,t)P_{X|ST}(x|s,t). \tag{4.19}$$

Suppose that Alice measures first. Then, the measurement postulate says that

$$P_{X|ST}(x|s,t)=P_{X|S}(x|s)=\langle\phi^+|E_s(x)_{\mathsf{A}}\otimes\mathbb{1}_{\mathsf{B}}|\phi^+\rangle=\|E_s(x)_{\mathsf{A}}\otimes\mathbb{1}_{\mathsf{B}}|\phi^+\rangle\|^2, \tag{4.20}$$

and immediately after Alice's measurement, given $S=s$ and $X=x$, the state of AB is

$$E_s(x)_{\mathsf{A}}\otimes\mathbb{1}_{\mathsf{B}}|\phi^+\rangle/\|E_s(x)_{\mathsf{A}}\otimes\mathbb{1}_{\mathsf{B}}|\phi^+\rangle\|. \tag{4.21}$$

The projectors $E_s(x)$ are rank-1 for all s,x. For an arbitrary rank-1 projector, which we can write as $|\psi\rangle_{\mathsf{A}}$ where $\langle\psi|\psi\rangle=1$, we have

$$|\psi\rangle\langle\psi|_{\mathsf{A}}\otimes\mathbb{1}_{\mathsf{B}}|\phi^+\rangle_{\mathsf{AB}}=|\psi\rangle_{\mathsf{A}}(\langle\psi|_{\mathsf{A}}\otimes\mathbb{1}_{\mathsf{B}})(|0\rangle_{\mathsf{A}}\otimes|0\rangle_{\mathsf{B}}+|1\rangle_{\mathsf{A}}\otimes|1\rangle_{\mathsf{B}})/\sqrt{2} \tag{4.22}$$

$$=|\psi\rangle_{\mathsf{A}}(\langle\psi|0\rangle|0\rangle_{\mathsf{B}}+\langle\psi|0\rangle|0\rangle_{\mathsf{B}})/\sqrt{2} \tag{4.23}$$

$$=\frac{1}{\sqrt{2}}|\psi\rangle_{\mathsf{A}}\otimes|\psi\rangle^*_{\mathsf{B}}, \tag{4.24}$$

where $|\psi\rangle_{\mathsf{B}} = \mathbb{1}_{\mathsf{B}\leftarrow\mathsf{A}}|\psi\rangle_{\mathsf{A}}$. The linear map $\mathbb{1}_{\mathsf{B}\leftarrow\mathsf{A}} \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$ is the map which takes the computational basis vectors of $\mathcal{H}_{\mathsf{A}}$ to the corresponding computational basis vectors of $\mathcal{H}_{\mathsf{B}}$:

$$\forall i \in \{0, \ldots, d_{\mathsf{A}} - 1\} : \ \mathbb{1}_{\mathsf{B}\leftarrow\mathsf{A}}|i\rangle_{\mathsf{A}} = |i\rangle_{\mathsf{B}}.$$

In (4.24) we used the fact that we have declared computational basis vectors to be real (i.e. $|i\rangle^* = |i\rangle$) and the general fact $\langle\alpha|\beta\rangle = \langle\beta|\alpha\rangle^*$ to compute

$$|\psi\rangle^* = (|0\rangle\langle 0|\psi\rangle + |1\rangle\langle 1|\psi\rangle)^* = |0\rangle\langle\psi|0\rangle + |1\rangle\langle\psi|1\rangle.$$

Note that, for any unit vector $|\psi\rangle \in \mathcal{H}_{\mathsf{A}}$, $\||\psi\rangle\langle\psi|_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}}|\phi^+\rangle_{\mathsf{AB}}\|^2 = 1/2$. Therefore, when the state of AB is $|\phi^+\rangle_{\mathsf{AB}}$, for *any* basis measurement (i.e. any PVM with rank-1 projectors), the two values the result can take have the same probability! In particular,

$$\forall x, s : P_{X|S}(x|s) = 1/2 \tag{4.25}$$

and, immediately after Alice's measurement, given $X = x, S = s$, the state of B is $|\beta_{(s,x)}\rangle := |\eta[\pi(x + s/2)]\rangle^* = |\eta[-\pi(x + s/2)]\rangle$, or, as a projector

$$\beta_{(s,x)} = |\beta_{(s,x)}\rangle\langle\beta_{(s,x)}| = \eta[-\pi(x + s/2)]. \tag{4.26}$$

In figure 4.1, we plot these four states on the Bloch sphere for B. Since Bob's PVMs have rank-1 projectors, we can plot these projectors on the Bloch sphere, also.

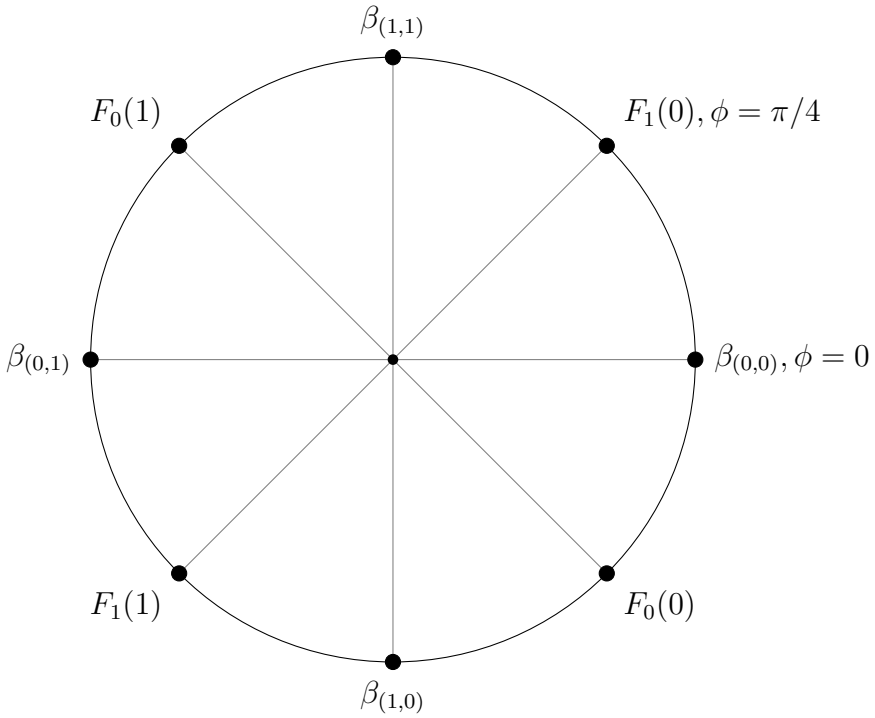


Figure 4.1: The circle is the equator of the Bloch sphere, viewed from above. On this we have plotted the states $|\beta_{(s,x)}\rangle$ of Bob's qubit after Alice's measurement, when $S = s$ (the question S she got had value s) and $X = x$ (her measurement result X had value x). We have also plotted the rank-1 projectors $F_t(y)$ corresponding to Bob's result having value y when he measures F_t (i.e. when the question T he receives has value t).

Now, given $S = s, X = x$ and $T = t$, the probability distribution for Bob's measurement result is

$$\begin{aligned} P_{Y|XST}(y|x,s,t) =& \langle \beta_{(s,x)}|F_t(y)|\beta_{(s,x)}\rangle = \mathrm{Tr}F_t(y)\beta_{(s,x)} && (4.27)\\ =& \mathrm{Tr}\eta[\pi(t/2 - y - 1/4)]\eta[\pi(x + s/2)] \text{ so, by (4.25),} && (4.28)\\ P_{XY|ST}(x,y|s,t) =& P_{Y|XST}(y|x,s,t)P_{X|ST}(x|s,t) && (4.29)\\ =& \frac{1}{2}\mathrm{Tr}\eta[\pi(t/2 - y - 1/4)]\eta[\pi(x + s/2)]. && (4.30) \end{aligned}$$

If the state of B is $\eta[\phi]$, then the probability of an obtaining a measurement result which is associated to the projector $\eta[\phi']$ is

$$\begin{aligned} \mathrm{Tr}\eta[\phi']\eta[\phi] = |\langle \eta[\phi']|\eta[\phi]\rangle|^2 =& \frac{1}{4}|1 + e^{i(\phi-\phi')}|^2 = \frac{1}{4}(1 + \mathrm{Re}[e^{i(\phi-\phi')}]) && (4.31)\\ =& \frac{1}{2}(1 + \cos(\phi - \phi')) = \cos^2\left(\frac{\phi - \phi'}{2}\right). && (4.32) \end{aligned}$$

This is zero when $|\phi - \phi'| = \pi$, and increases as $|\phi - \phi'|$ goes to zero. Using (4.30) and (4.32), and looking at the figure, we see that

$$\begin{aligned} \Pr(\mathbf{WIN}|ST = 0) =& \Pr(X = Y|ST = 0) = \cos^2(\pi/8) = \frac{1}{2}(1 + \frac{1}{\sqrt{2}}), \text{ and}\\ \Pr(\mathbf{WIN}|ST = 1) =& \Pr(X \neq Y|ST = 1) = \cos^2(\pi/8). \end{aligned}$$

5 Density operators

5.1 Uncertainty about the state vector

Suppose that we know that the state vector of a system Q belongs to some set $\{|\psi_x\rangle : x \in \mathcal{A}_X\}$ but that we aren't sure which. We represent the identity of the state as a random variable X with values in $\mathcal{A}_X$ and distribution P_X. Now, suppose that we perform a measurement with result Y (an RV taking values in $\mathcal{A}_Y$), and PVM E. Then $P_{Y|X}(y|x) = \langle\psi_x|E(y)|\psi_x\rangle = \mathrm{Tr}E(y)|\psi_x\rangle\langle\psi_x|$, and

$$P_Y(y) = \sum_{x\in\mathcal{A}_X} P_{Y|X}(y|x)P_X(x) = \mathrm{Tr}E(y)\rho$$

where $\rho = \sum_x P_X(x)|\psi_x\rangle\langle\psi_x| \in \mathcal{L}(\mathcal{H}_\mathsf{Q})$. So, the distribution, P_Y, of the measurement result Y does not depend on the particular **ensemble** $\{(P_X(x), |\psi_x\rangle\langle\psi_x|) : x \in \mathcal{A}_X\}$ but only on the **ensemble average** ρ. It is easily verified that any ensemble average ρ satisfies $\rho \geq 0$ and $\mathrm{Tr}\rho = 1$. Such an operator is called a **density operator**:

Definition 5.1 (Density operator)**.** A density operator on $\mathcal{H}_\mathsf{Q}$ an operator $\rho \in \mathrm{Herm}(\mathcal{H}_\mathsf{Q})$ with $\rho \geq 0$ (meaning ρ is a positive operator, see section 2.2.9) and $\mathrm{Tr}\rho = 1$. We will denote the set of all density operators on $\mathcal{H}_\mathsf{Q}$ by $\mathcal{D}(\mathcal{H}_\mathsf{Q})$.

For any state vector $|\psi\rangle$, there is a corresponding density operator $|\psi\rangle\langle\psi|$ which determines the state vector up to a (physically irrelevant) global phase. From now on, when we talk about a **state** of Q, we generally mean a density operator on $\mathcal{H}_\mathsf{Q}$. A **pure state** is a state of the form $|\psi\rangle\langle\psi|$ for some state vector $|\psi\rangle$. A state which is not pure is called **mixed**. Equivalently, a state is pure if its density operator is rank-1. The **maximally mixed state** of ρ is $\mathbb{1}_\mathsf{Q}/d_\mathsf{Q}$ (where $\mathbb{1}_\mathsf{Q} = \sum_{0\leq j<d_\mathsf{Q}} |j\rangle\langle j|_\mathsf{Q}$ denotes the identity operator on $\mathcal{H}_\mathsf{Q}$).

Proposition 5.2. Any density operator is the ensemble average of some ensemble of pure states.

Proof. Let ρ be any density operator on $\mathcal{H}_\mathsf{Q}$. ρ has an eigendecomposition $\sum_{0\leq j<d_\mathsf{Q}} \lambda_j|\alpha_j\rangle\langle\alpha_j|$[1] and since $\rho \geq 0$ iff $\lambda_j \geq 0$ for all j and $\mathrm{Tr}\rho = 1$ iff $\sum_{0\leq j<d_\mathsf{Q}} \lambda_j = 1$, the eigenvalues correspond to a probability distribution (on $\{0, \ldots, d_\mathsf{Q} - 1\}$, say), and ρ is the ensemble average of $\{(\lambda_j, |\alpha_j\rangle\langle\alpha_j|) : 0 \leq j < d_\mathsf{Q}\}$. □

We have shown that the set $\mathcal{D}(\mathcal{H}_\mathsf{Q})$ of density operators on $\mathcal{H}_\mathsf{Q}$ is the convex hull of the set of pure states (as projectors) $\{|\psi\rangle\langle\psi| : |\psi\rangle \in \mathcal{H}_\mathsf{Q}, \langle\psi|\psi\rangle = 1\}$.

5.2 Measuring PVMs when the state is mixed

Let's derive a "measurement postulate" for the measurement of PVMs when our state is described by a density operator.

[1] By this I mean that, for each i, j, $\langle\alpha_i|\alpha_j\rangle = \delta_{ij}$, $\rho|\alpha_i\rangle = \lambda_i|\alpha_i\rangle$ and $\rho = \sum_{0\leq j<d_\mathsf{Q}} \lambda_j|\alpha_j\rangle\langle\alpha_j|$.

Proposition 5.3. If the state of a system Q is ρ, where ρ is a density operator in $\mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ and a PVM $E : \mathcal{A}_Y \to \mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ is measured with result Y, then

1. $P_Y(y) = \mathrm{Tr}E(y)\rho$ and
2. Immediately after the measurement, if $Y = y$, then the state of Q is $\frac{E(y)\rho E(y)}{\mathrm{Tr}E(y)\rho}$.

Proof. We already established (1) in the previous section, so it remains to derive (2). Suppose that we have assigned ρ to Q because we know that Q has pure state $|\psi_x\rangle$ with probability $P_X(x)$, and

$$\sum_{x \in \mathcal{A}_X} P_X(x)|\psi_x\rangle\langle\psi_x| = \rho. \tag{5.1}$$

Using the measurement postulate, we know that for a measurement of E, the distribution of Y given X can is

$$P_{Y|X}(y|x) = \mathrm{Tr}E(y)|\psi_x\rangle\langle\psi_x| \tag{5.2}$$

(note that this is only defined when $P_X(x) > 0$) and immediately after the measurement, if $X = x$ and $Y = y$ then the state of Q is

$$|\psi_{x,y}\rangle = \frac{E(y)|\psi_x\rangle}{\sqrt{P_{Y|X}(y|x)}} \tag{5.3}$$

which is defined only when $P_{Y|X}(y|x) > 0$ and $P_X(x) > 0$, or equivalently, when $P_{XY}(x, y) > 0$. Conditional on $Y = y$ (we can assume $P_Y(y) > 0$) the state of Q is $|\psi_{x,y}\rangle$ with probability $P_{X|Y}(x|y)$, so we must assign to Q the state (density operator)

$$\sum_x P_{X|Y}(x|y)|\psi_{x,y}\rangle\langle\psi_{x,y}| = \sum_x P_{X|Y}(x|y)\frac{E(y)|\psi_x\rangle\langle\psi_x|E(y)}{P_{Y|X}(y|x)} \tag{5.4}$$

$$= \frac{E(y)\left(\sum_x P_X(x)|\psi_x\rangle\langle\psi_x|\right)E(y)}{P_Y(y)} = \frac{E(y)\rho E(y)}{\mathrm{Tr}E(y)\rho}. \tag{5.5}$$

Here we are summing only over x such that $P_{X|Y}(x|y) > 0$, and since we are assuming $P_Y(y) > 0$, we have $P_{XY}(x, y) > 0$ for all terms in the sum. We have used the fact that $\frac{P_{X|Y}(x|y)}{P_{Y|X}(y|x)} = \frac{P_{XY}(x,y)P_X(x)}{P_{XY}(x,y)P_Y(y)}$ and (5.1). Note that our derivation did not depend on the specific pure state ensemble $\{(P_X(x), |\psi_x\rangle\langle\psi_x|) : x \in \mathcal{A}_X\}$. □

5.3 Storing classical information in quantum systems

Evidently it is possible to reliably store information in quantum systems. If we accept that the universe is quantum mechanical, then *all* reliable information storage methods are evidence of this!

One way this can be done is to use an orthogonal basis of pure states to represent the value we wish to store. For example, we could store a classical bit in a qubit by using $|0\rangle$ to represent 0 and $|1\rangle$ represent 1. To retrieve the bit, one would measure the qubit in the computational basis. It would work just as well to use $|+\rangle$ and $|-\rangle$ to represent the two possible values, because they are orthogonal and, therefore, a measurement in the $\{|+\rangle, |-\rangle\}$ basis always gives the correct value for the stored bit.

More generally, given a value in some finite set $\mathcal{A}_X$, we can store it in a quantum system $\tilde{\mathsf{X}}$ with $d_{\tilde{\mathsf{X}}} = |\mathcal{A}_X|$. For convenience, we label the computational basis of the system $\tilde{\mathsf{X}}$ by the elements of $\mathcal{A}_X$ rather than numbers, and we store $x \in \mathcal{A}_X$ as $|x\rangle_{\tilde{\mathsf{X}}}$.

Suppose we store a particular value $x \in \mathcal{A}_X$ in $\tilde{\mathsf{X}}$ in this way. If we measure the PVM

$$C : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_{\tilde{\mathsf{X}}}) : x \mapsto |x\rangle\langle x|$$

(this is measurement in the computational basis) and call the result X', then

$$P_{X'}(x') = \mathrm{Tr} C(x')|x\rangle\langle x| = |\langle x'|x\rangle|^2 = \delta_{x'x}.$$

So, $X' = x$ with certainty: the storage is perfectly reliable. Furthermore, after measuring E, the state of $\tilde{\mathsf{X}}$ is certainly

$$C(x)|x\rangle\langle x|C(x) = |x\rangle\langle x|.$$

That is, the state (and hence, the value stored) has not been disturbed by the measurement.

If we say, without qualification, that a random variable X is stored in system $\tilde{\mathsf{X}}$, then we mean that the value of X has been stored in the way just described, and assume that the computational basis of $\tilde{\mathsf{X}}$ is labeled by $\mathcal{A}_X$. If X is stored in $\tilde{\mathsf{X}}$ then the state $\rho_{\tilde{\mathsf{X}}}$ of $\tilde{\mathsf{X}}$ is determined by the distribution P_X

$$\rho_{\tilde{\mathsf{X}}} = \sum_{x \in \mathcal{A}_X} P_X(x)|x\rangle\langle x|. \tag{5.6}$$

If I measure C on $\tilde{\mathsf{X}}$ obtaining result X' then, unsurprisingly,

$$P_{X'}(x') = \mathrm{Tr} C(x')\rho_{\tilde{\mathsf{X}}} = P_X(x').$$

What happens to the state of $\tilde{\mathsf{X}}$? Well, the previous section tells us that, if $X' = x'$ the state we should assign to $\tilde{\mathsf{X}}$ after the measurement is

$$\frac{C(x')\rho_{\tilde{\mathsf{X}}}C(x')}{\mathrm{Tr} C(x')\rho_{\tilde{\mathsf{X}}}} = |x'\rangle\langle x'|.$$

In this case, the form of the new state just reflects the fact that we have learnt that $X = x'$.

5.3.1 Copying classical information

There is nothing to stop us from building a device that measures C on $\tilde{\mathsf{X}}$ and the stores the result X' in another system $\tilde{\mathsf{X}}'$. The effect of this is to make a *copy* (or *clone*) the value stored in $\tilde{\mathsf{X}}$. If $\tilde{\mathsf{X}}$ has state $\rho_{\tilde{\mathsf{X}}}$, as in (5.6), before the device is applied, then afterwards we know that the state of $\tilde{\mathsf{X}}\tilde{\mathsf{X}}'$ must be

$$\sum_x P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |x\rangle\langle x|_{\tilde{\mathsf{X}}'}. \tag{5.7}$$

The form of the state reflects the fact that $\Pr(X' = X) = 1$: the original data and its copy are perfectly correlated.

5.4 States of subsystems

5.4.1 Partial trace

Recall that in section 2.2.7 we defined the trace $\mathrm{Tr}M$ of an operator $M \in \mathcal{L}(\mathcal{H})$

$$\mathrm{Tr}M := \sum_{0 \leq i < \dim(\mathcal{H})} \langle i|M|i\rangle.$$

Proposition 5.4. If Tr_{R} is a linear map from $\mathcal{L}(\mathcal{H}_{\mathsf{Q}} \otimes \mathcal{H}_{\mathsf{R}})$ to $\mathcal{L}(\mathcal{H}_{\mathsf{Q}})$, then the following statements are equivalent

1. $\forall M_{\mathsf{QR}} \in \mathcal{L}(\mathcal{H}_{\mathsf{Q}} \otimes \mathcal{H}_{\mathsf{R}}), J_{\mathsf{Q}} \in \mathcal{L}(\mathcal{H}_{\mathsf{Q}}) :\ \mathrm{Tr}J_{\mathsf{Q}} \otimes \mathbb{1}_{\mathsf{R}} M_{\mathsf{QR}} = \mathrm{Tr}J_{\mathsf{Q}}\mathrm{Tr}_{\mathsf{R}} M_{\mathsf{QR}}$;
2. $\forall K_{\mathsf{Q}} \in \mathcal{L}(\mathcal{H}_{\mathsf{Q}}), L_{\mathsf{R}} \in \mathcal{L}(\mathcal{H}_{\mathsf{R}}) :\ \mathrm{Tr}_{\mathsf{R}} K_{\mathsf{Q}} \otimes L_{\mathsf{R}} = K_{\mathsf{Q}}(\mathrm{Tr}L_{\mathsf{R}})$;
3. $\mathrm{Tr}_{\mathsf{R}} M_{\mathsf{QR}} = \sum_{0 \leq r < d_{\mathsf{R}}} (\mathbb{1}_{\mathsf{Q}} \otimes \langle r|_{\mathsf{R}}) M_{\mathsf{QR}} (\mathbb{1}_{\mathsf{Q}} \otimes |r\rangle_{\mathsf{R}})$;

and there a unique map Tr_{R} which satisfies them.

In the third statement, note that $\mathbb{1}_{\mathsf{Q}} \otimes \langle r|_{\mathsf{R}}$ is a linear map from $\mathcal{H}_{\mathsf{Q}} \otimes \mathcal{H}_{\mathsf{R}}$ to $\mathcal{H}_{\mathsf{Q}}$, while $\mathbb{1}_{\mathsf{Q}} \otimes |r\rangle_{\mathsf{R}}$ is a linear map from $\mathcal{H}_{\mathsf{Q}}$ to $\mathcal{H}_{\mathsf{Q}} \otimes \mathcal{H}_{\mathsf{R}}$. We can write them in terms of computational basis bras and kets as

$$\mathbb{1}_{\mathsf{Q}} \otimes \langle r|_{\mathsf{R}} = \sum_{0 \leq q < d_{\mathsf{Q}}} |q\rangle_{\mathsf{Q}}\langle q|_{\mathsf{Q}} \otimes \langle r|_{\mathsf{R}} \text{ and } \mathbb{1}_{\mathsf{Q}} \otimes |r\rangle_{\mathsf{R}} = \sum_{0 \leq q < d_{\mathsf{Q}}} |q\rangle_{\mathsf{Q}} \otimes |r\rangle_{\mathsf{R}}\langle q|_{\mathsf{Q}}. \tag{5.8}$$

Definition 5.5 (Partial trace)**.** The unique linear map Tr_{R} from $\mathcal{L}(\mathcal{H}_{\mathsf{Q}} \otimes \mathcal{H}_{\mathsf{R}})$ to $\mathcal{L}(\mathcal{H}_{\mathsf{Q}})$, which satisfies the three, equivalent, statements in the preceding proposition is called the "partial trace over R".

Example 5.6. Consider the state $|\phi^+\rangle_{\mathsf{AB}}$ we used in our quantum strategy for the CHSH game. Using the second definition of partial trace from Prop. 5.4, we compute

$$\begin{aligned}
\mathrm{Tr}_{\mathsf{B}}|\phi^+\rangle\langle\phi^+|_{\mathsf{AB}} =& \mathrm{Tr}_{\mathsf{B}} \frac{1}{2}\left(\sum_{i,j=0}^{1} |i\rangle_{\mathsf{A}} \otimes |i\rangle_{\mathsf{B}}\langle j|_{\mathsf{A}} \otimes \langle j|_{\mathsf{B}}\right) && (5.9)\\
=& \frac{1}{2}\mathrm{Tr}_{\mathsf{B}}\left(\sum_{i,j=0}^{1} |i\rangle\langle j|_{\mathsf{A}} \otimes |i\rangle\langle j|_{\mathsf{B}}\right) && (5.10)\\
=& \frac{1}{2}\sum_{i,j=0}^{1} |i\rangle\langle j|_{\mathsf{A}}(\mathrm{Tr}|i\rangle\langle j|_{\mathsf{B}}) && (5.11)\\
=& \frac{1}{2}\sum_{i,j=0}^{1} |i\rangle\langle j|_{\mathsf{A}}\langle i|j\rangle = \frac{1}{2}\sum_{i=0}^{1} |i\rangle\langle i|_{\mathsf{A}}. && (5.12)
\end{aligned}$$

So, $\mathrm{Tr}_{\mathsf{B}}|\phi^+\rangle\langle\phi^+|_{\mathsf{AB}} = \mathbb{1}_{\mathsf{A}}/2$ is the maximally mixed state of A.

Example 5.7. Suppose $d_{\mathsf{Q}} = d_{\mathsf{R}} = 2$, then we can write out any M_{QR} in the computational basis $M_{\mathsf{QR}} = \sum_{i,j,k,l=0}^{1} m_{ij,kl}|i\rangle_{\mathsf{Q}} \otimes |j\rangle_{\mathsf{R}}\langle k|_{\mathsf{Q}} \otimes \langle l|_{\mathsf{R}}$. The matrix representation of this is

$$M_{\mathsf{QR}} = \begin{pmatrix} m_{00,00} & m_{00,01} & m_{00,10} & m_{00,11} \\ m_{01,00} & m_{01,01} & m_{01,10} & m_{01,11} \\ m_{10,00} & m_{10,01} & m_{10,10} & m_{10,11} \\ m_{11,00} & m_{11,01} & m_{11,10} & m_{11,11} \end{pmatrix} \tag{5.13}$$

then

$$\mathrm{Tr}_{\mathsf{R}} M_{\mathsf{QR}} = \begin{pmatrix} m_{00,00}+m_{01,01} & m_{00,10}+m_{01,11} \\ m_{10,00}+m_{11,01} & m_{10,10}+m_{11,11} \end{pmatrix}, \text{ and}$$
$$\mathrm{Tr}_{\mathsf{Q}} M_{\mathsf{QR}} = \begin{pmatrix} m_{00,00}+m_{10,10} & m_{00,01}+m_{10,11} \\ m_{10,00}+m_{11,01} & m_{01,01}+m_{11,11} \end{pmatrix}.$$

5.4.2 States of subsystems

Suppose that we have a composite system QR which we know to have state vector $|\psi\rangle_{\mathsf{QR}}$. Unless $|\psi\rangle_{\mathsf{QR}}$ is a product vector, we cannot represent the state of subsystem Q (or R) by a state vector. *What is the state of* Q?

To be general, let's suppose the state of QR is any density operator ρ_{QR} (which could be the pure state $|\psi\rangle\langle\psi|_{\mathsf{QR}}$, for example). Consider measuring a PVM E on Q with result Y. Then, from our discussion of measurements on composite systems in section 4.3.6 of Handout 2, and from section 5.2 we know that

$$\Pr(Y = y) = \mathrm{Tr} E(y)_{\mathsf{Q}} \otimes \mathbb{1}_{\mathsf{R}} \rho_{\mathsf{QR}} = \mathrm{Tr} E(y)_{\mathsf{Q}} \rho_{\mathsf{Q}} \tag{5.14}$$

where $\rho_{\mathsf{Q}} = \mathrm{Tr}_{\mathsf{R}} \rho_{\mathsf{QR}}$ is given by the partial trace of ρ_{QR} over R. (Applying the map Tr_{R} is sometimes called "tracing out" R.) It is easy to check that ρ_{Q} is a density operator on $\mathcal{H}_{\mathsf{Q}}$, since it determines the probabilities of any PVM we might measure on the system Q, ρ_{Q} must be the state that we assign to Q.

Proposition 5.8 (♣♣ Prove this.)**.** If ρ_{QR} is a density operator, then $\rho_{\mathsf{Q}} = \mathrm{Tr}_{\mathsf{R}} \rho_{\mathsf{QR}}$ is a density operator.

5.5 Extensions and Purifications

Definition 5.9. If a state η_{QR} of QR satisfies $\mathrm{Tr}_{\mathsf{R}} \eta_{\mathsf{QR}} = \rho_{\mathsf{Q}}$ then we say that η_{QR} is an **extension** of the state ρ_{Q} to QR. An extension of ρ_{Q} which is pure is called a **purification** of ρ_{Q}.

Proposition 5.10. Suppose that we have a system Q with state ρ_{Q}. It is always possible to come up with a system R (we are not saying this system really exists) and a pure state η_{QR} such that $\mathrm{Tr}_{\mathsf{R}} \eta_{\mathsf{QR}} = \rho_{\mathsf{Q}}$ - that is, η_{QR} is a purification of ρ_{Q}.

Proof. We know there is at least one ensemble of pure states $\{(P(x), |\psi_x\rangle\langle\psi_x|_{\mathsf{Q}}) : 0 \le x < k\}$ such that $\rho_{\mathsf{Q}} = \sum_x P(x) |\psi_x\rangle\langle\psi_x|_{\mathsf{Q}}$. Let R be a k-dimensional system, and let $|\psi\rangle_{\mathsf{QR}} = \sum_{0 \le x < k} \sqrt{P(x)} |\psi_x\rangle_{\mathsf{Q}} \otimes |x\rangle_{\mathsf{R}}$. Then, using $\mathrm{Tr}_{\mathsf{R}} |x\rangle\langle x'|_{\mathsf{R}} = \langle x'|x\rangle = \delta_{x',x}$, the state of Q is

$$\mathrm{Tr}_{\mathsf{R}} |\psi\rangle\langle\psi|_{\mathsf{QR}} = \sum_{0 \le x, x' < k} \sqrt{P(x)P(x')} |\psi_x\rangle \langle\psi_{x'}|_{\mathsf{Q}} \langle x'|x\rangle_{\mathsf{R}} = \rho_{\mathsf{Q}} \tag{5.15}$$

as required. □

We should take care to note what Prop. 5.10 does *not* say. Suppose we have a system AB in some state ρ_{AB}. While we can certainly find a purification η_{ABR} (we are just taking $\mathsf{Q} = \mathsf{AB}$ in the Prop. 5.10) it is not necessarily possible to find any extension η_{ABR} such that the state $\eta_{\mathsf{AR}} = \mathrm{Tr}_{\mathsf{B}} \eta_{\mathsf{ABR}}$ of AR *alone* is pure. For example, if A and B are qubits and the state of AB is $|\phi^+\rangle$ (the state we used in the CHSH game) then there is no such extension.

6 Time evolution

6.1 Unitary evolution

The time evolution postulate of quantum mechanics states that, between measurements, the time evolution of the state vector of a closed quantum system Q is given by a linear differential equation called the **Schrödinger equation**: $i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle$. Here, $\hbar$ is the *reduced Planck constant*, we choose our units so that it's equal to one. $H \in \mathrm{Herm}(\mathcal{H}_\mathsf{Q})$ is the Hamiltonian - the observable corresponding to the total energy of a system. Solving the Schrödinger equation (for a time independent H) we obtain, for any $|\psi(t_1)\rangle$, $|\psi(t_2)\rangle = U(\Delta t)|\psi(t_1)\rangle$ where $U(\Delta t) := \exp(-iH\Delta t)$ is a *unitary* operator on $\mathcal{H}_\mathsf{Q}$, and $\Delta t := t_2 - t_1$. We won't talk about Hamiltonians again in this course. For us the important point is that the time evolution of a closed system is unitary and, furthermore, for *any* unitary transformation U we can find a Hamiltonian H and time interval Δt such that $U = \exp(-iH\Delta t)$. We will often imagine that we have sufficient control over systems to cause any unitary time evolution we like (by intervening to change the Hamiltonian). Sometimes this is not unrealistic: e.g. one can implement any unitary on a photon's polarisation (an example of a qubit) by passing the photon through various optical components.

Given a unitary time evolution U, we know how it acts on density operators corresponding to pure states of the system: $|\psi\rangle\langle\psi| \mapsto U|\psi\rangle\langle\psi|U^\dagger$. Since the time evolved ensemble average of an ensemble of pure states should be the ensemble average of the time evolved pure state, density operators evolve according to $\rho \mapsto U\rho U^\dagger$.

6.2 Operations

It is not hard to come up with some realistic examples of non-unitary evolution.

1. Uncertainty about unitary evolution, e.g. with probability p_i unitary U_i occurred. $\rho \mapsto \sum_i p_i U_i \rho U_i^\dagger$ (relevant to modelling *noise* in quantum computers, for example).

2. A PVM (with more than one possible outcome) is performed on the system (see example sheet 1).

3. Adding a system: $\rho_\mathsf{Q} \mapsto \rho_\mathsf{Q} \otimes \sigma_\mathsf{R}$.

4. **Isometric evolution**: $\rho_\mathsf{A} \mapsto V_{\mathsf{B}\leftarrow\mathsf{A}}\rho_\mathsf{A}V_{\mathsf{A}\leftarrow\mathsf{B}}^\dagger$ where $V \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ is an isometry, i.e. $V_{\mathsf{A}\leftarrow\mathsf{B}}^\dagger V_{\mathsf{B}\leftarrow\mathsf{A}} = \mathbb{1}_\mathsf{A}$. If $\sigma_\mathsf{R} = |\sigma\rangle\langle\sigma|_\mathsf{R}$ then adding a system is an example of isometric evolution with $V_{\mathsf{QR}\leftarrow\mathsf{R}} = \sum_{0\le i<d_\mathsf{Q}} |i\rangle_\mathsf{Q} \otimes |\sigma\rangle_\mathsf{R}\langle i|_\mathsf{Q} \in \mathcal{L}(\mathcal{H}_\mathsf{Q} \otimes \mathcal{H}_\mathsf{R}, \mathcal{H}_\mathsf{Q})$.

5. The **identity operation**, $\mathbf{id}^{\mathsf{B}\leftarrow\mathsf{A}} : X_\mathsf{A} \mapsto \mathbb{1}_{\mathsf{B}\leftarrow\mathsf{A}}X_\mathsf{A}\mathbb{1}_{\mathsf{A}\leftarrow\mathsf{B}}^\dagger$ is another simple example of isometric evolution.

6. Removing a system: $\rho_\mathsf{QR} \mapsto \mathrm{Tr}_\mathsf{R}\rho_\mathsf{QR}$.

7. Compositions of these, e.g. $\rho_\mathsf{QR} \mapsto \mathrm{Tr}_\mathsf{R}U_\mathsf{RQ}\rho_\mathsf{Q} \otimes \sigma_\mathsf{R}U_\mathsf{RQ}^\dagger$.

These are all examples of **operations**, that is, linear maps from one space of operators to another which are **completely positive** and **trace preserving**.

Definition 6.1. A linear map $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} : \mathcal{L}(\mathcal{H}_\mathsf{A}) \to \mathcal{L}(\mathcal{H}_\mathsf{B})$ is

1. **trace preserving (TP)** if for all $X_\mathsf{A} \in \mathcal{L}(\mathcal{H}_\mathsf{A})$ $\mathrm{Tr}\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_\mathsf{A} = \mathrm{Tr}X_\mathsf{A}$;
2. **positive** if for all $X_\mathsf{A} \in \mathcal{L}(\mathcal{H}_\mathsf{A})$ such that $X_\mathsf{A} \geq 0$, $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_\mathsf{A} \geq 0$;
3. **completely positive (CP)** if, for any system R, $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathbf{id}^{\mathsf{R}\leftarrow\mathsf{R}}$ is positive.
4. An **operation** (or CPTP map) if it is completely positive and trace preserving.

♣♣ Give an example of a map which is positive but not completely positive.

Proposition 6.2. If $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ and $\mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}$ are positive then their composition $\mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ is positive. If they are CP then their composition is CP. If they are TP then their composition is TP. Consequently, compositions of operations are operations.

Proposition 6.3. Given maps $\mathcal{M}_j^{\mathsf{B}\leftarrow\mathsf{A}}$, let $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} = \sum_i p_i\mathcal{M}_j^{\mathsf{B}\leftarrow\mathsf{A}}$ where $p_i \geq 0$ are real numbers. If the $\mathcal{M}_j^{\mathsf{B}\leftarrow\mathsf{A}}$ are positive then $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ is positive. If the $\mathcal{M}_j^{\mathsf{B}\leftarrow\mathsf{A}}$ are CP then $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ is CP.

Proposition 6.4. Maps of the form $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} : X_\mathsf{A} \mapsto ZX_\mathsf{A}Z^\dagger$, where $Z \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$, are CP.

Proof. If $X_\mathsf{A} \geq 0$, then $\langle\psi|_\mathsf{B}ZX_\mathsf{A}Z^\dagger|\psi\rangle_\mathsf{B} = \langle\psi'|_\mathsf{A}X_\mathsf{A}|\psi'\rangle_\mathsf{A} \geq 0$ where $|\psi'\rangle_\mathsf{A} = Z^\dagger|\psi\rangle_\mathsf{B}$ for all $|\psi\rangle_\mathsf{B}$, so the map is positive. Since $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathbf{id}^{\mathsf{R}\leftarrow\mathsf{R}}X_{\mathsf{AR}} = (Z \otimes \mathbb{1}_\mathsf{R})X_{\mathsf{AR}}(Z \otimes \mathbb{1}_\mathsf{R})^\dagger$ is positive (by the same reasoning) for any R, $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ is completely positive. □

Proposition 6.5. The following classes of maps are operations:

1. Adding a system in a fixed state, uncorrelated to the existing system: $\rho_\mathsf{Q} \mapsto \rho_\mathsf{Q} \otimes \sigma_\mathsf{R}$.
2. Isometric evolution: $\rho_\mathsf{A} \mapsto V\rho_\mathsf{A}V^\dagger$, where $V \in \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathcal{H}_\mathsf{B})$ and $V^\dagger V = \mathbb{1}_\mathsf{A}$.
3. Removing a system: $\rho_{\mathsf{QR}} \mapsto \mathrm{Tr}_\mathsf{R}\rho_{\mathsf{QR}}$.

Proof. That (2) is CP is a special case of Proposition 6.4. By decomposing σ as a convex combination of pure states we see that (1) can be written as a positive linear combination of isometries, so complete positivity follows from Props. 6.4 and 6.3. Looking at the third characterisation of the partial trace in the previous handout, we see that its complete positivity also follows from Props. 6.4 and 6.3. Adding a system is clearly TP; that isometric evolutions are TP follows from cyclicity of trace and $V^\dagger V = \mathbb{1}_\mathsf{Q}$. Partial trace is trace preserving simply because $\mathrm{Tr}_\mathsf{Q}\mathrm{Tr}_\mathsf{R} = \mathrm{Tr}_{\mathsf{QR}}$. □

Definition 6.6. For any Hilbert space $\mathcal{H}_\mathsf{A}$, we define a linear map $\mathrm{vec}_\mathsf{A} : \mathcal{L}(\mathcal{H}_\mathsf{A}, \mathbb{C}) \to \mathcal{H}_\mathsf{A}$ by its action on the computational basis:

$$\text{For } a \in \{0, \ldots, d_\mathsf{A} - 1\}, \mathrm{vec}_\mathsf{A} : \langle a|_\mathsf{A} \mapsto |a\rangle_\mathsf{A}.$$

vec_A is a bijection, whose inverse is $\mathrm{vec}_\mathsf{A}^{-1} : |a\rangle_\mathsf{A} \mapsto \langle a|_\mathsf{A}$.

Note that $\mathrm{vec}_{\mathsf{A}}\langle\psi|_{\mathsf{A}} = |\psi\rangle^*_{\mathsf{A}}$! If we apply $\mathrm{vec}_{\mathsf{A}}$ to $|i\rangle_{\mathsf{B}}\langle j|_{\mathsf{A}} \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}}) \cong \mathcal{H}_{\mathsf{B}} \otimes \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathbb{C})$, we get $\mathrm{vec}_{\mathsf{A}}|i\rangle_{\mathsf{B}}\langle j|_{\mathsf{A}} = |i\rangle_{\mathsf{B}} \otimes |j\rangle_{\mathsf{A}}$, and this extends by linearity to an isomorphism between $\mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$ and $\mathcal{H}_{\mathsf{B}} \otimes \mathcal{H}_{\mathsf{A}}$.

Given any two systems A and A' of equal dimension d let $|\Phi^+\rangle_{\mathsf{AA}'} := \sum_{0\leq j<d} |j\rangle_{\mathsf{A}} \otimes |j\rangle_{\mathsf{A}'}$ and $\Phi^+_{\mathsf{AA}'} := |\Phi^+\rangle\langle\Phi^+|_{\mathsf{AA}'}$, and let $|\phi^+\rangle_{\mathsf{AA}'} = |\Phi^+\rangle_{\mathsf{AA}'}/\sqrt{d}$.

Definition 6.7. Given a linear map $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} : \mathcal{L}(\mathcal{H}_{\mathsf{A}}) \to \mathcal{L}(\mathcal{H}_{\mathsf{B}})$, its **operator representation** in $\mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$ is defined to be

$$\mathbf{id}^{\mathsf{A}\leftarrow\mathsf{A}'}\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\Phi^+_{\mathsf{AA}'} = \sum_{0\leq j,k<d_{\mathsf{A}}} (\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}|j\rangle\langle k|_{\mathsf{A}})_{\mathsf{B}} \otimes |j\rangle\langle k|_{\mathsf{A}}$$

where A' is a system with the same dimension as A (the $\mathbf{id}^{\mathsf{A}\leftarrow\mathsf{A}'}$ is just for relabelling).

The action of $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ can be written in terms of its operator representation M_{BA}:

$$\begin{aligned}\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_{\mathsf{A}} =& \mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\left(\sum_{0\leq j,k<d_{\mathsf{A}}} |j\rangle\langle j|_{\mathsf{A}}X_{\mathsf{A}}|k\rangle\langle k|_{\mathsf{A}}\right) &(6.1)\\ =& \sum_{0\leq j,k<d_{\mathsf{A}}} \left(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}|j\rangle\langle k|_{\mathsf{A}}\right)_{\mathsf{B}} \mathrm{Tr}|j\rangle\langle k|_{\mathsf{A}}X_{\mathsf{A}}^{\mathrm{T}} = \mathrm{Tr}_{\mathsf{A}}M_{\mathsf{BA}}\mathbb{1}_{\mathsf{B}} \otimes X_{\mathsf{A}}^{\mathrm{T}} &(6.2)\end{aligned}$$

So, we have an isomorphism between the vector spaces $\mathcal{L}(\mathcal{L}(\mathcal{H}_{\mathsf{A}}), \mathcal{L}(\mathcal{H}_{\mathsf{B}}))$ and $\mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$ (sometimes called "Channel-state duality" or the "Choi-Jamiołkowski isomorphism").

Proposition 6.8. The map $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} : X_{\mathsf{A}} \mapsto ZX_{\mathsf{A}}Z^\dagger$, where $Z \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$, has the operator representation $|\zeta\rangle\langle\zeta|_{\mathsf{BA}} \in \mathcal{L}(\mathcal{H}_{\mathsf{B}} \otimes \mathcal{H}_{\mathsf{A}})$ where $|\zeta\rangle_{\mathsf{BA}} = \mathrm{vec}_{\mathsf{A}}Z$.

Proof. If $Z = \sum_{0\leq b<d_{\mathsf{B}},0\leq a<d_{\mathsf{A}}} z_{ba}|b\rangle_{\mathsf{B}}\langle a|_{\mathsf{A}}$, then $|\zeta\rangle_{\mathsf{BA}} = \mathrm{vec}_{\mathsf{A}}(Z) = \sum_{b,a} z_{ba}|b\rangle_{\mathsf{B}} \otimes |a\rangle_{\mathsf{A}}$, and $\mathrm{Tr}_{\mathsf{A}}|\zeta\rangle\langle\zeta|_{\mathsf{BA}}\mathbb{1}_{\mathsf{B}} \otimes X_{\mathsf{A}}^{\mathrm{T}} = \mathrm{Tr}_{\mathsf{A}}\left[\sum_{b,a,b',a'} z_{ba}z^*_{b'a'}|b\rangle\langle b'|_{\mathsf{B}} \otimes |a\rangle\langle a'|_{\mathsf{A}}X_{\mathsf{A}}^{\mathrm{T}}\right]$
$= \sum_{b,a,b',a'} z_{ba}z^*_{b'a'}|b\rangle\langle b'|_{\mathsf{B}}\langle a|_{\mathsf{A}}X_{\mathsf{A}}|a'\rangle_{\mathsf{A}} = \left(\sum_{b,a} z_{ba}|b\rangle_{\mathsf{B}}\langle a|_{\mathsf{A}}\right) X_{\mathsf{A}} \left(\sum_{b',a'} z^*_{ba}|a'\rangle_{\mathsf{A}}\langle b'|_{\mathsf{B}}\right)$. □

Proposition 6.9 (Representations of CP maps.)**.** Let M_{BA} be the operator representation of a map $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$. The following statements are equivalent:

1. $M_{\mathsf{BA}} \geq 0$.
2. There is a set of maps $\{K_j \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}}) : j \in \{1, \ldots, n\}\}$ such that $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_{\mathsf{A}} = \sum_{j=1}^n K_jX_{\mathsf{A}}K_j^\dagger$.
3. There is a system E and map $Z \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{E}} \otimes \mathcal{H}_{\mathsf{B}})$ such that $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_{\mathsf{A}} = \mathrm{Tr}_{\mathsf{E}}ZX_{\mathsf{A}}Z^\dagger$.
4. $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ is completely positive.

Proof. For (1) $\implies$ (2), we know that $M_{\mathsf{BA}} = \sum_j |\kappa_j\rangle\langle\kappa_j|_{\mathsf{BA}}$ for some $|\kappa_j\rangle_{\mathsf{BA}} \in \mathcal{H}_{\mathsf{B}} \otimes \mathcal{H}_{\mathsf{A}}$ (from the eigendecomposition of M_{BA}, for instance). So, using Proposition 6.8,

$$\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_{\mathsf{A}} = \mathrm{Tr}_{\mathsf{A}}M_{\mathsf{BA}}\mathbb{1}_{\mathsf{B}} \otimes X_{\mathsf{A}}^{\mathrm{T}} = \sum_j \mathrm{Tr}_{\mathsf{A}}|\kappa_j\rangle\langle\kappa_j|_{\mathsf{BA}}\mathbb{1}_{\mathsf{B}} \otimes X_{\mathsf{A}}^{\mathrm{T}} = \sum_j K_jX_{\mathsf{A}}K_j^\dagger$$

where $K_j = \text{vec}_{\mathsf{A}}^{-1}|\kappa_j\rangle_{\mathsf{BA}}$. For (1) $\implies$ (3), we use that $M_{\mathsf{BA}} = \text{Tr}_{\mathsf{E}}|\zeta\rangle\langle\zeta|_{\mathsf{EBA}}$ for some $|\zeta\rangle_{\mathsf{EBA}} \in \mathcal{H}_{\mathsf{E}} \otimes \mathcal{H}_{\mathsf{B}} \otimes \mathcal{H}_{\mathsf{A}}$. Again using Proposition 6.8,

$$\begin{aligned}\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_{\mathsf{A}} =& \text{Tr}_{\mathsf{A}}M_{\mathsf{BA}}\mathbb{1}_{\mathsf{B}} \otimes X_{\mathsf{A}}^{\text{T}} = \text{Tr}_{\mathsf{A}}(\text{Tr}_{\mathsf{E}}|\zeta\rangle\langle\zeta|_{\mathsf{EBA}})\mathbb{1}_{\mathsf{B}} \otimes X_{\mathsf{A}}^{\text{T}} &(6.3)\\ =& \text{Tr}_{\mathsf{E}}\text{Tr}_{\mathsf{A}}|\zeta\rangle\langle\zeta|_{\mathsf{EBA}}\mathbb{1}_{\mathsf{EB}} \otimes X_{\mathsf{A}}^{\text{T}} = \text{Tr}_{\mathsf{E}}ZX_{\mathsf{A}}Z^{\dagger} &(6.4)\end{aligned}$$

where $Z \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{E}} \otimes \mathcal{H}_{\mathsf{B}}) = \text{vec}_{\mathsf{A}}^{-1}|\zeta\rangle_{\mathsf{EBA}}$.

(2) $\implies$ (4) by Propositions 6.3 and 6.4. (3) $\implies$ (4) follows from the fact that isometries and partial traces are CP and from the composition of CP maps being CP. That (4) $\implies$ (1) is immediate from the definitions of complete positivity and the operator representation. □

An expression of the form $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_{\mathsf{A}} = \sum_j K_j X_{\mathsf{A}} K_j^{\dagger}$ is known as a **Kraus decomposition**, and the K_j are called **Kraus operators**, for $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$. An expression of the form $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_{\mathsf{A}} = \text{Tr}_{\mathsf{E}}ZX_{\mathsf{A}}Z^{\dagger}$ is known as a **Stinespring representation** for $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$.

Proposition 6.10. Given $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} : X_{\mathsf{A}} \mapsto \sum_{j=1}^{n} K_j X_{\mathsf{A}} K_j^{\dagger} = \text{Tr}_{\mathsf{E}}ZX_{\mathsf{A}}Z^{\dagger}$ where $K_j \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$ and $Z \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}} \otimes \mathcal{H}_{\mathsf{E}})$ then

$$\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \text{ is trace preserving } \iff \sum_{j=1}^{n} K_j^{\dagger}K_j = \mathbb{1}_{\mathsf{A}} \iff Z \text{ is an isometry.}$$

♣♣ Prove this.

Remark 6.11. [Stinespring representation of an operation] Propositions 6.9 and 6.10 tell us that any operation (CPTP map) $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ can be written $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}X_{\mathsf{A}} = \text{Tr}_{\mathsf{E}}VX_{\mathsf{A}}V^{\dagger}$ for some *isometry* $V \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}} \otimes \mathcal{H}_{\mathsf{E}})$.

Proposition 6.12. Any isometric evolution with $V \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$ can be written

$$VX_{\mathsf{A}}V^{\dagger} = \text{Tr}_{\mathsf{A}}U_{\mathsf{AB}}X_{\mathsf{A}} \otimes |0\rangle\langle 0|_{\mathsf{B}}U_{\mathsf{AB}}^{\dagger} \tag{6.5}$$

where U_{AB} is a unitary in $\mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$.

Proof. Since V is isometry, it can be written $V = \sum_{0\le j<d_{\mathsf{A}}} |v_j\rangle_{\mathsf{B}}\langle j|_{\mathsf{A}}$ where $\{|v_j\rangle_{\mathsf{B}} : 0 \le j < d_{\mathsf{A}}\}$ is an orthonormal set. Therefore, $\{|0\rangle_{\mathsf{A}} \otimes |v_j\rangle_{\mathsf{B}} : 0 \le j < d_{\mathsf{A}}\}$ is an orthonormal set in $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$. It is always possible to extend an orthonormal set to an orthonormal basis. Let $\mathfrak{B} = \{|0\rangle_{\mathsf{A}} \otimes |v_j\rangle_{\mathsf{B}} : 0 \le j < d_{\mathsf{A}}\} \cup \{|u_{kj}\rangle_{\mathsf{AB}} : 1 \le k < d_{\mathsf{B}}, 0 \le j < d_{\mathsf{A}}\}$ be such an extension, and let $U_{\mathsf{AB}} = \sum_{0\le j<d_{\mathsf{A}}} |0\rangle_{\mathsf{A}} \otimes |v_j\rangle_{\mathsf{B}}\langle j|_{\mathsf{A}} \otimes \langle 0|_{\mathsf{B}} + \sum_{0\le j<d_{\mathsf{A}}} \sum_{1\le k<d_{\mathsf{B}}} |u_{kj}\rangle_{\mathsf{AB}}\langle j|_{\mathsf{A}} \otimes \langle k|_{\mathsf{B}}$. This is a unitary operator in $\mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$ since it maps the computational basis for $\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ to the orthonormal basis $\mathfrak{B}$. Because $U_{\mathsf{AB}}|\psi\rangle_{\mathsf{A}} \otimes |0\rangle_{\mathsf{B}} = |0\rangle_{\mathsf{A}} \otimes (V|\psi\rangle_{\mathsf{A}})_{\mathsf{B}}$,

$$\text{Tr}_{\mathsf{A}}U_{\mathsf{AB}}|\psi\rangle\langle\psi|_{\mathsf{A}} \otimes |0\rangle\langle 0|_{\mathsf{B}}U_{\mathsf{AB}}^{\dagger} = V|\psi\rangle\langle\psi|_{\mathsf{A}}V^{\dagger}.$$

The result follows by linearity. □

From the last two results it follows that

Theorem 6.13. Any operation can be implemented by adding a system in a pure state, unitary evolution of the composite system, and removing a system.

7 Measurements

7.1 Instruments

Suppose we perform an operation $\mathcal{N}^{\tilde{\mathsf{X}}\mathsf{R}\leftarrow\mathsf{Q}}$ and then measure the system $\tilde{\mathsf{X}}$ in its computational basis (which we'll assume is labelled by the elements of some finite set $\mathcal{A}_X$) producing a result X. For each $x \in \mathcal{A}_X$, let $\mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}$ be the linear map

$$\mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}} : \rho_{\mathsf{Q}} \mapsto \mathrm{Tr}_{\tilde{\mathsf{X}}}|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{R}}(\mathcal{N}^{\tilde{\mathsf{X}}\mathsf{R}\leftarrow\mathsf{Q}}\rho_{\mathsf{Q}})|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{R}}. \tag{7.1}$$

Note that $\mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}\rho_{\mathsf{Q}} = \langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{R}}(\mathcal{N}^{\tilde{\mathsf{X}}\mathsf{R}\leftarrow\mathsf{Q}}\rho_{\mathsf{Q}})|x\rangle_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{R}}$. From the measurement postulate (for PVMs on density operators) we have

$$P_X(x) = \mathrm{Tr}_{\mathsf{R}}\mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}\rho_{\mathsf{Q}} \tag{7.2}$$

and, given $X = x$ the state of $\tilde{\mathsf{X}}\mathsf{R}$ is $|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}\rho_{\mathsf{Q}}/P_X(x)$, so the state of R is

$$\frac{\mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}\rho_{\mathsf{Q}}}{P_X(x)}. \tag{7.3}$$

The map $x \mapsto \mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}$ is an *instrument*.

Definition 7.1. We can represent a measurement on Q whose result takes values in $\mathcal{A}_X$ and which leaves behind a system R by an **instrument** $\mathcal{I}$, which is a map $\mathcal{I} : \mathcal{A}_X \to \mathcal{L}(\mathcal{L}(\mathcal{H}_{\mathsf{Q}}), \mathcal{L}(\mathcal{H}_{\mathsf{R}}))$ such that

1. For all $x \in \mathcal{A}_X$, $\mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}$ is completely positive;
2. $\sum_{x\in\mathcal{A}_X} \mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}$ is trace preserving.

The distribution of the result X is given by (7.2) while the state of R immediately after the measurement, given $X = x$ is (7.3).

Any instrument $\mathcal{I} : \mathcal{A}_X \to \mathcal{L}(\mathcal{L}(\mathcal{H}_{\mathsf{Q}}), \mathcal{L}(\mathcal{H}_{\mathsf{R}}))$ can be implemented by following an operation $\mathcal{N}^{\tilde{\mathsf{X}}\mathsf{R}\leftarrow\mathsf{Q}}$ by a computational basis measurement on $\tilde{\mathsf{X}}$ as described above. A suitable choice of $\mathcal{N}^{\tilde{\mathsf{X}}\mathsf{R}\leftarrow\mathsf{Q}}$ is simply $\mathcal{N}^{\tilde{\mathsf{X}}\mathsf{R}\leftarrow\mathsf{Q}} = \sum_{x\in\mathcal{A}_X} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}$. Note that

$$\mathcal{N}^{\tilde{\mathsf{X}}\mathsf{R}\leftarrow\mathsf{Q}} : \rho_{\mathsf{Q}} \mapsto \sum_{x\in\mathcal{A}_X} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathcal{I}(x)^{\mathsf{R}\leftarrow\mathsf{Q}}\rho_{\mathsf{Q}}. \tag{7.4}$$

For example, the measurement of a PVM $E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ on a system Q is represented by the instrument $\mathcal{I} : \mathcal{A}_X \to \mathcal{L}(\mathcal{L}(\mathcal{H}_{\mathsf{Q}}), \mathcal{L}(\mathcal{H}_{\mathsf{Q}}))$ such that

$$\mathcal{I}(x)^{\mathsf{Q}\leftarrow\mathsf{Q}} : \rho_{\mathsf{Q}} \to E(x)_{\mathsf{Q}}\rho_{\mathsf{Q}}E(x)_{\mathsf{Q}}. \tag{7.5}$$

7.2 POVMs

If we are not interested in the post-measurement state, but only in the distribution of the result, then a measurement is completely specified by its associated POVM (positive-operator-valued measure). In fact, any linear function taking states to probability distributions can be represented by a POVM.

Definition 7.2. We can represent a measurement on a system Q, whose result takes values in $\mathcal{A}_X$, by a **POVM** E, which is a map $E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_\mathsf{Q}) : x \mapsto E(x)$ such that

1. For all $x \in \mathcal{A}_X$, $E(x) \geq 0$ and
2. $\sum_{x\in\mathcal{A}_X} E(x) = \mathbb{1}_\mathsf{Q}$.

If the state of Q is ρ then $P_X(x) = \mathrm{Tr}E(x)\rho$.

We can write down the POVM for any instrument in terms of "adjoint maps".

Definition 7.3. On a space of operators $\mathcal{L}(\mathcal{H}_\mathsf{Q})$ we define the **Hilbert-Schmidt (HS) inner product** by $\langle L_\mathsf{Q}, J_\mathsf{Q}\rangle := \mathrm{Tr}L_\mathsf{Q}^\dagger J_\mathsf{Q}$. (Note: there is an exercise in the "Postulates..." handout which asks you to show that it is indeed an inner product).

Definition 7.4. The **adjoint map** of $\mathcal{M} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_\mathsf{Q}), \mathcal{L}(\mathcal{H}_\mathsf{R}))$ is its hermitian adjoint w.r.t. the Hilbert-Schmidt inner product. That is, it is the unique map $\mathcal{M}^\dagger \in \mathcal{L}(\mathcal{L}(\mathcal{H}_\mathsf{R}), \mathcal{L}(\mathcal{H}_\mathsf{Q}))$ such that, for all $L_\mathsf{R} \in \mathcal{L}(\mathcal{H}_\mathsf{R})$, $J_\mathsf{Q} \in \mathcal{L}(\mathcal{H}_\mathsf{Q})$, $\langle L_\mathsf{R}, \mathcal{M}J_\mathsf{Q}\rangle = \langle \mathcal{M}^\dagger L_\mathsf{R}, J_\mathsf{Q}\rangle$.

Proposition 7.5. For any $\mathcal{M}, \mathcal{N} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_\mathsf{Q}), \mathcal{L}(\mathcal{H}_\mathsf{R}))$:

1. $(\alpha\mathcal{M} + \beta\mathcal{N})^\dagger = \alpha^*\mathcal{M}^\dagger + \beta^*\mathcal{N}^\dagger$ for all $\alpha, \beta \in \mathbb{C}$.
2. If $\mathcal{M} : J_\mathsf{Q} \mapsto \sum_j K_j J_\mathsf{Q} K_j^\dagger$, where $K_j \in \mathcal{L}(\mathcal{H}_\mathsf{Q}, \mathcal{H}_\mathsf{R})$, then $\mathcal{M}^\dagger : L_\mathsf{R} \mapsto \sum_j K_j^\dagger L_\mathsf{R} K_j$.
3. $\mathcal{M}$ is completely positive iff $\mathcal{M}^\dagger$ is completely positive.
4. $\mathcal{M}$ is trace preserving iff $\mathcal{M}^\dagger$ is **unital**, which means that $\mathcal{M}^\dagger\mathbb{1}_\mathsf{R} = \mathbb{1}_\mathsf{Q}$.
5. $(\mathcal{M}^\dagger)^\dagger = \mathcal{M}$.

Proof. (1) follows from the properties of inner products. (2) follows from definition of the HS inner product and the cyclicity and linearity of trace. (3) follows from (2) and the characterisation of CP maps as those with Kraus decompositions. (4) $\mathrm{Tr}J_\mathsf{Q} = \langle \mathbb{1}_\mathsf{Q}, J_\mathsf{Q}\rangle$ and $\mathrm{Tr}\mathcal{M}J_\mathsf{Q} = \langle \mathbb{1}_\mathsf{R}, \mathcal{M}J_\mathsf{Q}\rangle = \langle \mathcal{M}^\dagger\mathbb{1}_\mathsf{R}, J_\mathsf{Q}\rangle$. Therefore, $\mathrm{Tr}J_\mathsf{Q} = \mathrm{Tr}\mathcal{M}J_\mathsf{Q}$ for all $J_\mathsf{Q} \in \mathcal{L}(\mathcal{H}_\mathsf{Q})$ iff $\langle \mathbb{1}_\mathsf{Q}, J_\mathsf{Q}\rangle = \langle \mathcal{M}^\dagger\mathbb{1}_\mathsf{R}, J_\mathsf{Q}\rangle$ for all J_Q, and this last statement is equivalent to $\mathcal{M}^\dagger\mathbb{1}_\mathsf{R} = \mathbb{1}_\mathsf{Q}$. □

Proposition 7.6 (POVM for an instrument)**.** Given an instrument $\mathcal{I} : \mathcal{A}_X \to \mathcal{L}(\mathcal{L}(\mathcal{H}_\mathsf{Q}), \mathcal{L}(\mathcal{H}_\mathsf{R}))$ the corresponding POVM is the map $E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_\mathsf{Q})$ with

$$E(x)_\mathsf{Q} = \mathcal{I}(x)^\dagger\mathbb{1}_\mathsf{R}. \tag{7.6}$$

Proof. Since $\mathcal{I}(x)$ is CP, its adjoint is too, and since $\mathbb{1}_\mathsf{R} \geq 0$, $E(x)_\mathsf{Q} \geq 0$. $\mathcal{N} := \sum_{x\in\mathcal{A}_X} \mathcal{I}(x)$ is trace preserving, so $\mathcal{N}^\dagger$ is unital, and $\sum_{x\in\mathcal{A}_X} E(x)_\mathsf{Q} = \sum_{x\in\mathcal{A}_X} \mathcal{I}(x)^\dagger\mathbb{1}_\mathsf{R} = \mathcal{N}^\dagger\mathbb{1}_\mathsf{R} = \mathbb{1}_\mathsf{Q}$. Therefore, (7.6) indeed defines a POVM, and

$$P_X(x) = \mathrm{Tr}_\mathsf{R}\mathcal{I}(x)\rho_\mathsf{Q} = \langle \mathbb{1}_\mathsf{R}, \mathcal{I}(x)\rho_\mathsf{Q}\rangle = \langle \mathcal{I}(x)^\dagger\mathbb{1}_\mathsf{R}, \rho_\mathsf{Q}\rangle = \mathrm{Tr}E(x)_\mathsf{Q}\rho_\mathsf{Q}$$

so its result has the same distribution as the instrument, for any state ρ_Q. □

Proposition 7.7. For any POVM $E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_\mathsf{Q})$ we can find some (non-unique) instrument with that POVM.

Proof. It is easy to check that setting

$$\mathcal{I}(x)^{\mathsf{Q}\leftarrow\mathsf{Q}} : \rho_\mathsf{Q} \mapsto E(x)_\mathsf{Q}^{1/2} \rho_\mathsf{Q} E(x)_\mathsf{Q}^{1/2}$$

defines an instrument $\mathcal{I} : \mathcal{A}_X \to \mathcal{L}(\mathcal{L}(\mathcal{H}_\mathsf{Q}), \mathcal{L}(\mathcal{H}_\mathsf{Q}))$ with the given POVM. □

The POVM for a PVM $E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_\mathsf{Q})$ is simply E itself. It is important to note that when we only give a POVM for a measurement, this does not give us enough information to determine the post-measurement states.

7.3 Summary of measurement representations

We have now seen all the ways of representing a measurement that we will encounter in this course: PVMs, Instruments and POVMs.

Name	Form	$P_X(x)$	State given $X = x$
PVM	$E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_\mathsf{Q})$, where $E(x)^\dagger E(x') = \delta_{x'x} E(x)$, and $\sum_{x\in\mathcal{A}_X} E(x) = \mathbb{1}$.	$\mathrm{Tr} E(x)\rho$	$E(x)\rho E(x)/P_X(x)$
Instrument	$\mathcal{I} : \mathcal{A}_X \to \mathcal{L}(\mathcal{L}(\mathcal{H}_\mathsf{Q}), \mathcal{L}(\mathcal{H}_\mathsf{R}))$, where $\mathcal{I}(x)$ is CP for all $x \in \mathcal{A}_X$, and $\sum_{x\in\mathcal{A}_X} \mathcal{I}(x)$ is TP.	$\mathrm{Tr}_\mathsf{R} \mathcal{I}(x)\rho$	$\mathcal{I}(x)\rho/P_X(x)$
POVM	$E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_\mathsf{Q})$, where $E(x) \geq 0$ for all $x \in \mathcal{A}_X$, and $\sum_{x\in\mathcal{A}_X} E(x) = \mathbb{1}$.	$\mathrm{Tr} E(x)\rho$	unspecified

Table 7.1: For measurements on system Q in state ρ_Q with result X we give the general form of the representation, the distribution P_X of the result, and the state immediately after the measurement when $X = x$.

8 State discrimination

8.1 Minimum error state discrimination

Suppose we know that a system Q is in the state $\rho(X)$ where X is a random variable taking values in $\mathcal{A}_X$. That is, we have a density operator $\rho(x)$ for each $x \in \mathcal{A}_X$ and we know that, with probability $P_X(x)$, the state of Q is $\rho(x)$. (For example, perhaps $\mathcal{A}_X = \{0, 1, 2\}$, $P_X(0) = 2/3, P_X(1) = 1/6, P_X(2) = 1/6$, and the states of Q are $\rho(0) = |+\rangle\langle+|, \rho(1) = |0\rangle\langle 0|, \rho(2) = |-\rangle\langle -|$, where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$.

We are interested in how well we can determine X by measuring the system. If $\hat{X}$ is the result of measuring some POVM $E : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_\mathsf{Q})$ then

$$\Pr(\hat{X} = X) = \sum_{x \in \mathcal{A}_X} \Pr(\hat{X} = x, X = x) = \sum_{x \in \mathcal{A}_X} \Pr(X = x) \Pr(\hat{X} = x | X = x) \tag{8.1}$$

$$= \sum_{x \in \mathcal{A}_X} P_X(x) \mathrm{Tr} E(x) \rho(x). \tag{8.2}$$

Maximising this success probability over all POVMs E, (i.e. E such that $E(x) \geq 0$ for all x and $\sum_x E(x) = \mathbb{1}$) is a type of optimisation called a *semidefinite program*, which can be solved efficiently on a computer (in time polynomial in k and d_Q) but which does not, in general, have a closed form solution. However, for the case $k = 2$, the **Holevo-Helstrom theorem** gives us a closed form solution for the maximum probability and a POVM which achieves it.

8.2 The Holevo-Helstrom theorem

8.2.1 Mathematical preliminaries

As usual, here we are assuming that we are dealing with operators on finite dimensional, complex Hilbert spaces.

1. Any operator of the form $L^\dagger L$ is positive. Any positive operator A has a unique positive square root $A^{1/2}$ such that $(A^{1/2})^\dagger A^{1/2} = A$ and $A \geq 0$. If $A = \sum_j \lambda_j |\alpha_j\rangle\langle\alpha_j|$ is an eigendecomposition, then $A^{1/2} = \sum_j \lambda_j^{1/2} |\alpha_j\rangle\langle\alpha_j|$.

2. If operators A and M satisfy $A \geq 0$ and $M \geq 0$ then $\mathrm{Tr} AM = \mathrm{Tr} A^{1/2} M A^{1/2} \geq 0$ because $A^{1/2} M A^{1/2} \geq 0$. But note that AM is not necessarily positive or even hermitian.

3. For any operator J, $|J| := (J^\dagger J)^{1/2}$ and the **trace norm** of J is $\|J\|_1 := \mathrm{Tr}|J|$.

4. The support of J is the orthogonal complement of $\ker(J)$.

5. If $A \in \mathcal{L}(\mathcal{H}_\mathsf{Q})$ is hermitian then A_+ and A_- are the unique operators such that $A_+ \geq 0$, $A_- \geq 0$, $A_+ - A_- = A$, and $A_+ A_- = 0$.

6. If $A \in \mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ is hermitian, with eigendecomposition $A = \sum_{j=1}^{d_{\mathsf{Q}}} \lambda_j |\alpha_j\rangle\langle\alpha_j|$, then
   (a) $A_+ = \sum_{j:\lambda_j>0} \lambda_j |\alpha_j\rangle\langle\alpha_j|$.
   (b) $A_- = -\sum_{j:\lambda_j<0} \lambda_j |\alpha_j\rangle\langle\alpha_j|$ (note the minus sign).
   (c) $|A| = A_+ + A_-$.
   (d) $\|A\|_1 = \sum_{j=1}^{d_{\mathsf{Q}}} |\lambda_j|$.
   (e) The projector onto $\mathrm{supp}(A_+)$ is $\sum_{j:\lambda_j>0} |\alpha_j\rangle\langle\alpha_j|$.
   (f) The projector onto $\mathrm{supp}(A_-)$ is $\sum_{j:\lambda_j<0} |\alpha_j\rangle\langle\alpha_j|$.
   (g) The projector onto $\ker(A)$ is $\sum_{j:\lambda_j=0} |\alpha_j\rangle\langle\alpha_j|$.
   (h) The projector onto $\mathrm{supp}(A)$ is $\sum_{j:\lambda_j\neq 0} |\alpha_j\rangle\langle\alpha_j|$.

Proposition 8.1. For any hermitian operator A, $\|A\|_1 = \max\{\mathrm{Tr}AT : -\mathbb{1} \leq T \leq \mathbb{1}\}$ and the maximum is attained when $T = \Pi_+ + \Pi_0 - \Pi_-$, where Π_+, Π_- and Π_0 are the projectors onto $\mathrm{supp}(A_+)$, $\mathrm{supp}(A_-)$ and $\ker(A)$ respectively.

Proof. The constraints on T are equivalent to $\mathbb{1} - T \geq 0$ and $\mathbb{1} + T \geq 0$. Because $A_+ \geq 0$, $A_- \geq 0$, by item 2 above, for any T which satisfies the constraints we have

$$\begin{aligned}\mathrm{Tr}AT =& \mathrm{Tr}(A_+ + A_-)T & (8.3)\\ =& \mathrm{Tr}A_+ + \mathrm{Tr}A_- - \mathrm{Tr}A_+(\mathbb{1} - T) - \mathrm{Tr}A_-(\mathbb{1} + T) \leq \mathrm{Tr}A_+ + \mathrm{Tr}A_- = \|A\|_1. & (8.4)\end{aligned}$$

Now let $T' = \Pi_+ + \Pi_0 - \Pi_-$. Since $\mathbb{1} = \Pi_+ + \Pi_0 + \Pi_-$, we have $\mathbb{1} - T' = 2\Pi_- \geq 0$ and $\mathbb{1} + T' = 2\Pi_+ + 2\Pi_0 \geq 0$, so T' satisfies the constraints in the maximisation and has

$$\begin{aligned}\mathrm{Tr}(A_+ - A_-)T' =& \mathrm{Tr}A_+(\Pi_+ + \Pi_0) + \mathrm{Tr}A_-\Pi_- - \mathrm{Tr}A_-(\Pi_+ + \Pi_0) - \mathrm{Tr}A_+\Pi_- & (8.5)\\ =& \mathrm{Tr}A_+ + \mathrm{Tr}A_-. & (8.6)\end{aligned}$$

□

With this fact in hand, it is relatively easy to give a simple formula for the maximum success probability for discriminating between two states of a system.

8.2.2 The Holevo-Helstrom theorem

Theorem 8.2 (The Holevo-Helstrom theorem)**.** Suppose we know that a system Q is in one of two states $\rho(0)$ or $\rho(1)$. If the state is $\rho(X)$ where X is a random variable with $\mathcal{A}_X = \{0, 1\}$, and $\hat{X}$ is the result of measuring some POVM $E : \{0, 1\} \to \mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ then

$$\max_{\text{POVMs } E} \Pr(\hat{X} = X) = \frac{1}{2}\left(1 + \|\Delta\|_1\right), \text{ where } \Delta = P_X(1)\rho(1) - P_X(0)\rho(0), \tag{8.7}$$

and this is attained by the POVM with $E(1) = \Pi_+ + \Pi_0$, $E(0) = \Pi_-$ where Π_+, Π_- and Π_0 are the projectors onto $\mathrm{supp}(\Delta_+)$, $\mathrm{supp}(\Delta_-)$ and $\ker(\Delta)$, respectively.

Proof. Since $E(0) + E(1) = \mathbb{1}$, we can write $E(1) = (\mathbb{1} + T)/2$ and $E(0) = (\mathbb{1} - T)/2$ where T is the hermitian operator $E(1) - E(0)$. In terms of T,

$$\begin{aligned}\Pr(\hat{X} = X) =& \frac{1}{2}P_X(0)\mathrm{Tr}(\mathbb{1} - T)\rho(0) + \frac{1}{2}P_X(1)\mathrm{Tr}(\mathbb{1} + T)\rho(1) & (8.8)\\ =& \frac{1}{2}\left(P_X(0)\mathrm{Tr}\rho(0) + P_X(1)\mathrm{Tr}\rho(1) + \mathrm{Tr}T\Delta\right) = \frac{1}{2}\left(1 + \mathrm{Tr}T\Delta\right). & (8.9)\end{aligned}$$

Since $E(1) \geq 0$ iff $T \geq -\mathbb{1}$ and $E(0) \geq 0$ iff $T \leq \mathbb{1}$, maximising over valid POVMs is equivalent to maximising over T such that $-\mathbb{1} \leq T \leq \mathbb{1}$. Therefore, according to Proposition 8.1, $\max_{\text{POVMs } E} \Pr(\hat{X} = X) = \frac{1}{2}(1 + \|\Delta\|_1)$ which is attained when $T = \Pi_+ + \Pi_0 - \Pi_-$ or, equivalently, when $E(1) = \Pi_+ + \Pi_0$, and $E(0) = \Pi_-$. □

8.3 Example

Let X be uniformly distributed bit i.e. $\mathcal{A}_X = \{0, 1\}$, $P_X(0) = P_X(1) = 1/2$. Suppose that system Q is a qubit, which we know is in the state $\rho(X)$ where

$$\rho(0) = \eta[0] = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \rho(1) = \eta[\pi/2] = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}.$$

Here the $\{\eta[\phi] : \phi \in \mathbb{R}\}$ are the pure states on the equator of the Bloch sphere defined in Handout 2. Figure 8.1 shows the position of the two states on the equator of the Bloch sphere.

8.3.1 Minimum error state discrimination

First, let us compute the maximum value of $\Pr(\hat{X} = X)$ which can be achieved by measuring a POVM $E : \{0, 1\} \to \mathcal{L}(\mathcal{H}_\mathsf{Q})$ to obtain an estimate $\hat{X}$ of X. The operator Δ which appears in the formula (8.7) is

$$\Delta = P_X(1)\rho(1) - P_X(0)\rho(0) = \frac{-1}{4}\begin{pmatrix} 0 & 1+i \\ 1-i & 0 \end{pmatrix}$$

Computing the eigenvectors and eigenvalues of Δ we find that it has the eigendecomposition

$$\Delta = \frac{1}{2\sqrt{2}}\eta[3\pi/4] - \frac{1}{2\sqrt{2}}\eta[-\pi/4], \quad \text{so } \Delta_- = \frac{1}{2\sqrt{2}}\eta[-\pi/4], \Delta_+ = \frac{1}{2\sqrt{2}}\eta[3\pi/4], \tag{8.10}$$

$$|\Delta| = \frac{1}{2\sqrt{2}}(\eta[-\pi/4] + \eta[3\pi/4]) = \frac{1}{2\sqrt{2}}\mathbb{1}, \text{ and } \|\Delta\|_1 = 1/\sqrt{2}. \tag{8.11}$$

Therefore, the Holevo-Helstrom theorem tells us that the POVM E' with

$$E'(0) = \eta[-\pi/4] \text{ and } E'(1) = \eta[3\pi/4]$$

is optimal and achieves $\Pr(\hat{X} = X) = (1+1/\sqrt{2})/2$. Since the POVM elements are rank-one, trace-one, positive operators we can represent them points on the Bloch sphere in Figure 8.1, along with the states.

8.3.2 Unambiguous state discrimination

Now suppose that instead we measure a POVM $F : \{0, 1, ?\} \to \mathcal{L}(\mathcal{H}_\mathsf{Q})$ of the form

$$F(0) = a\eta[-\pi/2], \; F(1) = a\eta[\pi], \; F(?) = \mathbb{1} - F(0) - F(1),$$

where a is some positive real number, obtaining a result Y. It is easy to compute

$$P_{Y|X}(0|0) = \text{Tr}F(0)\rho(0) = a/2, \; P_{Y|X}(1|0) = \text{Tr}F(1)\rho(0) = 0, \tag{8.12}$$

$$P_{Y|X}(0|1) = \text{Tr}F(0)\rho(1) = 0, \; P_{Y|X}(1|1) = \text{Tr}F(1)\rho(1) = a/2, \tag{8.13}$$

and it follows that

$$P_Y(0) = P_{YX}(0,0) + P_{YX}(0,1) = P_{Y|X}(0|0)P_X(0) + P_{Y|X}(0|1)P_X(1) = a/4, \tag{8.14}$$
$$P_Y(1) = P_{YX}(1,0) + P_{YX}(1,1) = P_{Y|X}(1|0)P_X(0) + P_{Y|X}(1|1)P_X(1) = a/4, \tag{8.15}$$
$$P_Y(?) = 1 - a/2. \tag{8.16}$$

So, if $Y = 0$ or $Y = 1$ then we know *for sure* that $X = Y$, even though the states we are discriminating are not orthogonal! This comes at the expense of having probability $1 - a/2$ that $Y = ?$. How large can we make a? The only constraint is that $F(?) \geq 0$ which is true iff

$$a(\eta[-\pi/2] + \eta[\pi]) \leq \mathbb{1}.$$

This is equivalent to saying that the largest eigenvalue of $a(\eta[-\pi/2] + \eta[\pi])$ must be less than or equal to one. The characteristic polynomial of

$$\eta[-\pi/2] + \eta[\pi] = \begin{pmatrix} 1 & (i-1)/2 \\ -(1+i)/2 & 1 \end{pmatrix}$$

is $(1-\lambda)^2 - (i-1)(i+1)/4 = 0$ so the eigenvalues are $\lambda_1 = 1 - 1/\sqrt{2}$ and $\lambda_2 = 1 + 1/\sqrt{2}$. So, the largest we can make a is $1/(1 + 1/\sqrt{2}) \approx 0.59$ for which $P_Y(?) = 1/\sqrt{2} \approx 0.71$.

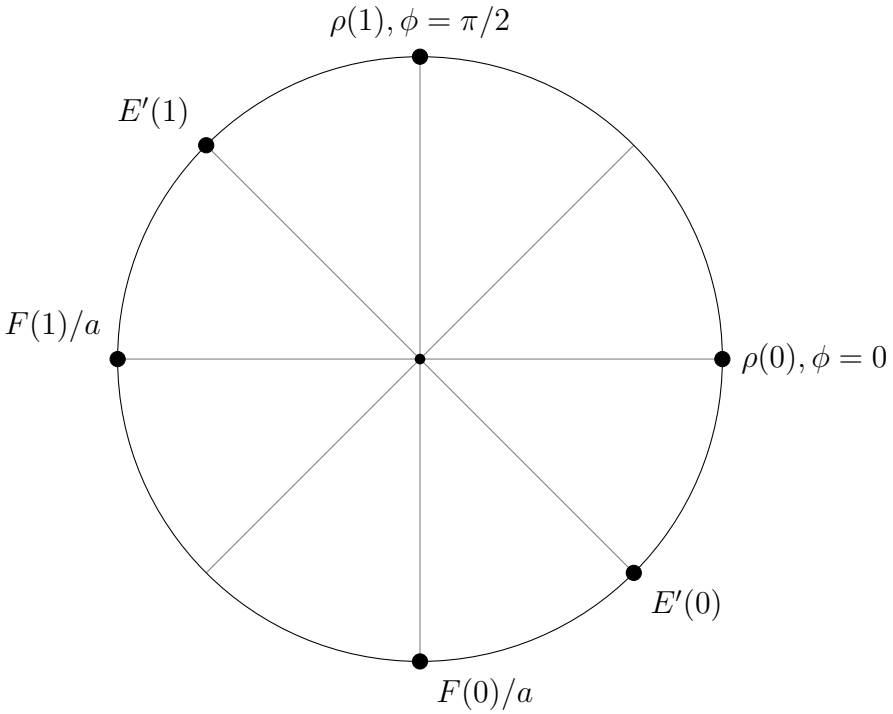


Figure 8.1: States and POVM elements described in the Section 8.3.

9 Entanglement

When we have a composite system, the pure states of that system can not, in general, be written as a tensor product of pure states of the subsystems. We say that such states are *entangled*. When we talk about a **bipartite system** we mean a composite of two systems. We have already seen an example of an entangled state of a bipartite system AB, namely the state

$$|\phi^+\rangle_{\mathsf{AB}} = (|0\rangle_{\mathsf{A}} \otimes |0\rangle_{\mathsf{B}} + |1\rangle_{\mathsf{A}} \otimes |1\rangle_{\mathsf{B}})/\sqrt{2}$$

which was used in the quantum strategy for the CHSH game. We will see later that without entanglement, Alice and Bob can do no better than the best classical strategy in any game like the CHSH game. We will also study a number of other quite different uses for entanglement.

9.1 The Schmidt decomposition

In this section, we will prove that any *pure state of a bipartite system* can be written in a certain standard form called a **Schmidt decomposition**. The Schmidt decomposition makes certain features of pure states of bipartite systems apparent, such as the fact that the states of the two parts have the same eigenvalues. It also lets us determine when two pure states of a bipartite system can be reversibly transformed into one another by *local* operations.

Note that we can think of a composite of any number of systems as being bipartite if we specify a partition of its subsystems into two sets. Such a partition is called a **bipartition**. For example, we might consider the bipartition $\mathsf{A} : \mathsf{BR}$ of a composite system ABR.

Lemma 9.1. Suppose $L \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$ and $W_{\mathsf{A}} := L^\dagger L$ has eigendecomposition

$$W_{\mathsf{A}} = \sum_{j=1}^{d_{\mathsf{A}}} \lambda_j |\alpha_j\rangle\langle\alpha_j|_{\mathsf{A}} \text{ where } \lambda_1 \geq \ldots \geq \lambda_{d_{\mathsf{A}}},$$

and $r := \operatorname{rank}(W_{\mathsf{A}})$. Then

$$L = \sum_{j=1}^{r} \lambda_j^{1/2} |\phi_j\rangle_{\mathsf{B}}\langle\alpha_j|_{\mathsf{A}}$$

for some orthonormal set $\{|\phi_j\rangle_{\mathsf{B}} : 1 \leq j \leq r\}$.

Proof. From $W_{\mathsf{A}} \geq 0$ and the ordering of the eigenvalues λ_j we know that $\lambda_j > 0$ for $j \leq r$ and $\lambda_j = 0$ for $j > r$. Since W_{A} is hermitian, $\{|\alpha_j\rangle_{\mathsf{A}} : 1 \leq j \leq d_{\mathsf{A}}\}$ is an orthonormal basis for $\mathcal{H}_{\mathsf{A}}$. So, $L = \sum_{j=1}^{d_{\mathsf{A}}} |\tilde{\phi}_j\rangle_{\mathsf{B}}\langle\alpha_j|_{\mathsf{A}}$ where $|\tilde{\phi}_j\rangle_{\mathsf{B}} = L|\alpha_j\rangle_{\mathsf{A}}$ and

$$\langle\tilde{\phi}_k|\tilde{\phi}_j\rangle = \langle\alpha_k|L^\dagger L|\alpha_j\rangle = \langle\alpha_k|W_{\mathsf{A}}|\alpha_j\rangle = \lambda_j \delta_{kj}.$$

Letting $|\phi_j\rangle_{\mathsf{B}} := \lambda^{-1/2}|\tilde{\phi}_j\rangle_{\mathsf{B}}$ for $1 \leq j \leq r$ defines an orthonormal set, and the result follows. □

Remark 9.2. Recall that $\mathrm{vec}_{\mathsf{B}}|\psi\rangle_{\mathsf{A}}\langle\phi|_{\mathsf{B}} = |\psi\rangle_{\mathsf{A}} \otimes |\phi\rangle^*_{\mathsf{B}}$.

Theorem 9.3 (Schmidt decomposition)**.** Any vector $|\psi\rangle_{\mathsf{AB}}$ in the Hilbert space of a bipartite system AB has a **Schmidt decomposition** $|\psi\rangle_{\mathsf{AB}} = \sum_{j=1}^{r} \lambda_j^{1/2} |\alpha_j\rangle_{\mathsf{A}} \otimes |\beta_j\rangle_{\mathsf{B}}$ where:

1. The λ_j are real, strictly positive and decreasing $\lambda_1 \geq \ldots \geq \lambda_r > 0$.
2. $\{|\alpha_j\rangle_{\mathsf{A}} : 1 \leq j \leq r\}$ and $\{|\beta_j\rangle_{\mathsf{B}} : 1 \leq j \leq r)\}$ are orthonormal sets;
3. $r = \mathrm{rank}(\psi_{\mathsf{A}})$ and $\sum_{1\leq j\leq r} \lambda_j |\alpha_j\rangle\langle\alpha_j|_{\mathsf{A}}$ is an eigendecomposition of $\psi_{\mathsf{A}} := \mathrm{Tr}_{\mathsf{B}}|\psi\rangle\langle\psi|_{\mathsf{AB}}$.
4. $r = \mathrm{rank}(\psi_{\mathsf{B}})$ and $\sum_{1\leq j\leq r} \lambda_j |\beta_j\rangle\langle\beta_j|_{\mathsf{B}}$ is an eigendecomposition of $\psi_{\mathsf{B}} := \mathrm{Tr}_{\mathsf{A}}|\psi\rangle\langle\psi|_{\mathsf{AB}}$.

The numbers $\lambda_j^{1/2}$ are the (non-zero) **Schmidt coefficients** of $|\psi\rangle_{\mathsf{AB}}$, and r is the **Schmidt rank** of $|\psi\rangle_{\mathsf{AB}}$. If $|\psi\rangle_{\mathsf{AB}}$ is a state vector (i.e. a unit vector) then $\sum_{1\leq j\leq r} \lambda_j = 1$.

Proof. We expand $|\psi\rangle_{\mathsf{AB}}$ in the computational basis,

$$|\psi\rangle_{\mathsf{AB}} = \sum_{a=0}^{d_{\mathsf{A}}-1} \sum_{b=0}^{d_{\mathsf{B}}-1} x_{ab} |a\rangle_{\mathsf{A}} \otimes |b\rangle_{\mathsf{B}}, \text{ and let } L^\dagger := \mathrm{vec}_{\mathsf{B}}^{-1}|\psi\rangle_{\mathsf{AB}} = \sum_{a,b} x_{ab}|a\rangle_{\mathsf{A}}\langle b|_{\mathsf{B}}.$$

Using $\mathrm{Tr}_{\mathsf{B}}|b\rangle\langle b'|_{\mathsf{B}} = \langle b'|b\rangle = \delta_{b'b} = \langle b|b'\rangle$, we have

$$\psi_{\mathsf{A}} = \mathrm{Tr}_{\mathsf{B}}|\psi\rangle\langle\psi|_{\mathsf{AB}} = \sum_{a,b}\sum_{a',b'} x_{ab}x^*_{a'b'}|a\rangle_{\mathsf{A}}\langle a'|_{\mathsf{A}}\mathrm{Tr}_{\mathsf{B}}|b\rangle\langle b'|_{\mathsf{B}} = L^\dagger L. \tag{9.1}$$

Let $\psi_{\mathsf{A}} = \sum_{j=1}^{r} \lambda_j |\alpha_j\rangle\langle\alpha_j|_{\mathsf{A}}$ be any eigendecomposition for ψ_{A} with eigenvalues λ_j in decreasing order. By Lemma 9.1 we have $L = \sum_{j=1}^{r} \lambda_j^{1/2} |\phi_j\rangle_{\mathsf{B}}\langle\alpha_j|_{\mathsf{A}}$, where $r = \mathrm{rank}(\psi_{\mathsf{A}})$ and $\{|\phi_j\rangle_{\mathsf{B}} : 0 \leq j \leq r\}$ is an orthonormal set. Therefore, $L^\dagger = \sum_{j=1}^{r} \lambda_j^{1/2} |\alpha_j\rangle_{\mathsf{A}}\langle\phi_j|_{\mathsf{B}}$ and

$$|\psi\rangle_{\mathsf{AB}} = \mathrm{vec}_{\mathsf{A}} L^\dagger = \sum_{j=1}^{r} \lambda_j^{1/2} |\alpha_j\rangle_{\mathsf{A}} \otimes |\beta_j\rangle_{\mathsf{B}}, \text{ where } |\beta_j\rangle_{\mathsf{B}} := |\phi_j\rangle^*_{\mathsf{B}}.$$

Since complex conjugation of vectors preserves orthogonality and norm, we have established points (1) to (3). (4) follows by taking the partial trace of $|\psi\rangle\langle\psi|_{\mathsf{AB}}$ over A and using the orthonormality of the $\{|\alpha_j\rangle_{\mathsf{A}}\}$. □

9.2 Mixed state entanglement

We say ρ_{AB} is a **product state** if it is a tensor product of local states $\rho_{\mathsf{AB}} = \alpha_{\mathsf{A}} \otimes \beta_{\mathsf{B}}$. A *pure* state $\rho_{\mathsf{AB}} = |\psi\rangle\langle\psi|_{\mathsf{AB}}$ is a product state iff the following equivalent conditions hold

- Its state vector[1] $|\psi\rangle_{\mathsf{AB}}$ is a product vector i.e. $|\psi\rangle_{\mathsf{AB}} = |\alpha\rangle_{\mathsf{A}} \otimes |\beta\rangle_{\mathsf{B}}$;
- Its state vector has Schmidt rank one;
- Both of its marginal states are pure.

[1] Only unique up to global phase, but the conditions given are independent of this.

Among states which are not pure, however, we regard not only product states as unentangled, but rather any convex combination of product states.

Definition 9.4. We say an operator $M_{\mathsf{AB}} \in \mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$ is **separable** (with respect to the $\mathsf{A} : \mathsf{B}$ bipartition) iff it can be written

$$M_{\mathsf{AB}} = \sum_x F(x)_{\mathsf{A}} \otimes G(x)_{\mathsf{B}}, \text{ where}$$
$$\forall x : F(x)_{\mathsf{A}} \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}), F(x)_{\mathsf{A}} \geq 0, G(x)_{\mathsf{B}} \in \mathcal{L}(\mathcal{H}_{\mathsf{B}}), G(x)_{\mathsf{B}} \geq 0.$$

We denote the set of all such operators by $\mathbf{sep}(\mathsf{A} : \mathsf{B})$. Note that a separable operator is necessarily positive.

A state (density operator) ρ_{AB} of AB belongs to $\mathbf{sep}(\mathsf{A} : \mathsf{B})$ if and only if it is a convex combination of product states, that is,

$$\rho_{\mathsf{AB}} = \sum_x P_X(x) \alpha(x)_{\mathsf{A}} \otimes \beta(x)_{\mathsf{B}} \tag{9.2}$$

where P_X is a probability distribution and the $\alpha(x)_{\mathsf{A}}$ and $\beta(x)_{\mathsf{B}}$ are density operators. Any state which is not separable, we call **entangled**.

If Alice and Bob both have access to a random variable X with distribution P_X, and Alice prepares A in the state $\alpha(X)_{\mathsf{A}}$ and Bob prepares B in the state $\beta(X)_{\mathsf{B}}$, then the state of AB will be the one given in (9.2).

9.2.1 A necessary condition for separability: PPT

Deciding whether a given state is separable or not is a computationally hard problem.[2] However, there is a simple, efficiently checkable necessary condition for separability, based on the fact that the transpose map $\mathbf{t}^{\mathsf{A} \leftarrow \mathsf{A}}$ is positive but not completely positive.

Definition 9.5. We say that an operator $M_{\mathsf{AB}} \in \mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$ is PPT (positive-partial transpose) with respect to the $\mathsf{A} : \mathsf{B}$ bipartition if $\mathbf{t}^{\mathsf{A} \leftarrow \mathsf{A}} M_{\mathsf{AB}} \geq 0$. We denote the set of positive operators which are also PPT by $\mathbf{ppt}(\mathsf{A} : \mathsf{B})$.

Remark 9.6. Since taking the (total) transpose of an operator does not change its eigenvalues, $\mathbf{t}^{\mathsf{A} \leftarrow \mathsf{A}} M_{\mathsf{AB}} \geq 0$ iff $\mathbf{t}^{\mathsf{AB} \leftarrow \mathsf{AB}} \mathbf{t}^{\mathsf{A} \leftarrow \mathsf{A}} M_{\mathsf{AB}} = \mathbf{t}^{\mathsf{B} \leftarrow \mathsf{B}} M_{\mathsf{AB}} \geq 0$. So, we could take the transpose on the B system rather than the A system in the definition and it would be equivalent. It is also easy to check that taking the transpose with respect to a different orthonormal basis doesn't change which states are PPT.

Proposition 9.7. $\mathbf{sep}(\mathsf{A} : \mathsf{B}) \subseteq \mathbf{ppt}(\mathsf{A} : \mathsf{B})$. That is, an operator which is separable with respect to a given bipartition is also PPT with respect to that bipartition.

Proof. M_{AB} is separable iff it can be written $M = \sum_j F(j)_{\mathsf{A}} \otimes G(j)_{\mathsf{B}}$ where $F(j)_{\mathsf{A}} \geq 0, G(j)_{\mathsf{B}} \geq 0$ for all j. Therefore, $F(j)_{\mathsf{A}}^{\mathrm{T}} \geq 0$ for all j and $\mathbf{t}^{\mathsf{A} \leftarrow \mathsf{A}} M_{\mathsf{AB}} = \sum_j F(j)_{\mathsf{A}}^{\mathrm{T}} \otimes G(j)_{\mathsf{B}} \geq 0$. □

Remark 9.8. It can be shown that if $d_{\mathsf{A}} + d_{\mathsf{B}} \leq 5$ then $\mathbf{sep}(\mathsf{A} : \mathsf{B}) = \mathbf{ppt}(\mathsf{A} : \mathsf{B})$.

[2] NP-hard, see e.g. `http://arxiv.org/abs/0810.4507`

10 Communication protocols using entanglement and the no-cloning theorem

10.1 Dense coding

10.1.1 The Bell basis

It will be convenient here to let $X := \sigma_x$ and $Z := \sigma_z$. These operators are hermitian and unitary. Therefore, $X^2 = \mathbb{1}$ and $Z^2 = \mathbb{1}$, and it follows that, for any $j \in \mathbb{Z}$,

$$X^j = \begin{cases} \mathbb{1} \text{ if } j \text{ is even, and} \\ X \text{ if } j \text{ is odd,} \end{cases} \tag{10.1}$$

and similarly for Z^j. Furthermore, $\mathrm{Tr}X = \mathrm{Tr}Z = \mathrm{Tr}XZ = 0$.

Given a system AB where A and B are qubits, we define for $i, j \in \{0, 1\}$

$$|\beta_{ij}\rangle_{\mathsf{AB}} := X^i_{\mathsf{A}} Z^j_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}} |\phi^+\rangle_{\mathsf{AB}}. \tag{10.2}$$

Using the properties of the Pauli X and Z operators mentioned above and the fact that $\mathrm{Tr}_{\mathsf{B}}\phi^+_{\mathsf{AB}} = \mathbb{1}_{\mathsf{A}}/d_{\mathsf{A}}$ (where $\phi^+_{\mathsf{AB}} = |\phi^+\rangle\langle\phi^+|_{\mathsf{AB}}$) we have

$$\langle\beta_{i'j'}|\beta_{ij}\rangle = \mathrm{Tr}_{\mathsf{AB}} X^i_{\mathsf{A}} Z^j_{\mathsf{A}} \phi^+_{\mathsf{AB}} Z^{j'}_{\mathsf{A}} X^{i'}_{\mathsf{A}} = \frac{1}{2}\mathrm{Tr}_{\mathsf{A}} X^i_{\mathsf{A}} Z^j_{\mathsf{A}} \mathbb{1}_{\mathsf{A}} Z^{j'}_{\mathsf{A}} X^{i'}_{\mathsf{A}} = \frac{1}{2}\mathrm{Tr}_{\mathsf{A}} X^{i+i'}_{\mathsf{A}} Z^{j+j'}_{\mathsf{A}}. \tag{10.3}$$

The operator $X^{i+i'}_{\mathsf{A}} Z^{j+j'}_{\mathsf{A}}$ has trace zero unless $i + i'$ and $j + j'$ are both even, in which case it is equal to $\mathbb{1}_{\mathsf{A}}$, which has trace 2. Since $i, j, i', j' \in \{0, 1\}$, this happens precisely when $i = i'$ and $j = j'$, so

$$\langle\beta_{i'j'}|\beta_{ij}\rangle = \delta_{i'i}\delta_{j'j}.$$

So $\{|\beta_{ij}\rangle_{\mathsf{AB}} : i, j \in \{0, 1\}\}$ is an orthonormal basis for a two qubit Hilbert space. It is known as the **Bell basis**.

10.1.2 The dense coding protocol

Suppose Alice wishes to transmit a message of two bits $M = (M_1, M_2)$, $\mathcal{A}_M = \{0, 1\}^2$, $P_M(m) = 1/4$ for all m, to Bob by sending just one qubit. The definition of the Bell basis (10.2) suggests a way to do this if Alice and Bob already possess qubits A and B, respectively, in the state ϕ^+_{AB}:

1. Alice performs unitary $X^{M_1}_{\mathsf{A}} Z^{M_2}_{\mathsf{A}}$ on A. The state of AB is now $|\beta_{M_1M_2}\rangle\langle\beta_{M_1M_2}|_{\mathsf{AB}}$.

2. Alice sends the qubit A to Bob.

3. Bob measures in the Bell basis to decode the message. That is, Bob measures the POVM $E : \{0, 1\}^2 \to \mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$ with $E(\,(i, j)\,) = |\beta_{ij}\rangle\langle\beta_{ij}|_{\mathsf{AB}}$ obtaining a result $\hat{M}$. Clearly, $\Pr(\hat{M} = M) = 1$.

10.1.3 The necessity of entanglement

Suppose Alice and Bob have systems A and B, respectively, in a state σ_{AB} and Alice wished to transmit a uniformly distributed message M to Bob by sending system A to him. The most general protocol has Alice perform some operation $\mathcal{N}(M)^{\mathsf{A}\leftarrow\mathsf{A}}$ on A depending on the message she wants to send, then send A to Bob, who measures a POVM $E : \mathcal{A}_M \to \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$ whose result $\hat{M}$ is his estimate of the message. The probability of success is

$$\Pr(\hat{M} = M) = \sum_{m\in\mathcal{A}_M} P_M(m) \Pr(\hat{M} = m|M = m) \tag{10.4}$$

$$= \sum_{m\in\mathcal{A}_M} P_M(m)\mathrm{Tr}E(m)_{\mathsf{AB}}\mathcal{N}(m)^{\mathsf{A}\leftarrow\mathsf{A}} \otimes \mathbf{id}^{\mathsf{B}\leftarrow\mathsf{B}}\sigma_{\mathsf{AB}}. \tag{10.5}$$

If σ_{AB} is separable, then

$$\sigma_{\mathsf{AB}} = \sum_x p(x)\mathcal{A}(x)_{\mathsf{A}} \otimes \beta(x)_{\mathsf{B}} \tag{10.6}$$

were p is a probability distribution and, for all x, $\alpha(x)$ and $\beta(x)$ are density operators. In this case,

$$\Pr(\hat{M} = M) = \sum_x p(x) \sum_{m\in\mathcal{A}_M} P_M(m)\mathrm{Tr}E(m)_{\mathsf{AB}}(\mathcal{N}(m)^{\mathsf{A}\leftarrow\mathsf{A}}\alpha(x)_{\mathsf{A}}) \otimes \beta(x)_{\mathsf{B}} \tag{10.7}$$

$$\leq \sum_{m\in\mathcal{A}_M} P_M(m)\mathrm{Tr}_{\mathsf{AB}}E(m)_{\mathsf{AB}}(\mathcal{N}(m)^{\mathsf{A}\leftarrow\mathsf{A}}\alpha(x^*)_{\mathsf{A}}) \otimes \beta(x^*)_{\mathsf{B}} \tag{10.8}$$

$$= \sum_{m\in\mathcal{A}_M} P_M(m)\mathrm{Tr}_{\mathsf{A}}E'(m)_{\mathsf{A}}(\mathcal{N}(m)^{\mathsf{A}\leftarrow\mathsf{A}}\alpha(x^*)_{\mathsf{A}}) \tag{10.9}$$

where x^* is the choice of x which maximises

$$\sum_{m\in\mathcal{A}_M} P_M(m)\mathrm{Tr}E(m)_{\mathsf{AB}}(\mathcal{N}(m)^{\mathsf{A}\leftarrow\mathsf{A}}\alpha(x)_{\mathsf{A}}) \otimes \beta(x)_{\mathsf{B}},$$

and E' is the POVM with

$$E'(m)_{\mathsf{A}} := \mathrm{Tr}_{\mathsf{B}}E(m)_{\mathsf{AB}}\mathbb{1}_{\mathsf{A}} \otimes \beta(x^*)_{\mathsf{B}}.$$

To see that this is a POVM we use a frequently useful fact about the partial trace

Proposition 10.1 (Partial trace cyclicity)**.** For any $J_{\mathsf{A}} \in \mathcal{L}(\mathcal{H}_{\mathsf{A}})$ and $L_{\mathsf{AB}} \in \mathcal{L}(\mathcal{H}_{\mathsf{AB}})$,

$$\mathrm{Tr}_{\mathsf{A}}J_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}}L_{\mathsf{AB}} = \sum_i (\mathrm{Tr}J_{\mathsf{A}}G_{\mathsf{A}}^{(i)}) \otimes F_{\mathsf{B}}^{(i)} = \sum_i (\mathrm{Tr}G_{\mathsf{A}}^{(i)}J_{\mathsf{A}}) \otimes F_{\mathsf{B}}^{(i)} = \mathrm{Tr}_{\mathsf{A}}L_{\mathsf{AB}}J_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}}$$

where we used the existence of a product basis for $\mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$ to write $L_{\mathsf{AB}} = \sum_i G_{\mathsf{A}}^{(i)} \otimes F_{\mathsf{B}}^{(i)}$ for some (not necessarily positive or hermitian) operators $G_{\mathsf{A}}^{(i)}$ and $F_{\mathsf{B}}^{(i)}$.

Note that it is certainly *not* generally true that $\mathrm{Tr}_{\mathsf{A}}K_{\mathsf{AB}}L_{\mathsf{AB}} = \mathrm{Tr}_{\mathsf{A}}L_{\mathsf{AB}}K_{\mathsf{AB}}$. Using partial trace cyclicity we have, for all m,

$E'(m)_{\mathsf{A}} := \mathrm{Tr}_{\mathsf{B}}E(m)_{\mathsf{AB}}\mathbb{1}_{\mathsf{A}} \otimes \beta(x^*)_{\mathsf{B}} = \mathrm{Tr}_{\mathsf{B}}\mathbb{1}_{\mathsf{A}} \otimes \beta(x^*)_{\mathsf{B}}^{1/2}E(m)_{\mathsf{AB}}\mathbb{1}_{\mathsf{A}} \otimes \beta(x^*)_{\mathsf{B}}^{1/2}$ and hence $E'(m)_{\mathsf{A}} \geq 0$.

Also, $\sum_m E'(m)_{\mathsf{A}} = \mathrm{Tr}_{\mathsf{B}}(\sum_m E(m)_{\mathsf{AB}})\mathbb{1}_{\mathsf{A}} \otimes \beta(x^*)_{\mathsf{B}} = \mathrm{Tr}_{\mathsf{B}}\mathbb{1}_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}}\mathbb{1}_{\mathsf{A}} \otimes \beta(x^*)_{\mathsf{B}} = \mathbb{1}_{\mathsf{A}}$, so E' is indeed a POVM.

Defining $\rho(m)_{\mathsf{A}} := \mathcal{N}(m)^{\mathsf{A}\leftarrow\mathsf{A}}\alpha(x^*)_{\mathsf{A}}$ we have

$$\Pr(\hat{M} = M) \leq \sum_{m\in\mathcal{A}_M} P_M(m)\mathrm{Tr}_{\mathsf{A}}E'(m)_{\mathsf{A}}\rho(m)_{\mathsf{A}}, \tag{10.10}$$

so there is a protocol where Bob measures only A which does as well as any protocol using a separable state σ_{AB}. In Example Sheet 1, Q. 13, you showed that in this situation, for a uniformly distributed message ($P_M(m) = 1/|\mathcal{A}_M|$), $\Pr(\hat{M} = M) \leq d_{\mathsf{A}}/|\mathcal{A}_M|$. So, if $M = 4$ and $d_{\mathsf{A}} = 2$ as in the superdense coding protocol, then without using entanglement, the success probability will be no more than one half.

10.2 The no-cloning theorem

Theorem 10.2 (The no-cloning theorem)**.** There is no "cloning operation" $\mathcal{C}^{\mathsf{QQ}'\leftarrow\mathsf{Q}}$ such that, for all state vectors $|\psi\rangle_{\mathsf{Q}}$,

$$\mathcal{C}^{\mathsf{QQ}'\leftarrow\mathsf{Q}}|\psi\rangle\langle\psi|_{\mathsf{Q}} = |\psi\rangle\langle\psi|_{\mathsf{Q}} \otimes |\psi\rangle\langle\psi|_{\mathsf{Q}'}$$

Proof. If $\mathcal{C}^{\mathsf{QQ}'\leftarrow\mathsf{Q}}$ is an operation, then there is an isometry $V \in \mathcal{L}(\mathcal{H}_{\mathsf{Q}}, \mathcal{H}_{\mathsf{Q}} \otimes \mathcal{H}_{\mathsf{Q}'} \otimes \mathcal{H}_{\mathsf{E}})$ such that $\mathcal{C}^{\mathsf{QQ}'\leftarrow\mathsf{Q}}|\psi\rangle\langle\psi|_{\mathsf{Q}} = \mathrm{Tr}_{\mathsf{E}}V|\psi\rangle\langle\psi|_{\mathsf{Q}}V^\dagger$. Since $\mathrm{Tr}_{\mathsf{E}}V|\psi\rangle\langle\psi|_{\mathsf{Q}}V^\dagger$ is supposed to be a pure state for any $|\psi\rangle_{\mathsf{Q}}$, we know (from the Schmidt decomposition) that $V|\psi\rangle_{\mathsf{Q}}$ must be a product state w.r.t. the QQ' / E bipartition, and should satisfy

$$V|0\rangle_{\mathsf{Q}} = |0\rangle_{\mathsf{Q}} \otimes |0\rangle_{\mathsf{Q}'} \otimes |\eta_0\rangle_{\mathsf{E}}, V|1\rangle_{\mathsf{Q}} = |1\rangle_{\mathsf{Q}} \otimes |1\rangle_{\mathsf{Q}'} \otimes |\eta_1\rangle_{\mathsf{E}} \tag{10.11}$$

for some $|\eta_0\rangle_{\mathsf{E}}$ and $|\eta_1\rangle_{\mathsf{E}}$, and, with $|+\rangle_{\mathsf{Q}} := \frac{1}{\sqrt{2}}(|0\rangle_{\mathsf{Q}} + |1\rangle_{\mathsf{Q}})$, we would like $V|+\rangle_{\mathsf{Q}}$ to be equal to $|+\rangle_{\mathsf{Q}} \otimes |+\rangle_{\mathsf{Q}'} \otimes |\eta_+\rangle_{\mathsf{E}}$, for some $|\eta_+\rangle_{\mathsf{E}}$. However, by (10.11) and the linearity of V, we have

$$V|+\rangle_{\mathsf{Q}} = \frac{1}{\sqrt{2}}(|0\rangle_{\mathsf{Q}} \otimes |0\rangle_{\mathsf{Q}'} \otimes |\eta_0\rangle_{\mathsf{E}} + |1\rangle_{\mathsf{Q}} \otimes |1\rangle_{\mathsf{Q}'} \otimes |\eta_1\rangle_{\mathsf{E}}). \tag{10.12}$$

This state is clearly not of the form that we require: In fact, the state of subsystem Q is the maximally mixed state, when it should be the pure state $|+\rangle\langle+|_{\mathsf{Q}}$! □

10.3 Teleportation

Suppose that Alice is given a system Q. She wants to transmit the state of Q to Bob, but she can only transmit classical information to him. The protocol has to work whatever the state of Q is, but Alice doesn't know which state it is.

Suppose Alice simply measures some POVM E on Q alone and sends the result M to Bob, who prepares a state $\sigma(M)_{\mathsf{B}}$. The overall operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{Q}}$ takes states of Q to states of B is

$$\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{Q}} : \rho_{\mathsf{Q}} \mapsto \sum_{m\in\mathcal{A}_M} \sigma(m)_{\mathsf{B}}\mathrm{Tr}_{\mathsf{Q}}E(m)_{\mathsf{Q}}\rho_{\mathsf{Q}}.$$

An operation of this form is called a **measure-prepare operation**. The procedure works iff $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{Q}} = \mathbf{id}^{\mathsf{B}\leftarrow\mathsf{Q}}$. But if this were true, then Alice could copy the message M and send it to Bob (with system B) and Bertha (with system B') who could also prepare $\sigma(M)$. The operation from Q to $\mathsf{B}'\mathsf{B}$

$$\mathcal{C}^{\mathsf{B}'\mathsf{B}\leftarrow\mathsf{Q}} : \rho_{\mathsf{Q}} \mapsto \sum_{m\in\mathcal{A}_M} \rho(m)_{\mathsf{B}'} \otimes \rho(m)_{\mathsf{B}}\mathrm{Tr}_{\mathsf{Q}}E(m)_{\mathsf{Q}}\rho_{\mathsf{Q}}$$

would be a cloning operation, and we know that these don't exist!

10.3.1 Teleporting the state of a qubit

Suppose that $d_{\mathsf{Q}} = 2$ and that Alice and Bob start with qubits A and B, respectively, in the state ϕ^+_{AB}. Alice is given a qubit Q whose state she must transmit to Bob. Consider the following protocol:

- Alice measures the Bell basis PVM on QA obtaining a result $M = (M_1, M_2)$. To be precise, we mean the PVM with

$$E(\,(i,j)\,) = |\beta_{ij}\rangle\langle\beta_{ij}|_{\mathsf{QA}} = (X^i_{\mathsf{Q}} Z^j_{\mathsf{Q}} \otimes \mathbb{1}_{\mathsf{A}})\phi^+_{\mathsf{QA}}(Z^j_{\mathsf{Q}} X^i_{\mathsf{Q}} \otimes \mathbb{1}_{\mathsf{A}}). \tag{10.13}$$

- She sends M to Bob.

- Bob performs the unitary $X^{M_1}_{\mathsf{B}} Z^{M_2}_{\mathsf{B}}$ on his qubit.

We claim that the overall operation from Q to B is $\mathbf{id}^{\mathsf{B}\leftarrow\mathsf{Q}}$. Let's work out the state of B conditional on $M = (i,j)$, which we will call $\rho^{(i,j)}_{\mathsf{B}}$, immediately after Alice measures. To do this we will use the "transpose trick", which you were asked to prove in example sheet 1. A proof is given in the solutions for that sheet.

Proposition 10.3 (Transpose trick)**.** If $d_{\mathsf{Q}} = d_{\mathsf{R}}$ and $J_{\mathsf{Q}} \in \mathcal{L}(\mathcal{H}_{\mathsf{Q}})$ then

$$J_{\mathsf{Q}} \otimes \mathbb{1}_{\mathsf{R}} |\phi^+\rangle_{\mathsf{QR}} = \mathbb{1}_{\mathsf{Q}} \otimes J^T_{\mathsf{R}} |\phi^+\rangle_{\mathsf{QR}} \text{ where } J_{\mathsf{R}} = \mathbf{id}^{\mathsf{R}\leftarrow\mathsf{Q}} J_{\mathsf{Q}}. \tag{10.14}$$

Using the measurement postulate and the expression (10.13) we find

$$\begin{aligned}
\Pr(M = (i,j))\rho^{(i,j)}_{\mathsf{B}} =& \mathrm{Tr}_{\mathsf{QA}} E(\,(i,j)\,)_{\mathsf{QA}} \otimes \mathbb{1}_{\mathsf{B}} \rho_{\mathsf{Q}} \otimes \phi^+_{\mathsf{AB}} E(\,(i,j)\,)_{\mathsf{QA}} \otimes \mathbb{1}_{\mathsf{B}} && (10.15)\\
=& \mathrm{Tr}_{\mathsf{QA}} \rho_{\mathsf{Q}} \otimes \phi^+_{\mathsf{AB}} E(\,(i,j)\,)_{\mathsf{QA}} \otimes \mathbb{1}_{\mathsf{B}} && (10.16)\\
=& \mathrm{Tr}_{\mathsf{QA}} (Z^j_{\mathsf{Q}} X^i_{\mathsf{Q}} \rho_{\mathsf{Q}} X^i_{\mathsf{Q}} Z^j_{\mathsf{Q}}) \otimes \phi^+_{\mathsf{AB}} \phi^+_{\mathsf{QA}} \otimes \mathbb{1}_{\mathsf{B}} && (10.17)
\end{aligned}$$

Let J_{Q} be any operator on $\mathcal{H}_{\mathsf{Q}}$. Using the transpose trick twice we have

$$J_{\mathsf{Q}} \otimes \phi^+_{\mathsf{AB}} \phi^+_{\mathsf{QA}} \otimes \mathbb{1}_{\mathsf{B}} = \mathbb{1}_{\mathsf{Q}} \otimes \phi^+_{\mathsf{AB}} \mathbb{1}_{\mathsf{Q}} \otimes J^T_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}} \phi^+_{\mathsf{QA}} \otimes \mathbb{1}_{\mathsf{B}} = \mathbb{1}_{\mathsf{Q}} \otimes \phi^+_{\mathsf{AB}} \phi^+_{\mathsf{QA}} \otimes J_{\mathsf{B}} \tag{10.18}$$

where $J_{\mathsf{A}} := \mathbf{id}^{\mathsf{A}\leftarrow\mathsf{Q}} J_{\mathsf{Q}}$ and $J_{\mathsf{B}} := \mathbf{id}^{\mathsf{B}\leftarrow\mathsf{Q}} J_{\mathsf{Q}}$. Futhermore, we note that

$$\mathrm{Tr}_{\mathsf{QA}} \mathbb{1}_{\mathsf{Q}} \otimes \phi^+_{\mathsf{AB}} \phi^+_{\mathsf{QA}} \otimes J_{\mathsf{B}} = \frac{1}{d_{\mathsf{A}}} \mathrm{Tr}_{\mathsf{A}} \phi^+_{\mathsf{AB}} \mathbb{1}_{\mathsf{A}} \otimes J_{\mathsf{B}} = \frac{1}{d_{\mathsf{A}} d_{\mathsf{B}}} J_{\mathsf{B}}.$$

Taking $J_{\mathsf{Q}} = Z^j_{\mathsf{Q}} X^i_{\mathsf{Q}} \rho_{\mathsf{Q}} X^i_{\mathsf{Q}} Z^j_{\mathsf{Q}}$ in (10.17) and using these facts we have

$$\Pr(M = (i,j))\rho^{(i,j)}_{\mathsf{B}} = \frac{1}{4} Z^j_{\mathsf{B}} X^i_{\mathsf{B}} \rho_{\mathsf{B}} X^i_{\mathsf{B}} Z^j_{\mathsf{B}},$$

so, for all $(i,j) \in \{0,1\}^2$, $\Pr(M = (i,j)) = 1/4$ and $\rho^{(i,j)}_{\mathsf{B}} = Z^j_{\mathsf{B}} X^i_{\mathsf{B}} \rho_{\mathsf{B}} X^i_{\mathsf{B}} Z^j_{\mathsf{B}}$. By applying the unitary operation, Bob always ends up with the state $\rho_{\mathsf{B}} = \mathbf{id}^{\mathsf{B}\leftarrow\mathsf{Q}} \rho_{\mathsf{Q}}$ as described.

10.4 Comments on the teleportation protocol

1. In the protocol, whatever the state ρ_{Q}, the result M of Alice's measurement on QA is uniformly distributed. Therefore M, by itself, doesn't tell us anything about the state of Q.

2. Without conditioning on the result of Alice's measurement, the state of B is the same immediately after the measurement as it was before. Namely, it is $\mathrm{Tr}_{\mathsf{A}}\phi^+_{\mathsf{AB}}$, the maximally mixed state. On learning M, Bob knows that the state of B is $Z_{\mathsf{B}}^{M_1}X_{\mathsf{B}}^{M_2}\rho_{\mathsf{B}}X_{\mathsf{B}}^{M_2}Z_{\mathsf{B}}^{M_1}$, which allows him to apply an appropriate unitary operation to recover ρ_{B}

11 Cloning and superluminal communication

11.1 Cloning allows superluminal communication

Suppose Alice has system A and Bob has system B and the state of AB in the state ϕ^+_{AB}. Suppose that if $M=0$ Alice measures A in the computational basis, and if $M=1$, Alice measures the PVM E_1 on A

$$E_1(0) = |+\rangle\langle+|_{\mathsf{A}}, E_1(1) = |-\rangle\langle-|_{\mathsf{A}}.$$

Let us call the result of Alice's measurement X. Suppose that Bob did have a cloning machine and that Alice and Bob use clocks to arrange that Bob's operations occur after Alice's measurement (note that in special relativity the only way this even makes sense is if the two events separated by a time-like interval).

At a time when he knows Alice has measured, Bob applies his cloning machine to B. If $M=0$, then Bob's state is

$$\frac{1}{2}\left(|0\rangle\langle 0|^{\otimes 2} + |1\rangle\langle 1|^{\otimes 2}\right) = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

while if $M=1$, then Bob's state is

$$\frac{1}{2}\left(|+\rangle\langle+|^{\otimes 2} + |-\rangle\langle-|^{\otimes 2}\right) = \begin{pmatrix} \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

These states are not the same, so the success probability to distinguish them is better than one half. If Bob applies his cloning machine a greater number of times to obtain n copies of his state after Alice's measurement, then probability that he can successfully determine whether $M=0$ or $M=1$ from a measurement of the n-fold copy approaches one.

An argument of this kind was used in a proposal for a superluminal signalling system by Nick Herbert (Foundations of Physics Volume 12, Issue 12 , pp 1171-1179). There is a nice summary of this paper here `http://www.scientificamerican.com/article/mistakes-faster-than-light-telegraph-that-wasnt/`

11.2 Quantum mechanics doesn't allow superluminal communication

Suppose Alice has system A and Bob has system B and the state of AB is ρ_{AB}. Now suppose that Alice performs a measurement on A which yields a result X and leaves her with quantum system A'. We can represent this by an instrument $\mathcal{I}^{\mathsf{A}'\leftarrow\mathsf{A}}$. The probability distribution of the result is

$$P_X(x) = \mathrm{Tr}_{\mathsf{AB}}\mathcal{I}(x)^{\mathsf{A}'\leftarrow\mathsf{A}} \otimes \mathbf{id}^{\mathsf{B}\leftarrow\mathsf{B}}\rho_{\mathsf{AB}} = \mathrm{Tr}_{\mathsf{A}}\mathcal{I}(x)^{\mathsf{A}'\leftarrow\mathsf{A}}\mathrm{Tr}_{\mathsf{B}}\rho_{\mathsf{AB}},$$

so it only depends on the state of A, $\mathrm{Tr}_{\mathsf{B}}\rho_{\mathsf{AB}}$. Conditional on the result having a particular value, say $X = x$, the state of $\mathsf{A}'\mathsf{B}$ is $\mathcal{I}(x)^{\mathsf{A}'\leftarrow\mathsf{A}} \otimes \mathbf{id}^{\mathsf{B}\leftarrow\mathsf{B}}\rho_{\mathsf{AB}}/P_X(x)$ and the state of B, which we denote by $\rho(x)_{\mathsf{B}}$, is

$$\rho(x)_{\mathsf{B}} = \mathrm{Tr}_{\mathsf{A}'}\mathcal{I}(x)^{\mathsf{A}'\leftarrow\mathsf{A}} \otimes \mathbf{id}^{\mathsf{B}\leftarrow\mathsf{B}}\rho_{\mathsf{AB}}/P_X(x).$$

Suppose we have some linear map $\mathcal{N}^{\mathsf{A}'\leftarrow\mathsf{A}} : \mathcal{L}(\mathcal{H}_{\mathsf{A}}) \to \mathcal{L}(\mathcal{H}_{\mathsf{A}'})$ and operator $M_{\mathsf{AB}} \in \mathcal{L}(\mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}})$. As always, we can write $M_{\mathsf{AB}} = \sum_x L_{\mathsf{A}}^{(x)} \otimes K_{\mathsf{B}}^{(x)}$, for some operators $L^{(x)}$ and $K^{(x)}$ (which are not necessarily positive). Let $E_{\mathsf{A}} = \mathcal{N}^\dagger \mathbb{1}_{\mathsf{A}'}$.

$$\mathrm{Tr}_{\mathsf{A}'}\mathcal{N}^{\mathsf{A}'\leftarrow\mathsf{A}} \otimes \mathbf{id}^{\mathsf{B}\leftarrow\mathsf{B}}M_{\mathsf{AB}} = \sum_x \mathrm{Tr}_{\mathsf{A}'}(\mathcal{N}^{\mathsf{A}'\leftarrow\mathsf{A}}L_{\mathsf{A}}^{(x)}) \otimes K_{\mathsf{B}}^{(x)} = \sum_x (\mathrm{Tr}\mathcal{N}^{\mathsf{A}'\leftarrow\mathsf{A}}L_{\mathsf{A}}^{(x)})K_{\mathsf{B}}^{(x)}$$
$$= \sum_x (\mathrm{Tr}E_{\mathsf{A}}L_{\mathsf{A}}^{(x)})K_{\mathsf{B}}^{(x)} = \mathrm{Tr}_{\mathsf{A}}\sum_x E_{\mathsf{A}}L_{\mathsf{A}}^{(x)} \otimes K_{\mathsf{B}}^{(x)} = \mathrm{Tr}_{\mathsf{A}}E_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}}M_{\mathsf{AB}}.$$

Therefore,

$$\rho(x)_{\mathsf{B}} = \mathrm{Tr}_{\mathsf{A}}E(x)_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}}\rho_{\mathsf{AB}}/P_X(x),$$

where $E(x)_{\mathsf{A}} = \mathcal{I}(x)^\dagger \mathbb{1}_{\mathsf{A}'}$, for all x. That is, E is the POVM for the instrument $\mathcal{I}$. Note that in this situation, where we want to know the state of a subsystem conditional on the outcome of a measurement on a *different* subsystem, the state only depends on the POVM for the measurement.

Without conditioning on the result of the measurement, then the state of B after the measurement is

$$\sum_x P_X(x)\rho(x)_{\mathsf{B}} = \sum_x \mathrm{Tr}_{\mathsf{A}}E(x)_{\mathsf{A}} \otimes \mathbb{1}_{\mathsf{B}}\rho_{\mathsf{AB}} = \mathrm{Tr}_{\mathsf{A}}\rho_{\mathsf{AB}},$$

because $\sum_x E(x)_{\mathsf{A}} = \mathbb{1}_{\mathsf{A}}$. This is the same as the state of B before the measurement. We conclude that if Bob knows nothing about the result of Alice's measurement, then he has no way of knowing which measurement was performed, or even if Alice performed any measurement at all. Since the predictions of quantum mechanics in the scenario described here are independent of Alice and Bob's location, this conclusion is reassuring - otherwise quantum mechanics would be predicting some form of superluminal communication!

12 Fidelity

Definition 12.1. The **fidelity** of two states ρ and σ is defined by

$$F(\rho,\sigma) := \mathrm{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}} = \|\sigma^{1/2}\rho^{1/2}\|_1.$$

It is a measure of how similar the states are.

Remark 12.2. For any isometry $V_{\mathsf{B}\leftarrow\mathsf{A}}$ and positive operator M_{A}, $(VMV^\dagger)^{1/2} = VM^{1/2}V^\dagger$. It follows that, for any operator L_{A}, $\|VLV^\dagger\|_1 = \|L\|_1$ and for any states $\rho_{\mathsf{A}}, \sigma_{\mathsf{A}}$, $F(V\rho V^\dagger, V\sigma V^\dagger) = F(\rho,\sigma)$.

Remark 12.3. When one or both of the states is pure, the fidelity simplifies, $F(|\psi\rangle\langle\psi|,\sigma) = (\langle\psi|\sigma|\psi\rangle)^{1/2}$ and $F(|\psi\rangle\langle\psi|,|\phi\rangle\langle\phi|) = |\langle\psi|\phi\rangle|$.

For example the fidelity between the maximally mixed state $\mathbb{1}_{\mathsf{Q}}/d_{\mathsf{Q}}$ and any pure state $|\psi\rangle\langle\psi|_{\mathsf{Q}}$ is $d_{\mathsf{Q}}^{-1/2}$. It turns out that the fidelity of two states is equal to the largest absolute value of the inner-product between state vectors corresponding to purifications of the two states. This result is called "Uhlmann's theorem" and it makes it easy to derive a number of properties of the fidelity.

Theorem 12.4 (Uhlmann's theorem)**.** For any R with $\dim\mathsf{R} \geq \dim\mathsf{Q}$,

$$F(\rho_{\mathsf{Q}},\sigma_{\mathsf{Q}}) = \max\{|\langle\psi|\phi\rangle| : \mathrm{Tr}_{\mathsf{R}}|\psi\rangle\langle\psi|_{\mathsf{QR}} = \rho_{\mathsf{Q}}, \mathrm{Tr}_{\mathsf{R}}|\phi\rangle\langle\phi|_{\mathsf{QR}} = \sigma_{\mathsf{Q}}\}.$$

To prove Uhlmann's theorem, we'll need a few definitions and results which we'll find other uses for. Let's denote by $\mathcal{U}(\mathcal{H})$ the set of all unitary operators in $\mathcal{L}(\mathcal{H})$

Lemma 12.5. Any $L \in \mathcal{L}(\mathcal{H})$ has a **polar decomposition** $L = U|L|$ for some $U \in \mathcal{U}(\mathcal{H})$.

Proof. We have $L^\dagger L = |L|^2$. Since $|L|^2 \geq 0$ it has an eigendecomposition of the form $|L|^2 = \sum_{j=1}^d \lambda_j^2|\alpha_j\rangle\langle\alpha_j|$, with $\lambda_j \geq \lambda_{j+1}$, $\lambda_j \geq 0$, and $|L| = \sum_{j=1}^d \lambda_j|\alpha_j\rangle\langle\alpha_j|$. By Lemma 1 from section 9.1, $L = \sum_{j=1}^r \lambda_j|\phi_j\rangle\langle\alpha_j|$, where $\{|\phi_j\rangle : 1 \leq j \leq r\}$ is an orthonormal set and $r = \mathrm{rank}(|L|^2) = \mathrm{rank}(|L|)$. Extending this to an orthonormal basis $\{|\phi_j\rangle : 1 \leq j \leq d\}$ and setting $U = \sum_{j=1}^d |\phi_j\rangle\langle\alpha_j|$, we have $L = U|L|$, and U is unitary. □

Definition 12.6. For any $L \in \mathcal{L}(\mathcal{H}_{\mathsf{A}}, \mathcal{H}_{\mathsf{B}})$, the **operator norm** of L is

$$\|L\|_{op} := \max_{\||\psi\rangle_{\mathsf{A}}\|\leq 1} \|L|\psi\rangle\|.$$

Proposition 12.7. The operator norm has the following properties:

1. If V is an isometry then $\|V\|_{op} = 1$.
2. $\|LR\|_{op} = \max_{\||\psi\rangle\|\leq 1} \|R|\psi\rangle\| \left\|L\frac{R|\psi\rangle}{\|R|\psi\rangle\|}\right\| \leq \|R\|_{op}\|L\|_{op}$.
3. $\|L\|_{op} = \max_{\||\phi\rangle_{\mathsf{B}}\|\leq 1,\||\psi\rangle_{\mathsf{A}}\|\leq 1} |\langle|\phi\rangle, L|\psi\rangle\rangle| = \max_{\||\phi\rangle_{\mathsf{B}}\|\leq 1,\||\psi\rangle_{\mathsf{A}}\|\leq 1} |\langle L^\dagger|\phi\rangle, |\psi\rangle\rangle| = \|L^\dagger\|_{op}$.

4. For any $U \in \mathcal{U}(\mathcal{H})$, $\|LU\|_{op} \leq \|L\|_{op}\|U\|_{op} = \|L\|_{op} = \|LUU^\dagger\|_{op} \leq \|LU\|_{op}$, so $\|LU\|_{op} = \|L\|_{op}$.

Lemma 12.8. For any $L \in \mathcal{L}(\mathcal{H})$, with polar decomposition $L = U|L|$,

$$\|L\|_1 = \max_{Z \in \mathcal{L}(\mathcal{H}):\|Z\|_{op}\leq 1} |\mathrm{Tr}ZL| = \max_{Z\in\mathcal{U}(\mathcal{H})} |\mathrm{Tr}ZL|$$

and the maximum is achieved for $Z = U^\dagger \in \mathcal{U}(\mathcal{H})$.

Proof. Making the change of variables $Z = Z'U^\dagger$, $\|Z\|_{op} = \|Z'\|_{op}$ by property (4) of the operator norm, and the RHS is equal to $\max_{\|Z'\|_{op}\leq 1} |\mathrm{Tr}Z'|L||$. We need to show this is no more than $\|L\|_1$: Let $|L| = \sum_j \lambda_j |\alpha_j\rangle\langle\alpha_j|$ be an eigendecomposition for $|L|$. Then, for any Z' with $\|Z'\|_{op} \leq 1$,

$$|\mathrm{Tr}Z'|L|| = \left|\sum_j \lambda_j \langle\alpha_j|Z'|\alpha_j\rangle\right| \leq \sum_j \lambda_j |\langle |\alpha_j\rangle, Z'|\alpha_j\rangle\rangle| \tag{12.1}$$

$$\leq \sum_j \lambda_j \||\alpha_j\rangle\|\|Z'|\alpha_j\rangle\| \leq \sum_j \lambda_j = \mathrm{Tr}|L| = \|L\|_1. \tag{12.2}$$

using the triangle inequality, Cauchy-Schwarz and $\||\alpha_j\rangle\| = 1$. Equality is achieved when $Z' = \mathbb{1}$, which means $Z = U^\dagger$. □

Proposition 12.9. If $|\psi\rangle_{\mathsf{QR}}$ and $|\psi'\rangle_{\mathsf{QR}'}$ are both purifications of a state ρ_{Q}, that is

$$\mathrm{Tr}_{\mathsf{R}}|\psi\rangle\langle\psi|_{\mathsf{QR}} = \mathrm{Tr}_{\mathsf{R}'}|\psi'\rangle\langle\psi'|_{\mathsf{QR}'} = \rho_{\mathsf{Q}} \tag{12.3}$$

and $d_{\mathsf{R}} \geq d_{\mathsf{R}'}$, then there is an isometry $V \in \mathcal{L}(\mathcal{H}_{\mathsf{R}'}, \mathcal{H}_{\mathsf{R}})$ such that

$$|\psi\rangle_{\mathsf{QR}} = V_{\mathsf{R}\leftarrow\mathsf{R}'}|\psi'\rangle_{\mathsf{QR}'}.$$

Proof. Given the equation (12.3), our proof of the Schmidt decomposition (Theorem 3, section 9.1) shows that we can write $|\psi'\rangle_{\mathsf{QR}'} = \sum_{j=1}^r \sqrt{\lambda_j}|\alpha_j\rangle_{\mathsf{Q}}\otimes|\beta'_j\rangle_{\mathsf{R}'}$ and $|\psi\rangle_{\mathsf{QR}} = \sum_{j=1}^r \sqrt{\lambda_j}|\alpha_j\rangle_{\mathsf{Q}}\otimes |\beta_j\rangle_{\mathsf{R}}$, where $\sum_{j=1}^r \lambda_j|\alpha_j\rangle\langle\alpha_j|_{\mathsf{Q}}$ is an eigendecomposition of ρ_{Q}, and where $\{|\beta'_j\rangle_{\mathsf{R}'} : 1 \leq j \leq r\}$ and $\{|\beta_j\rangle_{\mathsf{R}} : 1 \leq j \leq r\}$ are both orthonormal sets. Extending the first of these to an orthonormal *basis* $\{|\beta'_j\rangle_{\mathsf{R}'} : 1 \leq j \leq d_{\mathsf{R}'}\}$ for $\mathcal{H}_{\mathsf{R}'}$, we see that the isometry $V = \sum_{j=1}^{d_{\mathsf{R}'}} |\beta_j\rangle_{\mathsf{R}}\langle\beta'_j|_{\mathsf{R}'}$ does the job. □

12.0.1 Proof of Uhlmann's theorem

Proof. Recall that $\mathrm{Tr}_{\mathsf{R}'}\Phi^+_{\mathsf{QR}'} = \mathbb{1}_{\mathsf{Q}}$. It follows that, for any density operator μ_{Q}, $\mu_{\mathsf{Q}}^{1/2} \otimes \mathbb{1}_{\mathsf{R}'}|\Phi^+\rangle_{\mathsf{QR}'}$ is a purification of μ_{Q}, because

$$\mathrm{Tr}_{\mathsf{R}'}\mu_{\mathsf{Q}}^{1/2}\otimes\mathbb{1}_{\mathsf{R}'}\ \Phi^+_{\mathsf{QR}'}\ \mu_{\mathsf{Q}}^{1/2}\otimes\mathbb{1}_{\mathsf{R}'} = \mu_{\mathsf{Q}}^{1/2}(\mathrm{Tr}_{\mathsf{R}'}\Phi^+_{\mathsf{QR}'})\mu_{\mathsf{Q}}^{1/2} = \mu_{\mathsf{Q}}.$$

From Proposition 12.9 we know that, for $\dim \mathsf{R} \geq \dim \mathsf{Q}$, $|\psi\rangle_{\mathsf{QR}}$ is a purification of ρ_{Q} iff $|\psi\rangle_{\mathsf{QR}} = \rho_{\mathsf{Q}}^{1/2}\otimes V_{\mathsf{R}\leftarrow\mathsf{R}'}|\Phi^+\rangle_{\mathsf{QR}'}$ for some isometry V and, likewise, $|\phi\rangle_{\mathsf{QR}} = \sigma_{\mathsf{Q}}^{1/2}\otimes U_{\mathsf{R}\leftarrow\mathsf{R}'}|\Phi^+\rangle_{\mathsf{QR}'}$ for some isometry U. Using these expressions for the purifications and the "transpose trick"

$$\langle\rho|\sigma\rangle = \langle\Phi^+|\sigma_{\mathsf{Q}}^{1/2}\rho_{\mathsf{Q}}^{1/2}\otimes W_{\mathsf{R}'}^{\mathrm{T}}|\Phi^+\rangle_{\mathsf{QR}'} = \langle\Phi^+|\sigma_{\mathsf{Q}}^{1/2}\rho_{\mathsf{Q}}^{1/2}W_{\mathsf{Q}}\otimes\mathbb{1}_{\mathsf{R}'}|\Phi^+\rangle_{\mathsf{QR}'} \tag{12.4}$$

$$= \mathrm{Tr}_{\mathsf{QR}'}|\Phi^+\rangle\langle\Phi^+|_{\mathsf{QR}'}\sigma_{\mathsf{Q}}^{1/2}\rho_{\mathsf{Q}}^{1/2}W_{\mathsf{Q}}\otimes\mathbb{1}_{\mathsf{R}'} = \mathrm{Tr}_{\mathsf{Q}}\sigma_{\mathsf{Q}}^{1/2}\rho_{\mathsf{Q}}^{1/2}W_{\mathsf{Q}}, \tag{12.5}$$

where $W_{\mathsf{R}'}^{\mathrm{T}} := U^\dagger V$ and $W_{\mathsf{Q}} := \mathbf{id}^{\mathsf{Q}\leftarrow\mathsf{R}'} W_{\mathsf{R}'}$. Since U and V are isometries, from the properties of the operator norm, $\|W_{\mathsf{Q}}\|_{op} \leq 1$. Furthermore, provided $d_{\mathsf{R}} \geq d_{\mathsf{Q}}$, given any unitary W_{Q} we can find a suitable choice of U and V such that $W_{\mathsf{R}'}^{\mathrm{T}} := U^\dagger V$ (e.g. taking $V = \mathbb{1}_{\mathsf{R}\leftarrow\mathsf{R}'} W_{\mathsf{R}'}^T$ and $U = \mathbb{1}_{\mathsf{R}\leftarrow\mathsf{R}'}$ does the trick). Therefore, by Lemma 12.8,

$$\max\{|\langle\psi|\phi\rangle| : \mathrm{Tr}_{\mathsf{R}}|\psi\rangle\langle\psi|_{\mathsf{QR}} = \rho_{\mathsf{Q}}, \mathrm{Tr}_{\mathsf{R}}|\phi\rangle\langle\phi|_{\mathsf{QR}} = \sigma_{\mathsf{Q}}\} \tag{12.6}$$

$$= \max_{\|W\|_{op}\leq 1} |\mathrm{Tr}_{\mathsf{Q}}\sigma_{\mathsf{Q}}^{1/2}\rho_{\mathsf{Q}}^{1/2}W_{\mathsf{Q}}| = \|\sigma_{\mathsf{Q}}^{1/2}\rho_{\mathsf{Q}}^{1/2}\|_1 = F(\rho_{\mathsf{Q}}, \sigma_{\mathsf{Q}}). \tag{12.7}$$

□

Proposition 12.10. The fidelity has the following properties (♣♣: Prove this)

1. $F(\rho,\sigma) = F(\sigma,\rho)$.
2. $0 \leq F(\rho,\sigma) \leq 1$, and $F(\rho,\sigma) = 1$ iff $\rho = \sigma$.
3. $F(V\rho_{\mathsf{A}}V^\dagger, V\sigma_{\mathsf{A}}V^\dagger) = F(\rho,\sigma)$ for any isometry $V_{\mathsf{B}\leftarrow\mathsf{A}}$.
4. $F(\mathrm{Tr}_{\mathsf{B}}\rho_{\mathsf{AB}}, \mathrm{Tr}_{\mathsf{B}}\sigma_{\mathsf{AB}}) \geq F(\rho_{\mathsf{AB}}, \sigma_{\mathsf{AB}})$.
5. $F(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\rho_{\mathsf{A}}, \mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\sigma_{\mathsf{A}}) \geq F(\rho_{\mathsf{A}}, \sigma_{\mathsf{A}})$ for any operation $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$.
6. $F(\rho\otimes\tau, \sigma\otimes\tau) = F(\rho,\sigma)$.

12.0.2 Relationship to trace norm

The fidelity gives us a way quantify the similarity between two states. We have already seen a way to measure their distinguishability: The trace norm.

Definition 12.11. The **trace distance** between two states is the function

$$D(\rho,\sigma) := \frac{1}{2}\|\rho - \sigma\|_1. \tag{12.8}$$

The factor of $1/2$ means that $0 \leq D(\rho,\sigma) \leq 1$ for any two states.

Note that if we know that a system is either in state σ or state ρ, and each case has equal probability, then the Holevo-Helstrom theorem says that the probability of correctly identifying which state the system is in is $(1 + D(\rho,\sigma))/2$.

The fidelity and the trace distance between two states are related by the **Fuchs-van de Graaf inequalities**:

Proposition 12.12. $1 - D(\rho,\sigma) \leq F(\rho,\sigma) \leq \sqrt{1 - D(\rho,\sigma)^2}$. (♣♣: See example sheet 2.)

12.0.3 Fidelity of PPT states with ϕ^+

Proposition 12.13. Let σ_{AB} be a state of AB where $d_{\mathsf{A}} = d_{\mathsf{B}} = d$. If $\sigma_{\mathsf{AB}} \in \mathbf{ppt}(\mathsf{A}:\mathsf{B})$ then $F(\phi^+_{\mathsf{AB}}, \sigma_{\mathsf{AB}}) \leq 1/\sqrt{d}$.

Proof. It is easy to check that the transpose map $\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}$ is its own adjoint w.r.t. the Hilbert-Schmidt inner product, and it is clearly its own inverse. Since the adjoint of a tensor product of maps is the tensor product of the adjoints of the maps, the same remarks apply to the partial transposition $\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}} \otimes \mathbf{id}^{\mathsf{B}\leftarrow\mathsf{B}}$. Therefore,

$$\begin{aligned} F(\phi^+_{\mathsf{AB}}, \sigma_{\mathsf{AB}})^2 =& \mathrm{Tr}\phi^+_{\mathsf{AB}}\sigma_{\mathsf{AB}} = \langle \phi^+_{\mathsf{AB}}, \sigma_{\mathsf{AB}}\rangle = \langle \phi^+_{\mathsf{AB}}, \mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\sigma_{\mathsf{AB}}\rangle \\ =& \langle \mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\phi^+_{\mathsf{AB}}, \mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\sigma_{\mathsf{AB}}\rangle = \frac{1}{d}\mathrm{Tr}[(\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\Phi^+_{\mathsf{AB}})(\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\sigma_{\mathsf{AB}})] = \frac{1}{d}\mathrm{Tr}\mathbb{F}_{\mathsf{AB}}\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\sigma_{\mathsf{AB}} \\ \leq& \frac{1}{d}\max_{\|Z_{\mathsf{AB}}\|_{op}\leq 1} \left|\mathrm{Tr}Z_{\mathsf{AB}}\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\sigma_{\mathsf{AB}}\right| = \frac{1}{d}\|\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\sigma_{\mathsf{AB}}\|_1 = \frac{1}{d}\mathrm{Tr}\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\sigma_{\mathsf{AB}} = \frac{1}{d}. \end{aligned}$$

We used the fact that $\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}\Phi^+_{\mathsf{AB}} =: \mathbb{F}_{\mathsf{AB}} = \sum_{0\leq i,j<d} |i\rangle\langle j|_{\mathsf{A}}\otimes|j\rangle\langle i|_{\mathsf{B}}$ is the unitary 'flip' operator, and therefore has operator norm equal to one; the characterisation of the trace norm proven in handout 8; and the fact that the trace norm of a positive operator is simply its trace, and that $\mathbf{t}^{\mathsf{A}\leftarrow\mathsf{A}}$ preserves trace. □

12.1 The fidelity of an operation

Definition 12.14. For any operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ where $d_{\mathsf{A}} = d_{\mathsf{B}}$ and state ρ_{A} we define

$$\begin{aligned} F_{op}(\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}, \rho_{\mathsf{A}}) :=& \inf_{\mathsf{R},\ \rho_{\mathsf{RA}}} \{F(\mathbf{id}^{\mathsf{B}\leftarrow\mathsf{A}}\rho_{\mathsf{RA}}, \mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}\rho_{\mathsf{RA}}) : \mathrm{Tr}_{\mathsf{R}}\rho_{\mathsf{RA}} = \rho_{\mathsf{A}}\} && (12.9) \\ =& F(\mathbf{id}^{\mathsf{B}\leftarrow\mathsf{A}}\psi_{\mathsf{RA}}, \mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}\psi_{\mathsf{RA}}), && (12.10) \end{aligned}$$

where ψ_{RA} is any purification of ρ_{A}. The equality is because we can always purify ρ_{RA} without increasing the fidelity (see property 4), and since any two purifications are equivalent up to an isometry between the purifying systems, which does not change the fidelity (property 3), it doesn't matter which one we use.

F_{op} measures how well the operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ preserves the state of any *composite* system RA when the part on which the operation acts is initially in the state ρ_{A}. This quantity (or its square) is sometimes called the "entanglement fidelity". Given a Kraus decomposition for the operation, F_{op} has a simple expression in terms of the Kraus operators. For simplicity we take $\mathsf{B} = \mathsf{A}$.

Proposition 12.15. If $\mathcal{N}^{\mathsf{A}\leftarrow\mathsf{A}} : \rho_{\mathsf{A}} \mapsto \sum_m K_m\rho_{\mathsf{A}}K_m^\dagger$ then $F_{op}(\mathcal{N}^{\mathsf{A}\leftarrow\mathsf{A}}, \rho_{\mathsf{A}}) = \sqrt{\sum_m \left|\mathrm{Tr}K_m\rho_{\mathsf{A}}\right|^2}$.

Proof. Let $\psi_{\mathsf{RA}} = |\psi\rangle\langle\psi|_{\mathsf{RA}}$ be a purification of ρ_{A}.

$$\begin{aligned} F(\psi_{\mathsf{RA}}, \mathcal{N}^{\mathsf{A}\leftarrow\mathsf{A}}\psi_{\mathsf{RA}})^2 =& \langle\psi|(\mathbf{id}^{\mathsf{R}\leftarrow\mathsf{R}}\otimes\mathcal{N}^{\mathsf{A}\leftarrow\mathsf{A}}|\psi\rangle\langle\psi|_{\mathsf{RA}})|\psi\rangle_{\mathsf{RA}} && (12.11) \\ =& \sum_m |\langle\psi|\mathbb{1}_{\mathsf{R}}\otimes K_m|\psi\rangle_{\mathsf{RA}}|^2. && (12.12) \end{aligned}$$

Now, let $|\psi\rangle_{\mathsf{RA}} = \sum_k \sqrt{\lambda_k}|\phi_k\rangle_{\mathsf{R}}\otimes|\alpha_k\rangle_{\mathsf{A}}$ be a Schmidt decomposition for $|\psi\rangle_{\mathsf{RA}}$. Using the orthonormality of the $|\phi_k\rangle$:

$$\langle\psi|\mathbb{1}_{\mathsf{R}}\otimes K_m|\psi\rangle_{\mathsf{RA}} = \sum_{j,k}\sqrt{\lambda_j}\langle\phi_j|_{\mathsf{R}}\otimes\langle\alpha_j|_{\mathsf{A}}\mathbb{1}_{\mathsf{R}}\otimes K_m\sqrt{\lambda_k}|\phi_k\rangle_{\mathsf{R}}\otimes|\alpha_k\rangle_{\mathsf{A}} \quad (12.13)$$

$$= \sum_j \lambda_j\langle\alpha_j|_{\mathsf{A}}K_m|\alpha_j\rangle_{\mathsf{A}} = \mathrm{Tr}K_m\left(\sum_j\lambda_j|\alpha_j\rangle\langle\alpha_j|\right) = \mathrm{Tr}K_m\rho_{\mathsf{A}}. \quad (12.14)$$

□

Suppose that the state of A is initially $\omega(X)$ where X is a random variable which is stored in the system R. The state of RA is

$$\rho_{\mathsf{RA}} = \sum_{x \in \mathcal{A}_X} P_X(x) |x\rangle\langle x|_{\mathsf{R}} \otimes \omega(x)_{\mathsf{A}}$$

and $\rho_{\mathsf{A}} = \sum_{x \in \mathcal{A}_X} P_X(x) \omega(x)_{\mathsf{A}}$. If the system A now undergoes a time evolution represented by operation $\mathcal{N}^{\mathsf{A} \leftarrow \mathsf{A}}$, the state of RA becomes

$$\rho'_{\mathsf{RA}} = \sum_{x \in \mathcal{A}_X} P_X(x) |x\rangle\langle x|_{\mathsf{R}} \otimes \sigma(x)_{\mathsf{A}}, \text{ where } \sigma(x)_{\mathsf{A}} := \mathcal{N}^{\mathsf{A} \leftarrow \mathsf{A}} \omega(x)_{\mathsf{A}}.$$

In this case,

$$\begin{aligned} F(\rho_{\mathsf{RA}}, \rho'_{\mathsf{RA}}) =& \mathrm{Tr} \sqrt{\sum_x P_X(x)^2 |x\rangle\langle x|_{\mathsf{R}} \otimes \omega(x)^{1/2} \sigma(x) \omega(x)^{1/2}} && (12.15) \\ =& \mathrm{Tr} \sum_x P_X(x) |x\rangle\langle x|_{\mathsf{R}} \otimes \sqrt{\omega(x)^{1/2} \sigma(x) \omega(x)^{1/2}} && (12.16) \\ =& \sum_x P_X(x) F(\omega(x)_{\mathsf{A}}, \sigma(x)_{\mathsf{A}}) = \mathbb{E} F(\omega(X)_{\mathsf{A}}, \sigma(X)_{\mathsf{A}}), && (12.17) \end{aligned}$$

i.e. the expectation of the fidelity between the state of A before the operation and the state of A after the operation. The quantity $F_{op}(\mathcal{N}^{\mathsf{A} \leftarrow \mathsf{A}}, \rho_{\mathsf{A}})$ is a lower bound on this expectation.

Now, suppose that $\omega(x) = |x\rangle\langle x|$, so that prior to the operation system A also stores the random variable X

$$\rho_{\mathsf{RA}} = \sum_{x \in \mathcal{A}_X} P_X(x) |x\rangle\langle x|_{\mathsf{R}} \otimes |x\rangle\langle x|_{\mathsf{A}} \text{ and } \rho_{\mathsf{A}} = \sum_{x \in \mathcal{A}_X} P_X(x) |x\rangle\langle x|_{\mathsf{A}}.$$

Furthermore, let us suppose that the operation applied to A consists of measuring A in the computational basis, obtaining the result X, and then storing a random variable X' in A. The dependence of X' on X is described by the conditional probability distribution $P_{X'|X}$. Explicitly, this operation is

$$\mathcal{N}^{\mathsf{A} \leftarrow \mathsf{A}} : \rho_{\mathsf{A}} \mapsto \sum_{x', x} P_{X'|X}(x'|x) |x'\rangle\langle x|_{\mathsf{A}} \rho_{\mathsf{A}} |x\rangle\langle x'|_{\mathsf{A}}.$$

Note that this is a Kraus representation with Kraus operators

$$\{P_{X'|X}(x'|x)^{1/2} |x'\rangle\langle x| : x', x \in \mathcal{A}_X\}.$$

Now, $F(\omega(x)_{\mathsf{A}}, \sigma(x)_{\mathsf{A}}) = P_{X'|X}(x|x)^{1/2}$ and

$$F(\rho_{\mathsf{RA}}, \rho'_{\mathsf{RA}}) = \sum_x P_X(x) \sqrt{P_{X'|X}(x|x)} \leq \sqrt{\sum_x P_X(x) P_{X'|X}(x|x)} = \sqrt{\Pr(X' = X)},$$

where the inequality is by Jensen's inequality. So, in this situation, when X is distributed according to P_X, $\Pr(X' = X)$ is bounded below by $F_{op}(\mathcal{N}^{\mathsf{A} \leftarrow \mathsf{A}}, \rho_{\mathsf{A}})^2$.

13 Data compression

Definition 13.1 (Classical data compression)**.** Suppose we have a probability distribution p on $\mathcal{A}$ Let Z be a random variable with $\mathcal{A}_Z = \mathcal{A}$ and distribution $P_Z = p$. We denote by $s_\epsilon(p)$ the smallest number k such that there exists an

- encoding function $c : \mathcal{A} \to \{1, \ldots, k\}$ and
- decoding function $d : \{1, \ldots k\} \to \mathcal{A}$

such that, if $\hat{Z} = d(c(Z))$, then $\Pr(\hat{Z} = Z) \geq 1 - \epsilon$.

Clearly, for any set $S \subseteq \mathcal{A}$ there is a d with image S if and only if $k \geq |S|$. If d has image S, then any c has $\Pr(\hat{Z} = Z) \leq \Pr(Z \in S)$ and this bound can be attained by picking c so that $c(z)$ is an element of $d^{-1}(z)$ for all $z \in S$. These observations show that $s_\epsilon(p)$ is simply the size of the smallest possible ϵ-*sufficient set* for p, i.e.

$$s_\epsilon(p) = \min \left\{ |S| : S \subseteq \mathcal{A}, \sum_{z \in S} p(z) \geq 1 - \epsilon \right\}.$$

Remark 13.2 (Form of minimal ϵ-sufficient sets)**.** It isn't hard to see that we can obtain a minimal ϵ-sufficient set by taking elements from $\mathcal{A}$ in order of decreasing probability until those accumulated have total probability $\geq 1 - \epsilon$.

Definition 13.3 (Quantum data compression)**.** Suppose we have a density operator ρ on $\mathcal{H}$. Let Q be a quantum system, with $\mathcal{H}_\mathsf{Q} = \mathcal{H}$, which has state ρ. We denote by $s_\epsilon(\rho)$ the smallest number k such that if K is a system with $d_\mathsf{K} = k$, then there exists an

- encoding operation $\mathcal{C}^{\mathsf{K} \leftarrow \mathsf{Q}}$ and
- decoding operation $\mathcal{D}^{\mathsf{Q} \leftarrow \mathsf{K}}$

such that, $F_{op}^2(\mathcal{D}^{\mathsf{Q} \leftarrow \mathsf{K}} \mathcal{C}^{\mathsf{K} \leftarrow \mathsf{Q}}, \rho_\mathsf{Q}) \geq 1 - \epsilon$. **Note that** we are using the **square** of F_{op} here.

13.1 Relating the quantum and classical cases

Theorem 13.4. Given a density operator ρ on $\mathcal{H}$ with $\dim(\mathcal{H}) = d$, let $\rho = \sum_{k=1}^{d} p(k) |\alpha_k\rangle\langle\alpha_k|$ be an eigendecomposition of ρ (so p is a distribution on $\{1, \ldots, d\}$). Then, for any $\epsilon \in [0, 1]$,

$$s_\epsilon(p) \leq s_\epsilon(\rho) \leq s_{\epsilon/2}(p). \tag{13.1}$$

13.1.1 Proof of the upper bound in Theorem 13.4

Proposition 13.5. Let $\rho_{\mathsf{Q}} = \sum_k p(k)|\alpha_k\rangle\langle\alpha_k|_{\mathsf{Q}}$ be an eigendecomposition for ρ_{Q}, so p is a distribution on $\mathcal{A} = \{1, \ldots, d_{\mathsf{Q}}\}$. Given any $S \subseteq \mathcal{A}$, let $d_{\mathsf{K}} = |S|$ and let $V_{\mathsf{Q}\leftarrow\mathsf{K}}$ be an isometry which maps $\mathcal{H}_{\mathsf{K}}$ to $\text{span}\{|\alpha_k\rangle_{\mathsf{Q}} : k \in S\}$ and let $\Pi_{\mathsf{Q}} := VV^\dagger = \sum_{k\in S} |\alpha_k\rangle\langle\alpha_k|_{\mathsf{Q}}$. Then

$$\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}} : \sigma_{\mathsf{K}} \mapsto V\sigma_{\mathsf{K}}V^\dagger, \tag{13.2}$$

$$\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}} : \rho_{\mathsf{Q}} \mapsto V^\dagger\rho_{\mathsf{Q}}V + |0\rangle\langle 0|_{\mathsf{K}}\text{Tr}(\mathbb{1}_{\mathsf{Q}} - \Pi_{\mathsf{Q}})\rho_{\mathsf{Q}}. \tag{13.3}$$

are operations which achieve

$$F_{op}^2(\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}, \rho_{\mathsf{Q}}) \geq \left(\sum_{k\in S} p(k)\right)^2,$$

and it follows that $s_\epsilon(\rho) \leq s_{\epsilon/2}(p)$.

Proof. The isometric evolution $\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}$ is clearly an operation. The map $\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}$ is the sum of two completely positive maps and, as such, is completely positive. It is also trace preserving:

$$\text{Tr}_{\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}\rho_{\mathsf{Q}} = \text{Tr}_{\mathsf{K}}V^\dagger\rho_{\mathsf{Q}}V + \text{Tr}(\mathbb{1}_{\mathsf{Q}} - \Pi_{\mathsf{Q}})\rho_{\mathsf{Q}} = \text{Tr}_{\mathsf{Q}}VV^\dagger\rho_{\mathsf{Q}} + \text{Tr}(\mathbb{1}_{\mathsf{Q}} - \Pi_{\mathsf{Q}})\rho_{\mathsf{Q}} = \text{Tr}\rho_{\mathsf{Q}}$$

So $\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}$ is an operation, and there is a Kraus represention for $\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}$ of the form

$$\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}\rho_{\mathsf{Q}} = VV^\dagger\rho_{\mathsf{Q}}VV^\dagger + \ldots = \Pi_{\mathsf{Q}}\rho_{\mathsf{Q}}\Pi_{\mathsf{Q}} + \ldots.$$

From the expression for F_{op} in terms of Kraus operators (Proposition 12.15 in the previous handout), $F_{op}(\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}, \rho_{\mathsf{Q}})^2 \geq |\text{Tr}\Pi_{\mathsf{Q}}\rho_{\mathsf{Q}}|^2 = \left(\sum_{k\in S} p(k)\right)^2$, and we know that there is a set S of size $s_{\epsilon/2}(p)$ such that $\left(\sum_{k\in S} p(k)\right)^2 \geq (1-\epsilon/2)^2 \geq 1-\epsilon$, so we achieve this fidelity with $d_{\mathsf{K}} = s_{\epsilon/2}(p)$. □

13.1.2 Proof of the lower bound in Theorem 13.4

We start with an almost obvious fact:

Proposition 13.6. Given d real numbers λ_j for $j \in \{0, \ldots, d-1\}$, such that $\lambda_j \geq \lambda_{j+1}$ and $r \in \mathbb{N}$,

$$\max\left\{\sum_{0\leq j<d} \lambda_j t_j : 0 \leq t_j \leq 1, \sum_{0\leq j<d} t_j = r\right\} = \sum_{0\leq j<r} \lambda_j.$$

Proof. Letting $l_j := 1 - t_j$, $\sum_{0\leq j<d} t_j = r$ is equivalent to $\sum_{0\leq j<r} l_j = \sum_{r\leq j<d} t_j =: R$, and

$$\sum_{0\leq j<d} \lambda_j t_j = \sum_{0\leq j<r} \lambda_j(1-l_j) + \sum_{r\leq j<d} \lambda_j t_j = \sum_{0\leq j<r} \lambda_j + \sum_{r\leq j<d} \lambda_j t_j - \sum_{0\leq j<r} \lambda_j l_j \tag{13.4}$$

$$\leq \sum_{0\leq j<r} \lambda_j + \lambda_r R - \lambda_{r-1} R \leq \sum_{0\leq j<r} \lambda_j, \tag{13.5}$$

where the inequalities are by the ordering of the eigenvalues and the positivity of the t_j, l_j. □

Proposition 13.7. For any $M \in \mathrm{Herm}(\mathcal{H})$ with eigendecomposition $M = \sum_{0\le j<d} \lambda_j |\alpha_j\rangle\langle\alpha_j|$, where $\lambda_j \ge \lambda_{j+1}$ and $d = \dim(\mathcal{H})$,

$$\max\{\mathrm{Tr} M\Pi : \Pi^\dagger\Pi = \Pi,\ \mathrm{rank}(\Pi) = r\} = \sum_{0\le j<r} \lambda_j \tag{13.6}$$

and this maximum is achieved by the projector $\Pi = \sum_{0\le j<r} |\alpha_j\rangle\langle\alpha_j|$.

Proof. It is clear that the projector $\sum_{0\le j<r} |\alpha_j\rangle\langle\alpha_j|$ achieves the stated bound, so it remains to establish that the bound holds for *any* rank-r projector Π.

$$\mathrm{Tr}\Pi M = \sum_{0\le j<d} \lambda_j t_j \text{ where } t_j := \langle\alpha_j|\Pi|\alpha_j\rangle.$$

Since the $|\alpha_j\rangle$ form an orthonormal basis, $\sum_{0\le j<d} t_j = \mathrm{Tr}\Pi = r$, and since $0 \le \Pi \le \mathbb{1}$ we have $0 \le t_j \le 1$. Therefore, the preceding proposition tells us that

$$\mathrm{Tr}\Pi M \le \sum_{0\le j<r} \lambda_j.$$

□

You may recognise the preceding proposition as a generalisation of the Rayleigh-Ritz theorem (which corresponds to the $r = 1$ case). We shall now use it prove that $s_\epsilon(p) \le s_\epsilon(\rho)$:

Proof. Without loss of generality we may assume that p satisfies $p(k) \ge p(k+1)$. Suppose that there exist encoding and decoding operations, $\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}$ and $\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}$, such that $1 - \epsilon \le F_{op}^2(\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}, \rho_{\mathsf{Q}})$. We want to show that $s_\epsilon(p) \le d_{\mathsf{K}}$. Let

$$\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}} : \rho_{\mathsf{Q}} \mapsto \sum_i K_i \rho_{\mathsf{Q}} K_i^\dagger,\ \text{where } K_i \in \mathcal{L}(\mathcal{H}_{\mathsf{Q}}, \mathcal{H}_{\mathsf{K}}),\ \text{and} \tag{13.7}$$

$$\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}} : \sigma_{\mathsf{K}} \mapsto \sum_j L_j \sigma_{\mathsf{K}} L_j^\dagger,\ \text{where } L_j \in \mathcal{L}(\mathcal{H}_{\mathsf{K}}, \mathcal{H}_{\mathsf{Q}}), \tag{13.8}$$

be Kraus representations for these operations. The operation $\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}$ has a Kraus representation

$$\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}} : \rho_{\mathsf{Q}} \mapsto \sum_{i,j} L_j K_i \rho_{\mathsf{Q}} K_i^\dagger L_j^\dagger,\ \text{so}\ \sum_{i,j} K_i^\dagger L_j^\dagger L_j K_i = \mathbb{1}_{\mathsf{Q}}. \tag{13.9}$$

Let Π_j denote the projector onto the image of L_j, so $L_j = \Pi_j L_j$. The image of each L_j has dimension no larger than d_{K}, so $\mathrm{rank}(\Pi_j) \le d_{\mathsf{K}}$.

$$\begin{aligned}
1 - \epsilon \le & F_{op}^2(\mathcal{D}^{\mathsf{Q}\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}}, \rho_{\mathsf{Q}}) \overset{(a)}{=} \sum_{j,i} |\mathrm{Tr}\Pi_j L_j K_i \rho_{\mathsf{Q}}|^2 \overset{(b)}{=} \sum_{j,i} \left|\langle \Pi_j \rho_{\mathsf{Q}}^{1/2}, L_j K_i \rho_{\mathsf{Q}}^{1/2}\rangle\right|^2 \\
& \overset{(c)}{\le} \sum_{j,i} \langle \Pi_j \rho_{\mathsf{Q}}^{1/2}, \Pi_j \rho_{\mathsf{Q}}^{1/2}\rangle \langle L_j K_i \rho_{\mathsf{Q}}^{1/2}, L_j K_i \rho_{\mathsf{Q}}^{1/2}\rangle \le \sum_{j,i} (\mathrm{Tr}\Pi_j \rho_{\mathsf{Q}})(\mathrm{Tr} K_i^\dagger L_j^\dagger L_j K_i \rho_{\mathsf{Q}}) \\
& \overset{(d)}{\le} \left(\sum_{k=1}^{d_{\mathsf{K}}} p(k)\right) \sum_{j,i} \mathrm{Tr} K_i^\dagger L_j^\dagger L_j K_i \rho_{\mathsf{Q}} \overset{(e)}{=} \sum_{k=1}^{d_{\mathsf{K}}} p(k).
\end{aligned}$$

Equality (a) is by the expression for F_{op} in terms of Kraus operators (Proposition 12.15 in the previous handout) and $L_j = \Pi_j L_j$; (b) is by definition of the Hilbert-Schmidt inner product; (c) is Cauchy-Schwarz; (d) is by Proposition 13.7; (e) is by (13.9). Therefore, $s_\epsilon(p) \le d_{\mathsf{K}}$.

□

Definition 13.8 (Classical information source)**.** We model an information source as an infinite sequence of "symbols" $Z_1, Z_2, Z_3, \ldots$, where each symbol Z_i is a random variable taking values in some "alphabet" $\mathcal{A}_Z$ (this is just a set). We write

$$Z^{(n)} := (Z_1, Z_2, \ldots, Z_n)$$

for the string consisting of the first n symbols produced by the source. This is a RV taking values in $\mathcal{A}_Z^n$.

In the simplest kind of model, there are no correlations between the symbols.

Definition 13.9. Given a distribution p on a set $\mathcal{A}$, p^n denotes the distribution on $\mathcal{A}^n$ with $p^n((x_1, \ldots, x_n)) = \Pi_{i=1}^n p(x_i)$.

Definition 13.10. A **memoryless source** is one where the Z_j are independently and identically distributed (i.i.d.), meaning that $P_{Z^{(n)}} = P_Z^n$ for some P_Z. This situation is often denoted by $Z_i \overset{iid}{\sim} P_Z$. Here Z is an RV, also distributed according to P_Z and independent of the Z_i, which we use as a representative symbol produced by the source.

Suppose we want to compress the first n symbols of an information source to a bit string of the smallest possible length k given the requirement that the probability of error is no greater than ϵ. This is simply $k = \lceil \log (s_\epsilon(P_{Z^{(n)}})) \rceil$, where log is to base 2 here and throughout these notes, and where $\lceil x \rceil$ denotes the smallest integer no more than x. The optimal *rate* of compression is

$$\frac{1}{n} \lceil \log (s_\epsilon(P_{Z^{(n)}})) \rceil$$

bits per source symbol. We'll show that for a memoryless source, the large blocklength limit of the compression rate is given by the *entropy* of a source symbol. First, we will introduce a different way of measuring probability.

Definition 13.11. The **surprisal** of an event E, in bits, is $\log \frac{1}{\Pr(E)} \in [0, \infty]$.

Clearly a smaller probability means a larger surprisal. Recalling that two events A and B are independent iff

$$\Pr(A \wedge B) = \Pr(A) \Pr(B) \text{ iff } \log \frac{1}{\Pr(A \wedge B)} = \log \frac{1}{\Pr(A)} + \log \frac{1}{\Pr(B)}.$$

Definition 13.12. The **entropy** $H(X)$ of a random variable X is the expectation of its surprisal

$$H(X) := \mathbb{E} \log \frac{1}{P_X(X)} = S(P_X)$$

where, for any distribution p on a set $\mathcal{A}$, $S(p)$ is the entropy of the distribution p

$$S(p) = \sum_{x \in \text{supp}(p)} p(x) \log \frac{1}{p(x)},$$

where $\text{supp}(p)$ is the *support* of the distribution p, $\text{supp}(p) := \{x \in \mathcal{A} : p(x) > 0\}$.

Definition 13.13. Given a distribution p on a finite set $\mathcal{A}$, we define the δ-typical set for p to be

$$T_\delta(p) := \left\{ x \in \mathcal{A} : \left| \log \frac{1}{p(x)} - S(p) \right| \leq \delta \right\}. \tag{13.10}$$

Theorem 13.14 (Weak law of large numbers (WLLN))**.** Given n i.i.d. real-valued random variables Y_i with finite expectation $\mathbb{E}Y_i = \mu$ and variance $\text{var}(Y_i) = \sigma^2$, for all $\delta > 0$

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu\right| \geq \delta\right) \leq \frac{\sigma^2}{n\delta^2}.$$

When $Z_i \overset{iid}{\sim} P_Z$, $P_{Z^{(n)}} = P_Z^n$ so $H(Z^{(n)}) = S(P_{Z^{(n)}}) = nS(P_Z^n) = nH(Z)$, and

$$T_{n\delta}(P_{Z^{(n)}}) = \left\{\underline{z} \in \mathcal{A}^n : \left|\log \frac{1}{P_{Z^{(n)}}(\underline{z})} - nH(Z)\right| > n\delta\right\} \tag{13.11}$$

$$= \left\{\underline{z} \in \mathcal{A}^n : 2^{-(H(Z)+\delta)n} \leq P_{Z^{(n)}}(\underline{z}) \leq 2^{-(H(Z)-\delta)n}\right\}. \tag{13.12}$$

Proposition 13.15. If $Z_i \overset{iid}{\sim} P_Z$ then, for any $\delta > 0$:

1. $\Pr(Z^{(n)} \notin T_{n\delta}(P_{Z^{(n)}})) \leq \frac{\sigma^2}{n\delta^2}$ where $\sigma^2 = \text{var}(\log \frac{1}{P_Z(Z)})$;
2. $T_{n\delta}(P_{Z^{(n)}}) \leq 2^{(H(Z)+\delta)n}$;
3. For all $\epsilon > 0$, $\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \leq H(Z)$.

Proof. For the first claim, we have

$$\Pr(Z^{(n)} \notin T_{n\delta}(P_{Z^{(n)}})) = \Pr\left(\left|\frac{1}{n}\log \frac{1}{P_Z^n(Z^{(n)})} - H(Z)\right| > \delta\right) \tag{13.13}$$

$$= \Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - H(Z)\right| > \delta\right) \tag{13.14}$$

where Y_i is the random variable $Y_i = \log \frac{1}{P_Z(Z_i)}$. Because the Z_i are i.i.d. the Y_i are also i.i.d. and, for all i, $\mathbb{E}Y_i = H(Z)$. Therefore, the WLLN tells us that

$$\Pr(Z^{(n)} \notin T_{n\delta}(P_{Z^{(n)}})) \leq \frac{\sigma^2}{n\delta^2}.$$

For the second claim, using (13.12), we have

$$1 \geq \Pr(Z^{(n)} \in T_{n\delta}(P_{Z^{(n)}})) = \sum_{\underline{z} \in T_{n\delta}(P_{Z^{(n)}})} P_{Z^{(n)}}(\underline{z}) \geq |T_{n\delta}(P_{Z^{(n)}})| 2^{-(H(Z)+\delta)n}. \tag{13.15}$$

The first claim tells us that, for any $\delta > 0$ and $\epsilon > 0$, there is some n_0 (which depends on ϵ and δ) such that for all $n \geq n_0$, $\Pr\left(Z^{(n)} \in T_{n\delta}(P_{Z^{(n)}})\right) \geq 1 - \epsilon$. Given this and using the second claim,

$$\frac{1}{n}\log s_\epsilon(P_{Z^{(n)}}) \leq \frac{1}{n}\log |T_{n\delta}(P_{Z^{(n)}})| \leq H(Z) + \delta.$$

Therefore, for all $\delta > 0$, $\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \leq H(Z)$. □

Note that things are quite different when $\epsilon = 0$: $s_0(P_{Z^{(n)}}) = |\text{supp}(P_{Z^{(n)}})|$.

Proposition 13.16. If $Z_i \overset{iid}{\sim} P_Z$ then, for any $\epsilon \in [0, 1)$, and $\delta > 0$:

1. $s_\epsilon(P_{Z^{(n)}}) \geq (1 - \epsilon - \frac{\sigma^2}{n\delta^2}) 2^{(H(Z)-\delta)n}$ where $\sigma^2 = \text{var}(\log \frac{1}{P_Z(Z)})$.

2. $\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \geq H(Z)$.

Proof. For any RV X and sets $A, B \subseteq \mathcal{A}_X$ we have

$$\Pr(X \in A \cap B) \geq \Pr(X \in A) - \Pr(X \notin B). \tag{13.16}$$

Suppose $A \subseteq \mathcal{A}_Z^n$ is a set such that $\Pr(Z^{(n)} \in A) \geq 1-\epsilon$. From the first claim in Proposition 13.15 we know that, for any $\delta > 0$, $\Pr(Z^{(n)} \notin T_{n\delta}(P_{Z^{(n)}})) \leq \frac{\sigma^2}{n\delta^2}$. Using (13.16) and (13.12) we find that

$$1 - \epsilon - \frac{\sigma^2}{n\delta^2} \leq \Pr(Z^{(n)} \in A \cap T_{n\delta}(P_{Z^{(n)}})) \leq |A| 2^{-(H(Z)-\delta)n},$$

and the first claim follows.

Because $\epsilon < 1$, for any $\delta > 0$, there exists n_0 (which, again, will depend on ϵ and δ) such that for all $n \geq n_0$, $\frac{\sigma^2}{n\delta^2} \leq (1-\epsilon)/2$, and by the first claim

$$s_\epsilon(P_{Z^{(n)}}) \geq \frac{1-\epsilon}{2} 2^{(H(Z)-\delta)n}$$

or, equivalently,

$$\frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \geq \frac{1}{n} \log \left(\frac{1-\epsilon}{2} \right) + H(Z) - \delta.$$

It follows that, for all $0 \leq \epsilon < 1$ and $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \geq H(Z) - \delta$$

which gives us the second claim. □

If $\epsilon = 1$, then there is no limit on how much we can compress the source. Putting Propositions 13.15 and 13.16 together, and noting that $\lim_{n\to\infty} \frac{1}{n} \lceil \log s_\epsilon(P_{Z^{(n)}}) \rceil$ is equal to $\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}})$, we have

Theorem 13.17 (Source coding theorem)**.** Given a memoryless source, $Z_i \overset{iid}{\sim} P_Z$,

$$\forall \epsilon \in (0,1) : \lim_{n\to\infty} \frac{1}{n} \lceil \log s_\epsilon(P_{Z^{(n)}}) \rceil = H(Z) = S(P_Z).$$

Note that we could express this theorem without explicit reference to random variables by saying that, for any distribution p on a finite set,

$$\forall \epsilon \in (0,1) : \lim_{n\to\infty} \frac{1}{n} \lceil \log s_\epsilon(p^n) \rceil = S(p). \tag{13.17}$$

13.2 Schumacher's quantum source coding theorem

Suppose we have a composite system $\mathsf{Q}^n := \mathsf{Q}_1 \dots \mathsf{Q}_n$ where each Q_i has the same dimension d_Q. We know that the state of Q^n is the product state $\rho^{\otimes n} = \rho_{\mathsf{Q}_1} \otimes \cdots \otimes \rho_{\mathsf{Q}_n}$, but we bear in mind that Q^n may be correlated with some reference system R and we want our compression procedure to preserve these correlations.

We want to know the minimum number k such that there is an encoding operation $\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}^n}$ where K is a system of k qubits, i.e. $\dim(\mathsf{K}) = 2^k$, and decoding operation $\mathcal{D}^{\mathsf{Q}^n\leftarrow\mathsf{K}}$, such that the operation $\mathcal{D}^{\mathsf{Q}^n\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}^n}$ transmits any state $\rho_{\mathsf{Q}^n\mathsf{R}}$ of the composite system which satisfies $\mathrm{Tr}_\mathsf{R}\rho_{\mathsf{Q}^n\mathsf{R}} = \rho^{\otimes n}$ with squared fidelity $1-\epsilon$. Recalling the definitions from the previous handout, we know that k is simply

$$k = \lceil \log\left(s_\epsilon(\rho^{\otimes n})\right)\rceil.$$

We have just described the quantum analog of encoding n symbols produced by a memoryless source. Again, we ask for the large n limit of the compression *rate* $\lim_{n\to\infty}\frac{1}{n}\lceil\log\left(s_\epsilon(\rho^{\otimes n})\right)\rceil$. Now this rate is measured in "qubits per source system".

Definition 13.18. The log of a positive operator M is defined on the subspace $\mathrm{supp}(M)$ and, if M has an eigendecomposition $M = \sum_j \lambda_j|\alpha_j\rangle\langle\alpha_j|$, then $\log(M)$ is given by

$$\log(M) = \sum_{j:\lambda_j>0}\log(\lambda_j)|\alpha_j\rangle\langle\alpha_j|.$$

Definition 13.19. The **von Neumann entropy** of a density operator ρ is

$$S(\rho) = -\mathrm{Tr}\rho\log(\rho) \tag{13.18}$$

where we are treating both ρ and $\log(\rho)$ as operators on the subspace $\mathrm{supp}(\rho)$.

If ρ has an eigendecomposition $\rho = \sum_k p(k)|\alpha_k\rangle\langle\alpha_k|$ then

$$S(\rho) = -\sum_{i:p(i)>0}\sum_{j:p(j)>0}\mathrm{Tr}p(i)|i\rangle\langle i|\log(p(j))|j\rangle\langle j| = -\sum_{i\in\mathrm{supp}(p)}p(i)\log(p(i)) = S(p). \tag{13.19}$$

Furthermore, $\rho^{\otimes n}$ has an eigendecomposition

$$\rho^{\otimes n} = \sum_{k_1}\dots\sum_{k_n}p(k_1)\cdots p(k_n)|\alpha_{k_1}\rangle\langle\alpha_{k_1}|\otimes\cdots\otimes|\alpha_{k_n}\rangle\langle\alpha_{k_n}|,$$

so the theorem proved in section 13.1 (in the previous handout) tells us that, for all strictly positive integers n,

$$s_\epsilon(p^n) \le s_\epsilon(\rho^{\otimes n}) \le s_{\epsilon/2}(p^n).$$

Using this fact, the classical source coding theorem (13.17), and $S(p) = S(\rho)$ we have

Theorem 13.20. For any ρ and all $\epsilon\in(0,1)$, $\lim_{n\to\infty}\dfrac{1}{n}\lceil\log s_\epsilon(\rho^{\otimes n})\rceil = S(\rho)$.

14 Hypothesis testing

Content which only appears in this chapter is not examinable.

14.1 Relative entropy

Definition 14.1 (Support of a probability distribution)**.** Given a distribution P on some finite set $\mathcal{A}$ the **support of** P is the subset $\text{supp}(P) := \{a \in \mathcal{A} : P(a) > 0\}$.

Definition 14.2 (Relative entropy a.k.a. Kullback-Leibler divergence)**.** Given distributions P and Q on some finite set $\mathcal{A}$, the *relative entropy* of Q from P, $D(P\|Q)$, is defined by

$$D(P\|Q) := \sum_{a \in \text{supp}(P)} P(a) \log \frac{P(a)}{Q(a)},$$

where in this definition we take $s/0 = \infty$ for $s > 0$ and $\log \infty = \infty$.

Remark 14.3. $D(P\|Q)$ is finite if and only if $\text{supp}(P) \subseteq \text{supp}(Q)$.

14.2 Hypothesis testing

Suppose we have two hypotheses about an information source: The "null hypothesis" $\mathbf{H}_P$ states that $Z_i \overset{iid}{\sim} P_Z$, while the "alternative hypothesis" $\mathbf{H}_Q$ states that $Z_i \overset{iid}{\sim} Q_Z$. That is,

$$\Pr(Z^{(n)} = z^{(n)} | \mathbf{H}_P) = P_{Z^{(n)}}(z^{(n)}) = P_Z(z_1)P_Z(z_2)\cdots P_Z(z_n), \tag{14.1}$$

$$\Pr(Z^{(n)} = z^{(n)} | \mathbf{H}_Q) = Q_{Z^{(n)}}(z^{(n)}) = Q_Z(z_1)Q_Z(z_2)\cdots Q_Z(z_n). \tag{14.2}$$

We can define a *test* which looks at $Z^{(n)}$ and accepts either $\mathbf{H}_P$ or $\mathbf{H}_Q$ by giving an *acceptance region* $\Pi_n \subseteq \mathcal{A}_Z^n$ for $\mathbf{H}_P$: When $Z^{(n)}$ belongs to Π_n the test accepts $\mathbf{H}_P$; otherwise it rejects $\mathbf{H}_P$ (and accepts $\mathbf{H}_Q$). The probability that the test rejects $\mathbf{H}_P$, given that $\mathbf{H}_P$ is true (a "type I error") is

$$\alpha_n(\Pi_n) := \Pr(Z^{(n)} \notin \Pi_n | \mathbf{H}_P). \tag{14.3}$$

The probability that the test accepts $\mathbf{H}_P$, given that $\mathbf{H}_Q$ is true (a "type II error") is

$$\beta_n(\Pi_n) := \Pr(Z^{(n)} \in \Pi_n | \mathbf{H}_Q).$$

There will be a trade-off between these two error probabilities. Let's consider the problem of minimising $\beta_n(\Pi_n)$ subject to the constraint $\alpha_n(\Pi_n) \leq \epsilon$. We define

$$\beta_{n,\epsilon} := \min\{\beta_n(\Pi_n) : \Pi_n \subseteq \mathcal{A}_Z^n, \alpha_n(\Pi_n) \leq \epsilon\}.$$

Remark 14.4. The "Neyman-Pearson lemma" says that the extremal points of the set of achievable pairs (α_n, β_n) correspond to acceptance regions of the form

$$R_{n,t} = \left\{ z^{(n)} \in \mathcal{A}_Z^n : \log \frac{P_{Z^{(n)}}(z^{(n)})}{Q_{Z^{(n)}}(z^{(n)})} \geq t \right\}.$$

This form of test is called a "likelihood ratio test" and the proof of the Neyman-Pearson lemma doesn't require any assumptions on the forms of $P_{Z^{(n)}}$ and $Q_{Z^{(n)}}$ (such as i.i.d.).

14.2.1 Asymptotics of hypothesis testing: Stein's lemma

We will work out the asymptotic behaviour of $\beta_{n,\epsilon}$ as n grows large. As was the case in our analysis of source coding, we will not directly analyse the optimal solution. The results will prove useful when we look at noisy channel coding.

Proposition 14.5. If $\text{supp}(P_Z) \not\subseteq \text{supp}(Q_Z)$ then for all $\epsilon > 0$ and all sufficiently large n, $\beta_{n,\epsilon} = 0$. Therefore, for all $\epsilon > 0$

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_{n,\epsilon}} = \infty = D(P_Z \| Q_Z).$$

Proof. $\text{supp}(P_Z) \not\subseteq \text{supp}(Q_Z)$ iff there is some $a \in \mathcal{A}_Z$ such that $P_Z(a) > 0$ and $Q_Z(a) = 0$. If the source produces an a we know with certainty that $\mathbf{H}_Q$ is false. Consider the acceptance region $\Pi_n = \left\{ z^{(n)} \in \mathcal{A}_Z^n : z_i = a \text{ for some } i \in \{1, \ldots, n\} \right\}$. Then $\beta_n(\Pi_n) = 0$ and $\alpha_n(\Pi_n) = (1 - P_Z(a))^n$. □

Proposition 14.6. If $\text{supp}(P_Z) \subseteq \text{supp}(Q_Z)$ then for all $\epsilon \in (0, 1)$, for all $\delta > 0$, and all sufficiently large n:

$$\frac{(1-\epsilon)}{2} 2^{-(D(P_Z\|Q_Z)+\delta)n} \leq \beta_{n,\epsilon} \leq 2^{-(D(P_Z\|Q_Z)-\delta)n}$$

and therefore

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_{n,\epsilon}} = D(P_Z \| Q_Z).$$

Proof. Conditioned on $\mathbf{H}_P$, $Z_i \overset{iid}{\sim} P_Z$ and (using (14.1) and (14.2))

$$\frac{1}{n} \log \frac{P_{Z^{(n)}}(Z^{(n)})}{Q_{Z^{(n)}}(Z^{(n)})} = \frac{1}{n} \sum_{i=1}^{n} \log \frac{P_Z(Z_i)}{Q_Z(Z_i)}. \tag{14.4}$$

The fact that $\text{supp}(P_Z) \subseteq \text{supp}(Q_Z)$ means that the i.i.d. RVs $\log \frac{P_Z(Z_i)}{Q_Z(Z_i)}$ each have finite expectation

$$\mathbb{E}\left[\log \frac{P_Z(Z_i)}{Q_Z(Z_i)} \middle| \mathbf{H}_P \right] = D(P_Z \| Q_Z)$$

and finite variance. The WLLN tells us that (14.4) will converge in probability to $D(P_Z\|Q_Z)$. So, let's define an acceptance region $\Pi_{n,\delta}$ such that $\mathbf{H}_P$ is accepted when $\frac{1}{n} \log \frac{P_{Z^{(n)}}(Z^{(n)})}{Q_{Z^{(n)}}(Z^{(n)})}$ is within δ of $D(P_Z\|Q_Z)$:

$$\Pi_{n,\delta} := \left\{ z^{(n)} \in \mathcal{A}_Z^n : \left| \frac{1}{n} \log \frac{P_{Z^{(n)}}(z^{(n)})}{Q_{Z^{(n)}}(z^{(n)})} - D(P_Z\|Q_Z) \right| \leq \delta \right\} \tag{14.5}$$

$$= \left\{ z^{(n)} \in \mathcal{A}_Z^n : 2^{(D(P_Z\|Q_Z)-\delta)n} \leq \frac{P_{Z^{(n)}}(z^{(n)})}{Q_{Z^{(n)}}(z^{(n)})} \leq 2^{(D(P_Z\|Q_Z)+\delta)n} \right\} \tag{14.6}$$

1. The WLLN says that, for any $\delta > 0, \lim_{n\to\infty} \alpha_n(\Pi_{n,\delta}) = 0$. This establishes that, for all $\epsilon > 0$, $\delta > 0$ and for all sufficiently large n, $\beta_{n,\epsilon} \leq \beta_n(\Pi_{n,\delta})$.

2. For all $\delta > 0$,

$$\beta_n(\Pi_{n,\delta}) = \sum_{z^{(n)} \in \Pi_{n,\delta}} Q_{Z^{(n)}}(z^{(n)}) \leq \sum_{z^{(n)} \in \Pi_{n,\delta}} P_{Z^{(n)}}(z^{(n)}) 2^{-(D(P_Z \| Q_Z) - \delta)n} \leq 2^{-(D(P_Z \| Q_Z) - \delta)n}.$$

   Together with point (1), this establishes the 'achievability' part of the result:

3. To prove the 'converse' part of the result: For any ϵ, and any acceptance region $\Pi_n \subseteq \mathcal{A}_Z^n$ satisfying $\alpha_n(\Pi_n) \leq \epsilon$, and $\delta > 0$:

$$\begin{aligned}
\beta_n(\Pi_n) \geq & \Pr(Z^{(n)} \in \Pi_n \cap \Pi_{n,\delta} | \mathbf{H}_Q) \\
= & \sum_{z^{(n)} \in \Pi_n \cap \Pi_{n,\delta}} Q_{Z^{(n)}}(z^{(n)}) \\
\geq & \sum_{z^{(n)} \in \Pi_n \cap \Pi_{n,\delta}} P_{Z^{(n)}}(z^{(n)}) 2^{-(D(P_Z \| Q_Z) + \delta)n} \\
= & \Pr(Z^{(n)} \in \Pi_n \cap \Pi_{n,\delta} | \mathbf{H}_P) 2^{-(D(P_Z \| Q_Z) + \delta)n} \\
\geq & \left(1 - \Pr(Z^{(n)} \notin \Pi_n | \mathbf{H}_P) - \Pr(Z^{(n)} \notin \Pi_{n,\delta} | \mathbf{H}_P)\right) 2^{-(D(P_Z \| Q_Z) + \delta)n} \\
= & (1 - \alpha_n(\Pi_n) - \alpha_n(\Pi_{n,\delta})) 2^{-(D(P_Z \| Q_Z) + \delta)n} \\
\geq & (1 - \epsilon - \alpha_n(\Pi_{n,\delta})) 2^{-(D(P_Z \| Q_Z) + \delta)n}
\end{aligned}$$

   where the 2nd inequality comes from (14.6). By point (1), for any $\epsilon < 1$, for all sufficiently large n, $\alpha_n(\Pi_{n,\delta}) \leq (1 - \epsilon)/2$.

□

Putting the last two propositions together we have

Theorem 14.7 (Stein's lemma)**.**

$$\forall \epsilon \in (0, 1) : \lim_{n \to \infty} \frac{1}{n} \log \frac{1}{\beta_{n,\epsilon}} = D(P_Z \| Q_Z).$$

15 Entropies and hypothesis testing

15.1 Entropies

15.1.1 Notation

Given more than one random variable, X and Y for example,

$$H(X,Y) := S(P_{XY})$$

denotes the "joint entropy" of X and Y. Note that this is simply the entropy of the random variable (X,Y).

If we have states ρ_{ABC} and σ_{ABD} then, unless otherwise specified, $\rho_{\mathsf{A}} := \mathrm{Tr}_{\mathsf{BC}}\rho_{\mathsf{ABC}}$, $\rho_{\mathsf{AC}} := \mathrm{Tr}_{\mathsf{B}}\rho_{\mathsf{ABC}}$, $\sigma_{\mathsf{BD}} := \mathrm{Tr}_{\mathsf{A}}\sigma_{\mathsf{ABD}}$, etc. denote the states of the various subsystems. In analogy with our notation for entropies of random variables we write the von Neumann entropies for the states of these systems $H(\mathsf{A})_\rho := S(\rho_{\mathsf{A}})$, $H(\mathsf{D})_\sigma := S(\sigma_{\mathsf{D}})$, $H(\mathsf{ABC})_\rho := S(\rho_{\mathsf{ABC}})$, $H(\mathsf{AB})_\rho := S(\rho_{\mathsf{AB}})$, $H(\mathsf{AB})_\sigma := S(\sigma_{\mathsf{AB}})$, and so on.

15.1.2 Classical information stored in quantum systems

Suppose we have a number of random variables, e.g. X, Y and Z, whose values are stored in quantum systems $\tilde{\mathsf{X}}$, $\tilde{\mathsf{Y}}$ and $\tilde{\mathsf{Z}}$ in the standard way, so that the state of $\tilde{\mathsf{X}}\tilde{\mathsf{Y}}\tilde{\mathsf{Z}}$ is

$$\rho_{\tilde{\mathsf{X}}\tilde{\mathsf{Y}}\tilde{\mathsf{Z}}} = \sum_{x,y,z} P_{XYZ}(x,y,z)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |y\rangle\langle y|_{\tilde{\mathsf{Y}}} \otimes |z\rangle\langle z|_{\tilde{\mathsf{Z}}}. \tag{15.1}$$

The quantum entropies of these systems correspond to the classical entropies of the RVs in the obvious way e.g. $H(\tilde{\mathsf{X}}\tilde{\mathsf{Y}}\tilde{\mathsf{Z}})_\rho = H(X,Y,Z)$, $H(\tilde{\mathsf{Y}}\tilde{\mathsf{Z}})_\rho = H(Y,Z)$, $H(\tilde{\mathsf{X}})_\rho = H(X)$ etc.

15.1.3 Basic bounds on entropy

Given any distribution p on a finite set $\mathcal{A}$,

$$0 \leq S(p) \leq \log|\mathrm{supp}(p)| \leq \log|\mathcal{A}|, \tag{15.2}$$

where the first equality holds iff $p(x) = 1$ for some $x \in \mathcal{A}$. This result can be established by using the strict concavity of log and Jensen's inequality. Doing so is an exercise on example sheet 3.

Given any state ρ on a finite dimensional Hilbert space $\mathcal{H}$,

$$0 \leq S(\rho) \leq \log\dim(\mathrm{supp}(\rho)) \leq \log\dim(\mathcal{H}), \tag{15.3}$$

where the first equality holds iff ρ is pure. These inequalities follow easily from those in (15.2) by using the fact that if $\rho = \sum_k p(k)|\alpha_k\rangle\langle\alpha_k|$ is an eigendecomposition for ρ then $S(\rho) = S(p)$.

15.2 Conditional entropy and the chain rule

Definition 15.1 (Classical conditional entropies)**.** Given random variables X and Y and Z:

1. $H(X|Y=y) := S(P_{X|Y=y})$;
2. $H(X|Y) := \sum_{y\in\mathcal{A}_Y} H(X|Y=y)\Pr(Y=y)$;
3. $H(X|Y,Z=z) = \sum_{y\in\mathcal{A}_Y} H(X|Y=y,Z=z)\Pr(Y=y|Z=z)$.

It follows from the results in section 15.1.3 that $H(X|Y)\geq 0$ with equality iff $X=f(Y)$ for some function $f:\text{supp}(P_Y)\to\mathcal{A}_X$. Similarly, $H(X|Y,Z=z)\geq 0$ with equality iff, given $Z=z$, $X=f(Y)$ for some function $f:\text{supp}(P_{Y|Z=z})\to\mathcal{A}_X$.

The product rule of probability says that $P_{XY}(x,y)=P_{X|Y=y}(x)P_Y(y)$, so

$$\log\frac{1}{P_{XY}(x,y)} = \log\frac{1}{P_{X|Y=y}(x)} + \log\frac{1}{P_Y(y)}. \tag{15.4}$$

Multiplying by $P_{XY}(x,y)$ and summing over x and y

$$\begin{aligned}\sum_{x,y} P_{XY}(x,y)\log\frac{1}{P_{XY}(x,y)} &= \sum_{x,y} P_{XY}(x,y)\log\frac{1}{P_{X|Y=y}(x)} + \sum_{x,y} P_{XY}(x,y)\log\frac{1}{P_Y(y)}\\ &= \sum_y P_Y(y)\sum_x P_{X|Y}(x|y)\log\frac{1}{P_{X|Y=y}(x)} + \sum_y P_Y(y)\log\frac{1}{P_Y(y)}\end{aligned}$$

which is equivalent to

Proposition 15.2 (The chain rule)**.** $H(X,Y)=H(X|Y)+H(Y)$.

By treating multiple random variables as a single tuple of random variables, we can extend the chain rule to more variables. For example

$$\begin{aligned}H(X,Y|Z) = H(X,Y,Z) - H(Z) &= (H(X,Y,Z)-H(Y,Z)) + H(Y,Z) - H(Z)\\ &= H(X|Y,Z) + H(Y|Z).\end{aligned}$$

The *quantum conditional entropy* is *defined* to obey the chain rule.

Definition 15.3. For systems A and B:

1. The **quantum conditional entropy** of A given B is $H(\mathsf{A}|\mathsf{B}) := H(\mathsf{AB}) - H(\mathsf{B})$.
2. The **coherent information** of A given B is $I(\mathsf{A}\rangle\mathsf{B}) := -H(\mathsf{A}|\mathsf{B})$.

Any identity between classical entropies derived using only the chain rule, can be derived in the same way for quantum entropies. For example

$$H(\mathsf{AB}|\mathsf{C}) = H(\mathsf{A}|\mathsf{BC}) + H(\mathsf{B}|\mathsf{C}).$$

A striking difference between classical and quantum conditional entropies is that while any classical conditional entropy is positive, quantum conditional entropy can be negative. For example, if $d_\mathsf{A}=d_\mathsf{B}=d$ and $\rho_\mathsf{AB}=\phi^+_\mathsf{AB}$ then $H(\mathsf{A}|\mathsf{B})_\rho = H(\mathsf{AB})_\rho - H(\mathsf{B})_\rho = S(\phi^+_\mathsf{AB}) - S(\mathbb{1}_\mathsf{B}/d) = 0 - \log d$.

15.3 Mutual information and conditional mutual information

Given two random variables X and Y, the entropy gives us a way to quantify how much Y tells us about X: $H(X)$ is a way of quantifying our uncertainty about X. Prior to learning the value of Y, our expectation for the entropy we will assign to X if we *do* learn the value of Y, is $H(X|Y)$. Therefore, $I(X:Y) := H(X) - H(X|Y)$ is the expectation of the *reduction in entropy* of X which will occur if we learn the value of Y. This is a measure of how much learning Y would *inform* us about X. By the chain rule,

$$I(X:Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y) = H(Y) - H(Y|X),$$

so $I(X:Y)$ is *symmetric* in the two random variables, and we call it the **mutual information** between X and Y.

Definition 15.4. For random variables X, Y and Z:

1. The mutual information between X and Y is
$$I(X:Y) := H(X) - H(X|Y). \tag{15.5}$$
2. The conditional mutual information between X and Y given $Z = z$ is
$$I(X:Y|Z=z) := H(X|Z=z) - H(X|Y,Z=z). \tag{15.6}$$
3. The conditional mutual information between X and Y given Z is
$$\begin{aligned} I(X:Y|Z) :=& \sum_z \Pr(Z=z) I(X:Y|Z=z) = H(X|Z) - H(X|Y,Z) \\ =& H(X|Z) + H(Y|Z) - H(X,Y|Z) \\ =& H(Y|Z) - H(Y|X,Z). \end{aligned}$$

Definition 15.5. For systems A, B and C:

1. The **quantum mutual information** between A and B is
$$I(\mathsf{A}:\mathsf{B}) := H(\mathsf{A}) - H(\mathsf{A}|\mathsf{B}) = H(\mathsf{A}) + H(\mathsf{B}) - H(\mathsf{AB}) = H(\mathsf{B}) - H(\mathsf{B}|\mathsf{A}). \tag{15.7}$$
2. **quantum conditional mutual information** between A and B given C is
$$I(\mathsf{A}:\mathsf{B}|\mathsf{C}) := H(\mathsf{A}|\mathsf{C}) - H(\mathsf{A}|\mathsf{BC}) = H(\mathsf{AC}) + H(\mathsf{BC}) - H(\mathsf{C}) - H(\mathsf{ABC}). \tag{15.8}$$

Ex. sheet 3 asks you to prove the **chain rule for conditional mutual information**:

Proposition 15.6. $I(X:Y,Z|W) = I(X:Z|W) + I(X:Y|Z,W)$.

Since this can be proved using the chain rule alone, it also holds for quantum entropies:

Proposition 15.7. $I(\mathsf{A}:\mathsf{BC}|\mathsf{D}) = I(\mathsf{A}:\mathsf{C}|\mathsf{D}) + I(\mathsf{A}:\mathsf{B}|\mathsf{CD})$.

15.4 Hypothesis testing and relative entropy

We now return briefly to a kind of simple state discrimination problem known as "quantum hypothesis testing". Given a quantum system Q, let hypothesis H_0 be that the state of Q is ρ, and hypothesis H_1 be that the state of Q is σ. Suppose we measure a POVM $E : \{0,1\} \to \mathcal{H}_\mathsf{Q}$ obtaining a result $\hat{X}$ which is supposed to identify which hypothesis is true. Since this is a binary POVM, if we set $E(0) = T$ then this determines $E(1) = \mathbb{1} - T$.

$$\text{The "type-I" error probability is } \Pr(\hat{X} = 1|H_0) = \alpha(T,\rho) := 1 - \mathrm{Tr}T\rho, \text{ and} \tag{15.9}$$

$$\text{the "type-II" error probability is } \Pr(\hat{X} = 0|H_1) = \beta(T,\sigma) := \mathrm{Tr}T\sigma. \tag{15.10}$$

In general, there is a trade-off between these two conditional probabilities. We use the following notation for the minimum value of β that can be attained subject to the requirement that $\alpha \leq \epsilon$.

Definition 15.8. $\beta_\epsilon(\rho\|\sigma) := \min\{\beta(T,\sigma) : \alpha(T,\rho) \leq \epsilon, 0 \leq T \leq \mathbb{1}\}$.

Proposition 15.9 (Data processing inequality for β_ϵ)**.** For all $\epsilon \in [0,1]$, states ρ_A and σ_A, and operations $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$, $\beta_\epsilon(\mathcal{N}\rho_\mathsf{A}\|\mathcal{N}\sigma_\mathsf{A}) \geq \beta_\epsilon(\rho_\mathsf{A}\|\sigma_\mathsf{A})$.

Proof. Because the linear map $\mathcal{N}$ is completely positive and trace preserving, its adjoint $\mathcal{N}^\dagger$ is completely positive and unital (identity preserving). So if $0 \leq T \leq \mathbb{1}$ then $0 \leq \mathcal{N}^\dagger T \leq \mathcal{N}^\dagger\mathbb{1} = \mathbb{1}$ and therefore

$$\beta_\epsilon(\mathcal{N}\rho\|\mathcal{N}\sigma) = \min\{\beta(T,\mathcal{N}\sigma) : \alpha(T,\mathcal{N}\rho) \leq \epsilon, 0 \leq T \leq \mathbb{1}\} \tag{15.11}$$

$$= \min\{\beta(\mathcal{N}^\dagger T,\sigma) : \alpha(\mathcal{N}^\dagger T,\rho) \leq \epsilon, 0 \leq T \leq \mathbb{1}\} \geq \beta_\epsilon(\rho\|\sigma). \tag{15.12}$$

□

Intuitively, if we are given n systems in the state ρ or n systems in the state σ then larger values of n should make it easier to distinguish between the two situations. The next theorem tells us that $\beta_\epsilon(\rho^{\otimes n}\|\sigma^{\otimes n})$ exhibits exponential decay as n increases, and gives us the rate of the decay.

Theorem 15.10 (Quantum Stein's lemma)**.** For all states ρ, σ of a given system

$$\forall\epsilon \in (0,1), \lim_{n\to\infty} \frac{1}{n} \log \beta_\epsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) = -D(\rho\|\sigma), \tag{15.13}$$

where $D(\rho\|\sigma)$ is the *quantum relative entropy* between ρ and σ...

Definition 15.11 (Quantum relative entropy)**.**

$$D(\rho\|\sigma) := \begin{cases} -S(\rho) - \mathrm{Tr}\rho\log\sigma & \text{if } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma), \\ +\infty & \text{otherwise.} \end{cases} \tag{15.14}$$

In $\mathrm{Tr}\rho\log\sigma$ we regard both ρ and $\log\sigma$ as operators on $\mathrm{supp}(\sigma)$.

Note that the relative entropy is not symmetric in its arguments and does not obey the triangle inequality.

Proposition 15.12. Density operators ρ and σ on $\mathcal{H}$ commute if and only if there are eigendecompositions $\rho = \sum_{i=1}^{\dim(\mathcal{H})} p(i)|\alpha_i\rangle\langle\alpha_i|$ and $\sigma = \sum_{i=1}^{\dim(\mathcal{H})} q(i)|\alpha_i\rangle\langle\alpha_i|$. In this case, $D(\rho\|\sigma) = D(p\|q)$, where $D(p\|q)$ is the *classical relative entropy* between the distributions p and q...

Definition 15.13 (Classical relative entropy)**.**

$$D(p\|q) := \begin{cases} -S(p) - \sum_{x\in\mathrm{supp}(q)} p(x)\log q(x) & \text{if } \mathrm{supp}(p) \subseteq \mathrm{supp}(q), \\ +\infty & \text{otherwise.} \end{cases} \tag{15.15}$$

Theorem 15.14 (Gibbs' inequality)**.** For any two probability distributions p and q on a finite set $\mathcal{A}$, $D(p\|q) \geq 0$ with equality iff $p = q$.

Proof. That $D(p\|q) \geq 0$ already follows from the operational meaning given to $D(p\|q)$ by Stein's lemma. Here is a direct proof which gives us the equality condition. If $\mathrm{supp}(p) \not\subseteq \mathrm{supp}(q)$ then $D(p\|q) = \infty$ and the result holds, so let's now assume that $\mathrm{supp}(p) \subseteq \mathrm{supp}(q)$. $D(p\|q) = \sum_{x\in\mathrm{supp}(p)} p(x)\log\frac{p(x)}{q(x)} = -\frac{1}{\ln(2)}\sum_{x\in\mathrm{supp}(p)} p(x)\ln\frac{q(x)}{p(x)}$. The line $y = x-1$ is tangent to the graph $y = \ln x$ at $x = 1$ so from the strict concavity of ln it follows that $\ln x \leq x - 1$ with equality iff $x = 1$. Therefore,

$$-D(p\|q) \leq \frac{1}{\ln(2)} \sum_{x\in\mathrm{supp}(p)} p(x)\left(\frac{q(x)}{p(x)} - 1\right) \leq 0.$$

Since the $p(x)$ in the sum are strictly positive, the first inequality is an equality if and only if $q(x)/p(x) = 1$ for all $x \in \mathrm{supp}(p)$, which is true iff $q(x) = p(x)$ for all x, in which case the second inequality is also satisfied. □

Theorem 15.15 (Data processing inequality for D)**.** For all states ρ_{A} and σ_{A}, and operations $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$, $D(\mathcal{N}\rho_{\mathsf{A}}\|\mathcal{N}\sigma_{\mathsf{A}}) \leq D(\rho_{\mathsf{A}}\|\sigma_{\mathsf{A}})$.

Proof.

$$\begin{aligned} D(\mathcal{N}\rho\|\mathcal{N}\sigma) &= -\lim_{n\to\infty}\frac{1}{n}\log\beta_\epsilon((\mathcal{N}\rho)^{\otimes n}\|(\mathcal{N}\sigma)^{\otimes n}) && (15.16)\\ &= -\lim_{n\to\infty}\frac{1}{n}\log\beta_\epsilon(\mathcal{N}^{\otimes n}\rho^{\otimes n}\|\mathcal{N}^{\otimes n}\sigma^{\otimes n}) && (15.17)\\ &\leq -\lim_{n\to\infty}\frac{1}{n}\log\beta_\epsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma). && (15.18) \end{aligned}$$

The first and last equalities are by Quantum Stein's lemma. The inequality is by the data processing inequality for β_ϵ, the fact that log is an increasing function, and an elementary fact about limits. □

Theorem 15.16 (Klein's inequality)**.** $D(\rho\|\sigma) \geq 0$ with equality iff $\rho = \sigma$.

Proof. Suppose that $X = 0$ with probability $1/2$ and $X = 1$ with probability $1/2$, and suppose that the state of Q is ρ when $X = 0$ and the state is σ when $X = 1$. Suppose we measure a POVM E to obtain an estimate $\hat{X}$ of X and then prepare Q in the computational basis state $|\hat{X}\rangle\langle\hat{X}|$. This corresponds to applying the operation

$$\mathcal{N} : \tau \mapsto |0\rangle\langle 0|\mathrm{Tr}E(0)\tau + |1\rangle\langle 1|\mathrm{Tr}E(1)\tau. \tag{15.19}$$

We choose the POVM which maximises $\Pr(\hat{X} = X)$.

Letting $p(i) = \Pr(\hat{X} = i|X = 0)$ and $q(i) = \Pr(\hat{X} = i|X = 1)$, $\mathcal{N}\rho = p(0)|0\rangle\langle 0| + p(1)|1\rangle\langle 1|$ and $\mathcal{N}\sigma = q(0)|0\rangle\langle 0| + q(1)|1\rangle\langle 1|$. By the data processing inequality for quantum relative entropy, Proposition 15.12, and Gibbs' inequality

$$D(\rho\|\sigma) \geq D(\mathcal{N}\rho\|\mathcal{N}\sigma) = D(p\|q) \geq 0 \tag{15.20}$$

so we require $p = q$ for $D(\rho\|\sigma) = 0$ to hold. The Holevo-Helstrom theorem tells us that

$$\Pr(\hat{X} = X) = \frac{1}{2}\left(p(0) + q(1)\right) = \frac{1}{2}\left(1 + q(1) - p(1)\right) = \frac{1}{2}\left(1 + \frac{1}{2}\|\sigma - \rho\|_1\right) \tag{15.21}$$

so $q(1) - p(1) = \frac{1}{2}\|\sigma - \rho\|_1$, and $q = p$ iff $\|\sigma - \rho\|_1 = 0$ iff $\sigma = \rho$. □

15.5 Entropic inequalities

Example sheet 3 asks you to show

Proposition 15.17. $I(X : Y) = D(P_{XY}\|P_X P_Y)$.

Therefore, Gibb's inequality tells us that $I(X : Y) \geq 0$ with equality iff $P_{XY} = P_X P_Y$. From this and the definitions of the classical quantities in section 15.3 we obtain

Proposition 15.18. For any random variables X, Y and Z (taking values in finite sets):

1. $I(X : Y) \geq 0$ with equality iff X and Y are independent, which means

$$\forall x, y : P_{XY}(x, y) = P_X(x)P_Y(y).$$

2. $I(X : Y|Z = z) \geq 0$ with equality iff X and Y are conditionally independent given $Z = z$, which means

$$\forall x, y : P_{XY|Z=z}(x, y) = P_{X|Z=z}(x)P_{Y|Z=z}(y).$$

3. $I(X : Y|Z) \geq 0$ with equality iff X and Y are conditionally independent given Z, which means

$$\forall z \in \mathrm{supp}(P_Z), x, y : P_{XY|Z}(x, y|z) = P_{X|Z}(x|z)P_{Y|Z}(y|z).$$

Example sheet 3 also asks you to show

Proposition 15.19. $I(\mathsf{A} : \mathsf{B})_\rho := D(\rho_{\mathsf{AB}}\|\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}})$.

So Klein's inequality gives us

Proposition 15.20. For any state ρ_{AB}, $I(\mathsf{A} : \mathsf{B})_\rho \geq 0$ with equality iff $\rho_{\mathsf{AB}} = \rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}$.

Quantum *conditional* mutual information is also positive, a fact known as the **strong subadditivity** of von Neumann entropy. Proving this from scratch is much more involved than the classical case, but it is a fairly easy consequence of the data processing inequality for quantum relative entropy.

Theorem 15.21. For any state ρ_{ABC}, $I(\mathsf{A}:\mathsf{B}|\mathsf{C}) \geq 0$.

Proof. By the chain rule and Proposition 15.19,

$$\begin{aligned} I(\mathsf{A}:\mathsf{B}|\mathsf{C}) =& I(\mathsf{A}:\mathsf{BC}) - I(\mathsf{A}:\mathsf{C}) \\ =& D(\rho_{\mathsf{ABC}}\|\rho_{\mathsf{A}} \otimes \rho_{\mathsf{BC}}) - D(\rho_{\mathsf{AC}}\|\rho_{\mathsf{A}} \otimes \rho_{\mathsf{C}}) \\ =& D(\rho_{\mathsf{ABC}}\|\rho_{\mathsf{A}} \otimes \rho_{\mathsf{BC}}) - D(\mathrm{Tr}_{\mathsf{B}}\rho_{\mathsf{ABC}}\|\mathrm{Tr}_{\mathsf{B}}\rho_{\mathsf{A}} \otimes \rho_{\mathsf{BC}}) \\ \geq& D(\rho_{\mathsf{ABC}}\|\rho_{\mathsf{A}} \otimes \rho_{\mathsf{BC}}) - D(\rho_{\mathsf{ABC}}\|\rho_{\mathsf{A}} \otimes \rho_{\mathsf{BC}}) = 0, \end{aligned}$$

where the inequality is the data processing inequality for quantum relative entropy. □

Proposition 15.22. If ρ_{AB} is a pure state, then $H(\mathsf{A})_\rho = H(\mathsf{B})_\rho$.

Proof. From the existence of a Schmidt decomposition for ρ_{AB} we know that ρ_{A} and ρ_{B} have the same non-zero eigenvalues with the same multiplicities, and therefore have the same von Neumann entropy. □

Theorem 15.23. Data processing inequality for the coherent information: If $\sigma_{\mathsf{AC}} = \mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}\rho_{\mathsf{AB}}$ for some operation $\mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}$ then $I(\mathsf{A}\rangle\mathsf{B})_\rho \geq I(\mathsf{A}\rangle\mathsf{C})_\sigma$.

Proof. Let ρ_{RAB} be a purification of ρ_{AB}, and let $\mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}X_{\mathsf{B}} = \mathrm{Tr}_{\mathsf{E}}VX_{\mathsf{B}}V^\dagger$ be a Stinespring representation for $\mathcal{N}^{\mathsf{C}\leftarrow\mathsf{B}}$ (so V is an isometry in $\mathcal{L}(\mathcal{H}_{\mathsf{B}}, \mathcal{H}_{\mathsf{C}} \otimes \mathcal{H}_{\mathsf{E}})$). Defining

$$\psi_{\mathsf{RACE}} = \mathbb{1}_{\mathsf{RA}} \otimes V\rho_{\mathsf{RAB}}\mathbb{1}_{\mathsf{RA}} \otimes V^\dagger,$$

we have $\mathrm{Tr}_{\mathsf{RE}}\psi_{\mathsf{RACE}} = \sigma_{\mathsf{AC}}$, so let's write $\sigma_{\mathsf{RACE}} := \psi_{\mathsf{RACE}}$. Note that σ_{RACE} is a pure state. Now, the coherent information before the operation (the 'data processing') is

$$I(\mathsf{A}\rangle\mathsf{B})_\rho = H(\mathsf{B})_\rho - H(\mathsf{AB})_\rho \overset{(a)}{=} H(\mathsf{RA})_\rho - H(\mathsf{R})_\rho \overset{(b)}{=} H(\mathsf{RA})_\sigma - H(\mathsf{R})_\sigma = H(\mathsf{A}|\mathsf{R})_\sigma, \quad (15.22)$$

where (a) is by two applications of Proposition 15.22 to the pure state ρ_{RAB}, (b) is because $\rho_{\mathsf{RA}} = \sigma_{\mathsf{RA}}$, and the other equalities are true by definition. After the operation,

$$\begin{aligned} I(\mathsf{A}\rangle\mathsf{C})_\sigma =& H(\mathsf{C})_\sigma - H(\mathsf{AC})_\sigma \overset{(c)}{=} H(\mathsf{RAE})_\sigma - H(\mathsf{RE})_\sigma \\ =& H(\mathsf{AE}|\mathsf{R})_\sigma + H(\mathsf{R})_\sigma - (H(\mathsf{E}|\mathsf{R})_\sigma + H(\mathsf{R})_\sigma) = H(\mathsf{AE}|\mathsf{R})_\sigma - H(\mathsf{E}|\mathsf{R})_\sigma, \end{aligned} \quad (15.23)$$

where (c) is by two applications of Proposition 15.22 to the pure state ρ_{RACE} and the other equalities are definitions. Subtracting equation (15.23) from (15.22) yields

$$I(\mathsf{A}\rangle\mathsf{B})_\rho - I(\mathsf{A}\rangle\mathsf{C})_\sigma = H(\mathsf{A}|\mathsf{R})_\sigma + H(\mathsf{E}|\mathsf{R})_\sigma - H(\mathsf{AE}|\mathsf{R})_\sigma = I(\mathsf{A}:\mathsf{E}|\mathsf{R}) \geq 0$$

by definition and positivity of QCMI. □

Theorem 15.24. Data processing inequality for the quantum mutual information: If $\sigma_{\mathsf{A'B'}} = \mathcal{N}^{\mathsf{A'}\leftarrow\mathsf{A}} \otimes \mathcal{M}^{\mathsf{B'}\leftarrow\mathsf{B}}\rho_{\mathsf{AB}}$ where $\mathcal{N}^{\mathsf{A'}\leftarrow\mathsf{A}}$ and $\mathcal{M}^{\mathsf{B'}\leftarrow\mathsf{B}}$ are operations then

$$I(\mathsf{A}:\mathsf{B})_\rho \geq I(\mathsf{A'}:\mathsf{B'})_\sigma.$$

Proof. With $\rho'_{\mathsf{AB'}} := \mathbf{id}^{\mathsf{A}\leftarrow\mathsf{A}} \otimes \mathcal{M}^{\mathsf{B'}\leftarrow\mathsf{B}}\rho_{\mathsf{AB}}$, $\sigma_{\mathsf{A'B'}} = \mathcal{N}^{\mathsf{A'}\leftarrow\mathsf{A}} \otimes \mathbf{id}^{\mathsf{B'}\leftarrow\mathsf{B'}}\rho'_{\mathsf{AB'}}$, Using $\rho'_{\mathsf{A}} = \rho_{\mathsf{A}}$ and $\rho'_{\mathsf{B'}} = \sigma_{\mathsf{B'}}$, the relationships between mutual information and coherent information, and applying the DPI for coherent information twice

$$\begin{aligned} & I(\mathsf{A}:\mathsf{B})_\rho = H(\mathsf{A})_\rho + I(\mathsf{A}\rangle\mathsf{B})_\rho \geq H(\mathsf{A})_{\rho'} + I(\mathsf{A}\rangle\mathsf{B'})_{\rho'} \\ =& I(\mathsf{A}:\mathsf{B'})_{\rho'} = H(\mathsf{B'})_{\rho'} + I(\mathsf{B'}\rangle\mathsf{A})_{\rho'} \geq H(\mathsf{B'})_\sigma + I(\mathsf{B'}\rangle\mathsf{A})_\sigma = I(\mathsf{A'}:\mathsf{B'})_\sigma. \end{aligned}$$

□

15.6 Fano's inequality

If $\mathcal{A}_X = \{0, 1\}$ then $H(X) = h(P_X(1))$ where h is the **binary entropy function**:

Definition 15.25. $h(\lambda) := (1-\lambda)\log\frac{1}{1-\lambda} + \lambda\log\frac{1}{\lambda}$.

Proposition 15.26 (Fano's inequality)**.** Given two random variables M and $\hat{M}$ which both take values in a finite set $\mathcal{A}$, we have

$$H(M|\hat{M}) \leq \Pr(\hat{M} \neq M)\log(|\mathcal{A}| - 1) + h(\Pr(\hat{M} \neq M)).$$

Proof. Let E be a binary RV which is 1 if $\hat{M} \neq M$ and 0 if $\hat{M} = M$. Using the chain rule and $H(E|M, \hat{M}) = 0$ (because E is a function of M and $\hat{M}$):

$$\begin{aligned} H(M|\hat{M}) =& H(M|\hat{M}, E) + H(E|\hat{M}) - H(E|M, \hat{M}) \\ =& H(M|\hat{M}, E = 1)\Pr(E = 1) + H(M|\hat{M}, E = 0)\Pr(E = 0) + H(E|\hat{M}) \end{aligned}$$

$H(M|\hat{M}, E = 0) = 0$ because, conditioned on $E = 0$ (i.e. $\hat{M} = M$) M is a function of $\hat{M}$. $H(M|\hat{M}, E = 1) \leq \log(|\mathcal{A}| - 1)$ because, for all $a \in \mathcal{A}$, the support of $P_{M|\hat{M}=a,E=1}$ is no larger than $|\mathcal{A}| - 1$. Finally, by positivity of mutual information, $H(E|\hat{M}) \leq H(E) = h(\Pr(\hat{M} \neq M))$. □

16 Channel coding

A **binary symmetric channel** with bit flip probability f is a model of a noisy classical communication channel. With each use of the channel we get to input a binary value, and the channel output is, with probability $1-f$, the same value or, with probability f, the other value. Furthermore, each use of the channel behaves independently of the other uses. If we make n uses of the channel and the input is $X^{(n)} = (X_1, \ldots, X_n)$ and the output is $Y^{(n)} = (Y_1, \ldots, Y_n)$

$$P_{Y^{(n)}|X^{(n)}}((y_1, \ldots, y_n)|(x_1, \ldots, x_n)) = \Pi_{i=1}^{n} P_{Y|X}(y_i|x_i), \tag{16.1}$$

where

$$P_{Y|X}(y|x) = \begin{cases} 1-f \text{ if } y = x \\ f \text{ if } y \neq x. \end{cases} \tag{16.2}$$

A channel where each use behaves identically and independently, which is the property described by equation (16.1), is called a **memoryless channel**.

Suppose that Alice is connected to Bob by this channel, so that she makes inputs and Bob receives outputs. Suppose Alice wants to transmit a single bit M to Bob using the channel. Let us denote by $\hat{M}$ the decoding of the message that Bob makes based on the channel output. If only one use of the channel is made then it is not hard to see that, for $f \leq 1/2$, the worst-case error probability

$$\max_{m \in \{0,1\}} \Pr(\hat{M} \neq M | M = m)$$

will be at least f. Provided $f < 1/2$, it is possible to do better by making more uses of the channel: A "repetition code" of "blocklength" n, makes n uses of the channel to send one bit. If $M = 0$ then Alice sends a string of n zeros and if $M = 1$ then she sends a string of n ones. Bob decodes the output of the channel by "majority vote": If more than half of the output symbols are ones, then he decodes to one, and otherwise to zero. For $n = 3$, we have

$$\Pr(\hat{M} = 0 | M = 1) = \Pr(\hat{M} = 1 | M = 0) = f^3 + \binom{3}{1} f^2(1-f), \tag{16.3}$$

which is roughly 0.03 when $f = 0.1$. It is not hard to show that, for any $f < 1/2$, the two error probabilities go to zero as n increases, but so does the *rate* of the code which is $1/n$ bits per channel use. As we will see, in general it is possible to do much better.

16.1 Coding over quantum channels

Definition 16.1. A (k, ϵ)-code for an operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ consists of

- an encoding function $\underline{\rho}$ from $\mathcal{A}_M = \{1, \ldots, k\}$ to states of A and
- a decoding POVM $E : \mathcal{A}_M \to \mathcal{L}(\mathcal{H}_\mathsf{B})$

such that

$$\forall m \in \mathcal{A}_M, \mathrm{Tr}(\mathbb{1} - E(m))\mathcal{N}\underline{\rho}(m) \leq \epsilon. \tag{16.4}$$

We call k the "size" of the code and define

$$k_\epsilon(\mathcal{N}) := \max\{k \in \mathbb{N} : \text{ There exists a } (k, \epsilon)\text{-code for } \mathcal{N}\}.$$

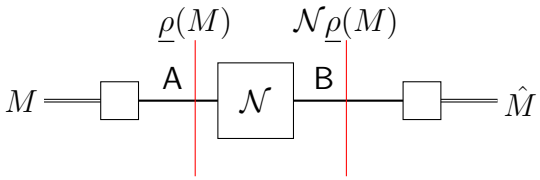


Figure 16.1: Alice uses $\underline{\rho}$ to encode a random variable M (taking values in $\mathcal{A}_M$) as a state $\underline{\rho}(M)$ of system A. The operation $\mathcal{N}$ is applied leaving B in the state $\mathcal{N}\underline{\rho}(M)$. Bob measures the POVM E on B producing a result $\hat{M}$ (also taking values in $\mathcal{A}_M$).

The condition (16.4) means that, if Alice uses a (k, ϵ)-code to transmit a random variable M, as in the figure, then the worst-case error probability is no more than ϵ:

$$\forall m \in \mathcal{A}_M : \Pr(\hat{M} \neq M | M = m) = \mathrm{Tr}(\mathbb{1} - E(m))\mathcal{N}\underline{\rho}(m) \leq \epsilon. \tag{16.5}$$

16.1.1 Memoryless quantum channels

A **memoryless quantum channel** is completely specified by its behaviour for a single use, which is given by some operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ where A is an input system for a single use of the channel and B the corresponding output system. Making n uses of the channel on n input systems $\mathsf{A}_1, \ldots, \mathsf{A}_n$ in an arbitrary state $\rho_{\mathsf{A}_1\cdots\mathsf{A}_n}$ results in output systems $\mathsf{B}_1, \ldots, \mathsf{B}_n$ (each of the same dimension as B) in the state

$$\mathcal{N}^{\otimes n}\rho_{\mathsf{A}_1\cdots\mathsf{A}_n} \text{ where } \mathcal{N}^{\otimes n} = \bigotimes_{i=1}^{n} \mathcal{N}^{\mathsf{B}_i\leftarrow\mathsf{A}_i}$$

and $\mathcal{N}^{\mathsf{B}_i\leftarrow\mathsf{A}_i} = \mathbf{id}^{\mathsf{B}_i\leftarrow\mathsf{B}}\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}\mathbf{id}^{\mathsf{A}\leftarrow\mathsf{A}_i}$. The **classical capacity** of the channel (in bits per channel use) is

Definition 16.2.

$$C(\mathcal{N}) := \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\lfloor\log(k_\epsilon(\mathcal{N}^{\otimes n}))\rfloor. \tag{16.6}$$

16.2 The Holevo bound and the HSW theorem

Suppose that X is a random variable, $\underline{\sigma}$ is a function from $\mathcal{A}_X$ to states of a system B, and the system B is in the state $\underline{\sigma}(X)_{\mathsf{B}}$. If X is stored in system $\tilde{\mathsf{X}}$ then the state of $\tilde{\mathsf{X}}\mathsf{B}$ is $\sigma_{\tilde{\mathsf{X}}\mathsf{B}} = \sum_{x\in\mathcal{A}_X} P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \underline{\sigma}(x)_{\mathsf{B}}$.

Proposition 16.3. In the situation just described $H(\mathsf{B}|\tilde{\mathsf{X}})_\sigma = \sum_x P_X(x) S(\underline{\sigma}(x)_{\mathsf{B}})$.

Proof. For each x, let $\underline{\sigma}(x) = \sum_{j=1}^{d_{\mathsf{B}}} \lambda(j|x)|\alpha_j^{(x)}\rangle\langle\alpha_j^{(x)}|$ be an eigendecomposition. $\sum_{j=1}^{d_{\mathsf{B}}} \lambda(j|x) = 1$ and $\lambda(j|x) \geq 0$. Therefore, we can imagine that there is a random variable J distributed such that $P_{J|X=x}(j) = \lambda(j|x)$ and, therefore, $P_{JX} = \lambda(j|x)P_X$ and

$$\sigma_{\tilde{\mathsf{X}}\mathsf{B}} = \sum_x \sum_j P_{JX}(j,x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |\alpha_j^{(x)}\rangle\langle\alpha_j^{(x)}|_{\mathsf{B}}$$

is an eigendecomposition. Therefore, using the chain rule,

$$\begin{aligned} S(\sigma_{\tilde{\mathsf{X}}\mathsf{B}}) =& S(P_{XJ}) = H(X,J) = H(X) + H(J|X) && (16.7)\\ =& H(X) + \sum_x H(J|X=x)P_X(x) && (16.8)\\ =& S(P_X) + \sum_x S(P_{J|X=x})P_X(x) && (16.9)\\ =& S(\sigma_{\tilde{\mathsf{X}}}) + \sum_x S(\underline{\sigma}(x)_{\mathsf{B}})P_X(x), && (16.10) \end{aligned}$$

and, since $H(\mathsf{B}|\tilde{\mathsf{X}}) = S(\sigma_{\tilde{\mathsf{X}}\mathsf{B}}) - S(\sigma_{\tilde{\mathsf{X}}})$, we have the result. □

If a POVM E is measured on B with result Y, then the **Holevo bound** says that

$$I(X:Y) \leq I(\tilde{\mathsf{X}}:\mathsf{B})_\sigma = S\left(\sum_{x\in\mathcal{A}_X} P_X(x)\underline{\sigma}(x)_{\mathsf{B}}\right) - \sum_{x\in\mathcal{A}_X} P_X(x) S\left(\underline{\sigma}(x)_{\mathsf{B}}\right). \tag{16.11}$$

Here the equality comes from $I(\tilde{\mathsf{X}}:\mathsf{B})_\sigma = H(\mathsf{B})_\sigma - H(\mathsf{B}|\tilde{\mathsf{X}})_\sigma$ and the previous proposition. The inequality follows almost immediately from the data processing inequality for QMI: Let $\mathcal{D}^{\tilde{\mathsf{Y}}\leftarrow\mathsf{B}}$ be the operation

$$\mathcal{D}^{\tilde{\mathsf{Y}}\leftarrow\mathsf{B}} : \tau_{\mathsf{B}} \mapsto \sum_{y\in\mathcal{A}_Y} |y\rangle\langle y|_{\tilde{\mathsf{Y}}} \mathrm{Tr} E(y)_{\mathsf{B}}\tau_{\mathsf{B}} \tag{16.12}$$

which measures the POVM E on B and store the result Y in system $\tilde{\mathsf{Y}}$ in the standard way. Then the DPI for QMI tells us that

$$I(X:Y) = I(\tilde{\mathsf{X}}:\tilde{\mathsf{Y}})_{\mathcal{D}^{\tilde{\mathsf{Y}}\leftarrow\mathsf{B}}\sigma_{\tilde{\mathsf{X}}\mathsf{B}}} \leq I(\tilde{\mathsf{X}}:\mathsf{B})_{\sigma_{\tilde{\mathsf{X}}\mathsf{B}}}. \tag{16.13}$$

16.2.1 The Holevo information of an operation

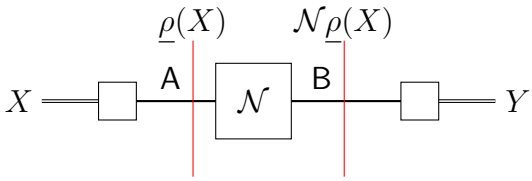
Suppose that, rather than the state $\underline{\sigma}(X)_{\mathsf{B}}$ being prepared directly, it is the result of first preparing the state $\underline{\rho}(X)_{\mathsf{A}}$ which is then which is then acted on by an operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$. If we maximise $I(X:Y)$ over all choices of the set $\mathcal{A}_X$, distribution P_X and function $\underline{\rho}$, as well as over all POVMs E, then we can regard the result as a measure of how large a classical correlation can be created via the operation. From the Holevo bound, it follows immediately that the maximal value of $I(X:Y)$ is bounded above by $\chi(\mathcal{N})$, the **Holevo information** of the operation $\mathcal{N}$, which is defined by

Definition 16.4.

$$\chi(\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}) := \max_{\mathcal{A}_X, P_X, \underline{\rho}} I(\tilde{\mathsf{X}}:\mathsf{B})_{\sigma_{\tilde{\mathsf{X}}\mathsf{B}}} = \max_{\mathcal{A}_X, P_X, \underline{\rho}} \left(S\left(\sum_x P_X(x)\mathcal{N}\underline{\rho}(x)\right) - \sum_x P_X(x) S\left(\mathcal{N}\underline{\rho}(x)\right)\right)$$

where $\sigma_{\tilde{\mathsf{X}}\mathsf{B}} = \sum_{x\in\mathcal{A}_X} P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes (\mathcal{N}\underline{\rho}(x))_{\mathsf{B}}$.

Remark 16.5. It is possible to show that the maximum is achieved for some P_X and $\underline{\rho}$ when $|\mathcal{A}_X| = d_{\mathsf{A}}^2$.



Proposition 16.6. Suppose a system A is prepared in the state $\underline{\rho}(X)$, and then operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ acts, leaving B in the state $\mathcal{N}\underline{\rho}(X)$, and then a POVM E is measured on B producing result Y. (The situation illustrated in the figure above.) For any $\mathcal{A}_X$, P_X, $\underline{\rho}$, $\mathcal{A}_Y$ and POVM E, the mutual information $I(X:Y)$ is no more than the Holevo information $\chi(\mathcal{N})$ of the operation $\mathcal{N}$.

♣♣The proof of the following proposition is an exercise for example class 4.

Proposition 16.7. For any two operations $\mathcal{M}$ and $\mathcal{N}$, $\chi(\mathcal{N}\otimes\mathcal{M}) \geq \chi(\mathcal{N}) + \chi(\mathcal{M})$.

16.2.2 The Holevo-Schumacher-Westmoreland (HSW) theorem

It turns out that the classical capacity of an operation $\mathcal{N}$ is equal to its "regularised" Holevo information:

Theorem 16.8. For any operation $\mathcal{N}$, $C(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n})$.

This result is known as the Holevo-Schumacher-Westmoreland (HSW) theorem, and we will prove it, starting with the "converse" which shows that the right-hand-side is an upper bound on the left-hand-side. (That the limit on the RHS exists follows from Proposition 16.7 and Fekete's lemma - a basic result in analysis.)

16.3 HSW theorem: Converse part

Proposition 16.9. For any $\epsilon \in [0, 1)$ and operation $\mathcal{N}$

$$\log(k_\epsilon(\mathcal{N})) \leq \frac{\chi(\mathcal{N}) + 1}{1 - \epsilon}. \tag{16.14}$$

Proof. Suppose that there is a (k, ϵ)-code for $\mathcal{N}$ which we use to transmit a uniformly distributed message M (taking values in $\{1, \ldots, k\}$). Let $\hat{M}$ be the result of measuring the decoding POVM. The bound on the worst-case error probability implies that

$$P(\hat{M} \neq M) \leq \epsilon. \tag{16.15}$$

Using the fact that M is uniformly distributed, Fano's inequality, and $h(x) \leq 1$ for all $x \in [0, 1]$, and (16.15), we have

$$\begin{aligned} I(M : \hat{M}) =& H(M) - H(M|\hat{M}) && (16.16)\\ =& \log(k) - H(M|\hat{M}) && (16.17)\\ \geq& \log(k) - \Big(\Pr(\hat{M} \neq M) \log(k - 1) + 1\Big) && (16.18)\\ \geq& (1 - \epsilon) \log(k) - 1. && (16.19) \end{aligned}$$

Rearranging this and using Proposition 16.6 we have

$$\log(k) \leq \frac{I(M : \hat{M}) + 1}{1 - \epsilon} \leq \frac{\chi(\mathcal{N}) + 1}{1 - \epsilon}, \tag{16.20}$$

as claimed. □

Corollary 16.10 (HSW theorem: Converse part)**.**

$$C(\mathcal{N}) \leq \lim_{n \to \infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}). \tag{16.21}$$

Proof.

$$C(\mathcal{N}) \leq \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \frac{\chi(\mathcal{N}^{\otimes n}) + 1}{1 - \epsilon} = \lim_{\epsilon \to 0} (1 - \epsilon)^{-1} \lim_{n \to \infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}).$$

□

We will return to the achievability part later, but before that, we will show how the formula given for the capacity simplifies for certain kinds of channel.

16.4 Capacity of entanglement breaking channels

Definition 16.11. An operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ is **entanglement breaking** if, for all systems R and states ρ_{RA}, $\sigma_{\mathsf{RB}} = \mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}\rho_{\mathsf{RA}}$ is separable.

Theorem 16.12. If $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ and $\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}$ are operations, and $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ is entanglement-breaking, then $\chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}) = \chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}) + \chi(\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}})$.

Proof. We already know that $\chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}) \geq \chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}) + \chi(\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}})$ so we just need to show the opposite inequality.

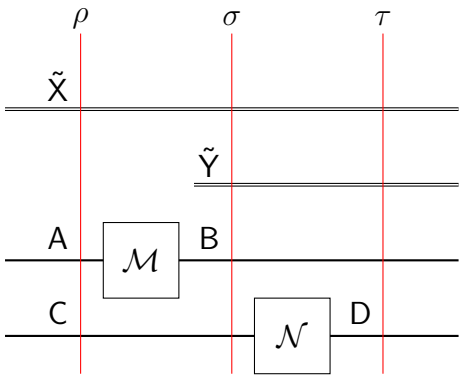


Figure 16.2: The systems, operations and states referred to in the proof.

Given some $\mathcal{A}_X$, P_X, and $\underline{\rho}$ let

1. $\rho_{\tilde{\mathsf{X}}\mathsf{AC}} = \sum_x P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \underline{\rho}(x)_{\mathsf{AC}}$.
2. $\sigma_{\tilde{\mathsf{X}}\mathsf{BC}} := \mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\rho_{\tilde{\mathsf{X}}\mathsf{AC}} = \sum_x P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes (\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\underline{\rho}(x)_{\mathsf{AC}})$. Because $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ is entanglement breaking, we know that, for all $x \in \mathcal{A}_X$,
$$\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\underline{\rho}(x)_{\mathsf{AC}} = \sum_r q(r|x)\underline{\beta}(x,r)_{\mathsf{B}} \otimes \underline{\gamma}(x,r)_{\mathsf{C}}$$
where, for all r, $\underline{\beta}(x,r)_{\mathsf{B}}$ and $\underline{\gamma}(x,r)_{\mathsf{C}}$ are states, $q(r|x) \geq 0$ and $\sum_r q(r|x) = 1$.
3. $\sigma_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BC}} := \sum_x \sum_r q(r|x)P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |r\rangle\langle r|_{\tilde{\mathsf{R}}}\underline{\beta}(x,r)_{\mathsf{B}} \otimes \underline{\gamma}(x,r)_{\mathsf{C}}$ defines an extension of $\sigma_{\tilde{\mathsf{X}}\mathsf{BC}}$.
4. $\tau_{\tilde{\mathsf{X}}\mathsf{BD}} := \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}\sigma_{\tilde{\mathsf{X}}\mathsf{BC}} = \mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}\rho_{\tilde{\mathsf{X}}\mathsf{AC}}$. So, $\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BD}} := \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}\sigma_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BC}}$ defines an extension of $\tau_{\tilde{\mathsf{X}}\mathsf{BD}}$.

Let $\mathcal{R}^{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B}\leftarrow\tilde{\mathsf{X}}\tilde{\mathsf{R}}}$ be the operation which measures the computational basis of $\tilde{\mathsf{X}}\tilde{\mathsf{R}}$ (measures the PVM) and adds system B prepared in the state $\underline{\beta}(x,r)_{\mathsf{B}}$, that is

$$\mathcal{R}^{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B}\leftarrow\tilde{\mathsf{X}}\tilde{\mathsf{R}}} : \eta_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}} \mapsto \sum_{x,r}(|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |r\rangle\langle r|_{\tilde{\mathsf{R}}}\eta_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}}|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |r\rangle\langle r|_{\tilde{\mathsf{R}}}) \otimes \underline{\beta}(x,r)_{\mathsf{B}}.$$

Since $\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BD}} = \mathcal{R}^{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B}\leftarrow\tilde{\mathsf{X}}\tilde{\mathsf{R}}}\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{D}}$ and $\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{D}} = \mathrm{Tr}_{\mathsf{B}}\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BD}}$, the DPI for QMI tells us that

$$I(\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B} : \mathsf{D})_\tau = I(\tilde{\mathsf{X}}\tilde{\mathsf{R}} : \mathsf{D})_\tau \tag{16.22}$$

By making use of these facts and the following instance of strong subadditivity

$$H(\mathsf{BD}|\tilde{\mathsf{X}})_\tau - H(\mathsf{B}|\tilde{\mathsf{X}})_\tau - H(\mathsf{D}|\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B})_\tau = H(\mathsf{D}|\tilde{\mathsf{X}}\mathsf{B})_\tau - H(\mathsf{D}|\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B})_\tau = I(\mathsf{D} : \tilde{\mathsf{R}}|\tilde{\mathsf{X}}\mathsf{B})_\tau \geq 0 \quad (16.23)$$

we find that

$$\begin{aligned} I(\tilde{\mathsf{X}} : \mathsf{BD})_\tau &\overset{(a)}{=} H(\mathsf{BD})_\tau - H(\mathsf{BD}|\tilde{\mathsf{X}})_\tau && (16.24)\\ &\overset{(b)}{\leq} H(\mathsf{B})_\tau + H(\mathsf{D})_\tau - H(\mathsf{BD}|\tilde{\mathsf{X}})_\tau && (16.25)\\ &\overset{(c)}{\leq} H(\mathsf{B})_\tau + H(\mathsf{D})_\tau - H(\mathsf{B}|\tilde{\mathsf{X}})_\tau - H(\mathsf{D}|\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B})_\tau && (16.26)\\ &\overset{(d)}{=} I(\tilde{\mathsf{X}} : \mathsf{B})_\tau + I(\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B} : \mathsf{D})_\tau && (16.27)\\ &\overset{(e)}{=} I(\tilde{\mathsf{X}} : \mathsf{B})_\tau + I(\tilde{\mathsf{X}}\tilde{\mathsf{R}} : \mathsf{D})_\tau && (16.28)\\ &\overset{(f)}{\leq} \chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}) + \chi(\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}). && (16.29) \end{aligned}$$

where (a) is by definition; (b) is $I(\mathsf{B} : \mathsf{D}) \geq 0$; (c) is (16.23); (d) is by definition; (e) is (16.22); and (f) is by definition of the Holevo information. Since this bound is true for *any* $\mathcal{A}_X$, P_X and $\underline{\rho}$, we have

$$\chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}) = \max_{\mathcal{A}_X, P_X, \underline{\rho}} I(\tilde{\mathsf{X}} : \mathsf{BD})_\tau \leq \chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}) + \chi(\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}})$$

as required.

□

It is easily verified that the tensor product of two (or more) entanglement breaking operations is entanglement breaking. Therefore, if $\mathcal{N}$ is entanglement breaking, we have $\chi(\mathcal{N}^{\otimes n}) = \chi(\mathcal{N}^{\otimes(n-1)}) + \chi(\mathcal{N}) = n\chi(\mathcal{N})$ and, by the HSW theorem,

Corollary 16.13. If $\mathcal{N}$ is any entanglement-breaking operation $C(\mathcal{N}) = \chi(\mathcal{N})$.

Proposition 16.14. Given a conditional probability distribution $N_{Y|X}$ we can define an associated operation $\mathcal{N}^{\tilde{\mathsf{Y}}\leftarrow\tilde{\mathsf{X}}}$ by

$$\mathcal{N}^{\tilde{\mathsf{Y}}\leftarrow\tilde{\mathsf{X}}} : \rho_{\tilde{\mathsf{X}}} \mapsto \sum_{y,x} N_{Y|X}(y|x)|y\rangle\langle x|\rho_{\tilde{\mathsf{X}}}|x\rangle\langle y|.$$

$$C(\mathcal{N}^{\tilde{\mathsf{Y}}\leftarrow\tilde{\mathsf{X}}}) = \max_{P_X} I(X : Y) \text{ where } P_{XY}(x,y) = N_{Y|X}(y|x)P_X(x)$$

and the maximisation is over probability distributions P_X on $\mathcal{A}_X$.

♣♣Proving this is an exercise for example class 4.

16.5 HSW Theorem: Achievability part

Given an Hermitian operation L with eigendecomposition $L = \sum_i \lambda_i|\alpha_i\rangle\langle\alpha_i|$, we define for $u \in \mathbb{R}$,

$$L^u = \sum_{i:\lambda_i\neq 0} \lambda_i^u |\alpha_i\rangle\langle\alpha_i|.$$

Note that for any $u, v \in \mathbb{R}$, $L^u L^v = L^{u+v}$ and that L^0 is the projector onto $\mathrm{supp}(L)$.

Lemma 16.15 (Hayashi-Nagaoka)**.** For any real $c > 0$ and operators S, R (on the same Hilbert space) such that $0 \leq S \leq \mathbb{1}$ and $0 \leq R$,

$$\mathbb{1} - (S+R)^{-1/2} S (S+R)^{-1/2} \leq (1+c)(\mathbb{1} - S) + (2 + c + c^{-1})R \tag{16.30}$$

Proof. **(Not examinable.)** For any real c and operators A and B

$$(A - cB)^\dagger (A - cB) = A^\dagger A + c^2 B^\dagger B - c(A^\dagger B + B^\dagger A) \geq 0. \tag{16.31}$$

If $c > 0$ then

$$A^\dagger B + B^\dagger A \leq c^{-1} A^\dagger A + c B^\dagger B. \tag{16.32}$$

Setting $A = R^{1/2}$ and $B = R^{1/2}(X - \mathbb{1})$ for some hermitian X this becomes

$$R(X - \mathbb{1}) + (X - \mathbb{1})R \leq c^{-1} R + c(X - \mathbb{1})R(X - \mathbb{1}). \tag{16.33}$$

Equivalently,

$$\begin{aligned} XRX =& (X - \mathbb{1})R(X - \mathbb{1}) + R + R(X - \mathbb{1}) + (X - \mathbb{1})R & (16.34)\\ \leq& (1+c)(X - \mathbb{1})R(X - \mathbb{1}) + (1 + c^{-1})R & (16.35) \end{aligned}$$

and, since $S \geq 0$,

$$XRX \leq (1+c)(X - \mathbb{1})(S+R)(X - \mathbb{1}) + (1 + c^{-1})R. \tag{16.36}$$

Since square-root is operator monotone, and $R \geq 0$ and $0 \leq S \leq \mathbb{1}$ we have

$$(S+R)^{1/2} \geq S^{1/2} \geq S. \tag{16.37}$$

If we take $X = (S+R)^{-1/2}$ then $\Pi := X(S+R)X$ is the projector onto $\mathrm{supp}(S+R)$ and

$$(X - \mathbb{1})(S+R)(X - \mathbb{1}) = \Pi + S + R - 2(S+R)^{1/2} \leq \Pi + R - S, \tag{16.38}$$

where the inequality is (16.37). Combining this upper bound with (16.36) yields

$$XRX \leq (1+c)(\Pi - S) + (2 + c + c^{-1})R \tag{16.39}$$

and $XRX = \Pi - (S+R)^{-1/2} S (S+R)^{-1/2}$ so

$$\Pi - (S+R)^{-1/2} S (S+R)^{-1/2} \leq (1+c)(\Pi - S) + (2 + c + c^{-1})R. \tag{16.40}$$

Because $\mathbb{1} - \Pi \geq 0$ and $c > 0$, $(\mathbb{1} - \Pi) \leq (1+c)(\mathbb{1} - \Pi)$. Adding this to (16.40) yields (16.30). □

Lemma 16.16. If we have a code which can send a uniformly distributed message M from a set $\mathcal{A}_M$ of size k with average error probability $\Pr(\hat{M} \neq M) = \bar{\epsilon}$ then the same code can send a message from a subset of $\mathcal{A}_M$ of size $\lceil \frac{k}{2} \rceil$ with worst-case error probability no more than $2\bar{\epsilon}$.

Proof. Label the elements of $\mathcal{A}_M$, $m_1, m_2, \ldots, m_k$, such that the probabilities

$$p_i := \Pr(\hat{M} \neq M | M = m_i)$$

are in increasing order $p_1 \leq p_2 \leq \cdots \leq p_k$.

$$\begin{aligned} \bar{\epsilon} =& \frac{1}{k} \left(\sum_{i=1}^{\lfloor k/2 \rfloor} p_i + \sum_{i=\lceil k/2 \rceil}^{k} p_i \right) && (16.41) \\ \geq& \frac{1}{k} \left(0 + (k + 1 - \lceil k/2 \rceil) p_{\lceil k/2 \rceil} \right) && (16.42) \\ \geq& \frac{1}{k} \frac{k+1}{2} p_{\lceil k/2 \rceil} \geq \frac{1}{2} p_{\lceil k/2 \rceil}. && (16.43) \end{aligned}$$

Therefore the subset $\{m_1, \ldots, m_{\lceil k/2 \rceil}\}$ has the required property. □

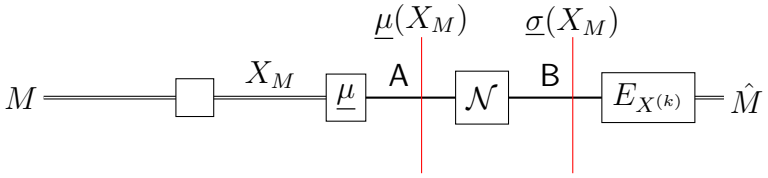


Figure 16.3: Depiction of the random codes used in the proof of Theorem 16.17.

Theorem 16.17. Suppose that $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ is an operation, $k \geq 1$ is an integer, and M is a uniformly distributed RV taking values in $\{1, \ldots, k\}$. Given any finite set $\mathcal{A}_X$, distribution P_X on that set, and map $\underline{\mu} : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_\mathsf{A})$ from $\mathcal{A}_X$ to states of A let

$$\sigma_{\mathsf{XB}} := \sum_{x \in \mathcal{A}_X} P_X(x) |x\rangle\langle x|_\mathsf{X} \otimes \underline{\sigma}(x)_\mathsf{B}, \tag{16.44}$$

where $\underline{\sigma}(x)_\mathsf{B} := \mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}} \underline{\mu}(x)_\mathsf{A}$. Given k "codewords" $(x_1, x_2, \ldots, x_k) \in \mathcal{A}_X^k$ we encode M as the state $\underline{\mu}(x_M)_\mathsf{A}$; the operation $\mathcal{N}$ is applied; then we measure a decoding POVM $E : \{0, 1, \ldots, k\} \to \mathcal{L}(\mathcal{H}_\mathsf{B})$ producing result $\hat{M}$. We claim that there exists a set of k codewords and a decoding POVM E such that the average probability of error $\Pr(\hat{M} \neq M)$ is equal to $\bar{\epsilon}$, where $\bar{\epsilon}$ satisfies

$$k \geq \frac{\bar{\epsilon}}{16} \frac{1}{\beta_{\bar{\epsilon}/2}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_\mathsf{B})} \tag{16.45}$$

and therefore[1]

$$k_\epsilon(\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}) \geq \frac{\epsilon}{64} \frac{1}{\beta_{\epsilon/4}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_\mathsf{B})}. \tag{16.46}$$

[1] I missed out a factor of half in the lecture.

Proof. **Random codes to specific codes:** Suppose that the k codewords are chosen at random from $\mathcal{A}_X$. We represent these as random variables $X_1, \ldots, X_k$ taking values in $\mathcal{A}_{\mathsf{X}}$ and write $X^{(k)} = (X_1, \ldots, X_k)$. We assume that both Alice and Bob know $X^{(k)}$, so Bob's decoding POVM can depend on $X^{(k)}$. We will use the following reasoning, which holds for any way of choosing the decoding POVM and any distribution of the k codewords:

$$\bar{\epsilon} = \min_{\underline{x} \in \mathcal{A}_X^k} \Pr(\hat{M} \neq M | X^{(k)} = \underline{x}) \tag{16.47}$$

$$\leq \Pr(\hat{M} \neq M) = \sum_{\underline{x} \in \mathcal{A}_X^k} P_{X^{(k)}}(\underline{x}) \Pr(\hat{M} \neq M | X^{(k)} = \underline{x}). \tag{16.48}$$

The expression (16.47) is the minimum average error probability attained by some particular choice of codewords $\underline{x} = (x_1, \ldots, x_k)$, so we can set this equal to the $\bar{\epsilon}$ of the theorem. This is a lower-bound on the average error probability (16.48) which is achieved when the codewords are chosen at random (with the decoding POVM chosen depending on the codewords).

The decoding POVM: Recall that

$$\beta_{\epsilon'}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}) = \min\{\beta(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}) : \alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) \leq \epsilon, 0 \leq T_{\tilde{\mathsf{X}}\mathsf{B}} \leq \mathbb{1}\} \tag{16.49}$$

where $\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) = 1 - \mathrm{Tr} T_{\tilde{\mathsf{X}}\mathsf{B}} \sigma_{\tilde{\mathsf{X}}\mathsf{B}}$ and $\beta(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}) = \mathrm{Tr} T_{\tilde{\mathsf{X}}\mathsf{B}} \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}$. In this particular case, defining

$$L(x)_{\mathsf{B}} := \mathrm{Tr}_{\tilde{\mathsf{X}}} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}} T_{\tilde{\mathsf{X}}\mathsf{B}} \tag{16.50}$$

we find that

$$\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) = 1 - \sum_{x \in \mathcal{A}_X} \mathrm{Tr} L(x)_{\mathsf{B}} \underline{\sigma}(x)_{\mathsf{B}} \text{ and} \tag{16.51}$$

$$\beta(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}) = \sum_{x, x' \in \mathcal{A}_X} P_X(x) P_X(x') \mathrm{Tr} L(x)_{\mathsf{B}} \underline{\sigma}(x')_{\mathsf{B}}. \tag{16.52}$$

By partial trace cyclicity, $L(x)_{\mathsf{B}} = \mathrm{Tr}_{\tilde{\mathsf{X}}} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}} T_{\tilde{\mathsf{X}}\mathsf{B}} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}}$, so $L(x)_{\mathsf{B}} \geq 0$ and, because $T_{\tilde{\mathsf{X}}\mathsf{B}} \leq \mathbb{1}_{\tilde{\mathsf{X}}\mathsf{B}}$,

$$|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}} T_{\tilde{\mathsf{X}}\mathsf{B}} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}} \leq |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}}$$

and, therefore, $L(x)_{\mathsf{B}} \leq \mathbb{1}_{\mathsf{B}}$.

In the random code, Bob will measure $E_{X^{(k)}}$ where, for each possible value $\underline{x} = (x_1, \ldots, x_k) \in \mathcal{A}_X^k$ of the random codewords $X^{(k)}$, we build a decoding POVM $E_{\underline{x}}$ based on the operators $L(x)$, as follows: For $m \in \{1, \ldots, k\}$ we set

$$E_{\underline{x}}(m) := \left(\sum_{i=1}^k L(x_i) \right)^{-1/2} L(x_m) \left(\sum_{i=1}^k L(x_i) \right)^{-1/2}. \tag{16.53}$$

Clearly these are positive operators, and

$$\sum_{m=1}^k E_{\underline{x}}(m) = \left(\sum_{m=1}^k L(x_m) \right)^0$$

which is the projector onto the support of $\sum_{m=1}^k L(x_m)$. Therefore, setting

$$E_{\underline{x}}(0) := \mathbb{1} - \sum_{i=1}^m E_{\underline{x}}(m) \tag{16.54}$$

we have $E_{\underline{x}}(0) \geq 0$ and $\sum_{m=0}^{k} E_{\underline{x}}(m) = \mathbb{1}$ so $E_{\underline{x}}$ is indeed a POVM.

Bounding $\Pr(\hat{M} \neq M)$ for the random code: Since the message M is uniformly distributed, we have

$$\Pr(\hat{M} \neq M) = \frac{1}{k} \sum_{m=1}^{k} \Pr(\hat{M} \neq M | M = m) \tag{16.55}$$

and because $X^{(k)}$ is independent of M,

$$\Pr(\hat{M} \neq M | M = m) = \sum_{\underline{x}} \Pr(\hat{M} \neq M | M = m, X^{(k)} = \underline{x}) \Pr(X^{(k)} = \underline{x} | M = m) \tag{16.56}$$

$$= \sum_{\underline{x}} \Pr(\hat{M} \neq M | M = m, X^{(k)} = \underline{x}) \Pr(X^{(k)} = \underline{x}) = \mathbb{E} p_m \tag{16.57}$$

where $\mathbb{E}p_m$ is the expectation of the random variable

$$p_m := \mathrm{Tr}(\mathbb{1} - E_{X^{(k)}}(m)) \underline{\sigma}(X_m). \tag{16.58}$$

For the decoding POVM we are using, this is

$$p_m = \mathrm{Tr}\left(\mathbb{1} - (S+R)^{-1/2} S (S+R)^{-1/2}\right) \underline{\sigma}(X_m) \tag{16.59}$$

where $S = L(X_m)$ and $R = \sum_{i \neq m} L(X_i)$. We already showed that $0 \leq S \leq \mathbb{1}$, and clearly $R \geq 0$, so the Hayashi-Nagaoka operator inequality tells us that

$$\forall c \geq 0,\ \mathbb{1} - (S+R)^{-1/2} S (S+R)^{-1/2} \leq (1+c)(\mathbb{1} - S) + (2 + c + c^{-1}) R. \tag{16.60}$$

Since $\underline{\sigma}(X_m) \geq 0$, for all m we have

$$\begin{aligned} p_m \leq & \mathrm{Tr}((1+c)(\mathbb{1} - S) + (2 + c + c^{-1}) R) \underline{\sigma}(X_m) \\ = & (1+c)\left(1 - \mathrm{Tr} L(X_m) \underline{\sigma}(X_m)\right) + (2 + c + c^{-1}) \sum_{i \neq m} \mathrm{Tr} L(X_i) \underline{\sigma}(X_m) \end{aligned} \tag{16.61}$$

where we used $\mathrm{Tr}\underline{\sigma}(X_m) = 1$. Now, using the fact that each X_m is distributed according to P_X, we have

$$\forall m\ \mathbb{E}\left(1 - \mathrm{Tr} L(X_m) \underline{\sigma}(X_m)\right) = 1 - \sum_{x \in \mathcal{A}_X} P_X(x) \mathrm{Tr}_{\mathsf{B}} L(x) \underline{\sigma}(x) = \alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) \tag{16.62}$$

Using, in addition, the independence of X_m and X_i for all $i \neq m$,

$$\forall i \neq m,\ \mathbb{E}\mathrm{Tr} L(X_i) \underline{\sigma}(X_m) = \sum_{x,x'} P_X(x) P_X(x') \mathrm{Tr} L(x) \underline{\sigma}(x') = \beta(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}). \tag{16.63}$$

Taking the expectation of both sides of (16.61) and using (16.62) and (16.63) we obtain

$$\forall m,\ \Pr(\hat{M} \neq M | M = m) = \mathbb{E} p_m \leq (1+c)\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) + (2 + c + c^{-1})(k-1)\beta(T_{\tilde{\mathsf{X}}\mathsf{B}} \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}). \tag{16.64}$$

and therefore

$$\bar{\epsilon} \leq \Pr(\hat{M} \neq M) \leq (1+c)\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) + (2 + c + c^{-1})(k-1)\beta(T_{\tilde{\mathsf{X}}\mathsf{B}} \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}).$$

Minimising the right-hand-side over all $T_{\tilde{\mathsf{X}}\mathsf{B}}$ such that $\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) \leq \epsilon'$ and $0 \leq T_{\tilde{\mathsf{X}}\mathsf{B}} \leq \mathbb{1}_{\tilde{\mathsf{X}}\mathsf{B}}$ we obtain

$$\bar{\epsilon} \leq (1+c)\epsilon' + (2 + c + c^{-1})(k-1)\beta_{\epsilon'}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}).$$

We have shown that, for any $c \geq 0$, there exists a code with k codewords and average error probability $\bar{\epsilon}$ where

$$k - 1 \geq \frac{\bar{\epsilon} - (1+c)\epsilon'}{2 + c + c^{-1}} \frac{1}{\beta_{\epsilon'}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}})} \tag{16.65}$$

Taking $\epsilon' = \bar{\epsilon}/2$ and (the optimal value) $c = 1/3$

$$k \geq \frac{\bar{\epsilon}}{16} \frac{1}{\beta_{\bar{\epsilon}/2}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}})}. \tag{16.66}$$

Finally, since there exists a set of codewords of k' and average probability of error $\epsilon/2$ such that

$$k' \geq \frac{\epsilon}{32} \frac{1}{\beta_{\epsilon/4}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}})},$$

we know from Lemma 16.16 that there is set of codewords of size $\lceil k'/2 \rceil$ with *worst-case* error probability ϵ. □

Remark 16.18. We can interpret the operators $L(x)$ in the following way: For each $x \in \mathcal{A}_X$ let $F_x : \{0,1\} \to \mathcal{L}(\mathcal{H}_{\mathsf{B}})$ by a POVM with

$$F_x(0)_{\mathsf{B}} = L(x)_{\mathsf{B}}, F_x(1)_{\mathsf{B}} = \mathbb{1}_{\mathsf{B}} - L(x)_{\mathsf{B}}.$$

Suppose we measure $\tilde{\mathsf{X}}$ in the computational basis, obtaining an outcome X, and then measure the POVM F_X on B, and use the outcome of that measurement as our guess of which hypothesis is true. This procedure has exactly the α and β given above. We can interpret $L(x)$ as specifying a test which tries to decide between the particular state $\underline{\sigma}(x)_{\mathsf{B}}$ and the average $\sigma_{\mathsf{B}} = \sum_{x \in \mathcal{A}_X} P_X(x) \underline{\sigma}(x)_{\mathsf{B}}$.

Proposition 16.19 (HSW Theorem: Achievability)**.**

$$C(\mathcal{N}) = \lim_{\ell \to \infty} \frac{1}{\ell} C(\mathcal{N}^{\otimes \ell}) \geq \frac{1}{\ell} \chi(\mathcal{N}^{\otimes \ell}). \tag{16.67}$$

Proof. We consider codes for the operation $\mathcal{N}^{\otimes n}$. For a given $\mathcal{A}_X$, distribution P_X and map $\underline{\mu}$ from $\mathcal{A}_X$ to states of A, we can use $\mathcal{A}_X^n$ and P_X^n as the set and distribution in the premises of Theorem 16.17 and for the map use

$$\underline{\mu}^{\otimes n} : \mathcal{A}_X^n \to \mathcal{L}(\mathcal{H}_{\mathsf{A}}^{\otimes n}) : (x_1, \ldots, x_n) \mapsto \underline{\mu}(x_1)_{\mathsf{A}_1} \otimes \ldots \otimes \underline{\mu}(x_n)_{\mathsf{A}_i} \text{ where } \underline{\mu}(x)_{\mathsf{A}_i} = \mathbf{id}^{\mathsf{A}_i \leftarrow \mathsf{A}} \underline{\mu}(x)_{\mathsf{A}}.$$

The "null hypothesis" state which appears in Theorem 16.17 is

$$\sigma_{\tilde{\mathsf{X}}_1 \mathsf{B}_1 \cdots \tilde{\mathsf{X}}_n \mathsf{B}_n} = \bigotimes_{i=1}^{n} \sigma_{\tilde{\mathsf{X}}_i \mathsf{B}_i} \text{ where } \sigma_{\tilde{\mathsf{X}}_i \mathsf{B}_i} = \sum_x P_X(x) |x\rangle\langle x|_{\tilde{\mathsf{X}}_i} \otimes \underline{\sigma}(x)_{\mathsf{B}_i}$$

where $\underline{\sigma}(x)_{\mathsf{B}_i} = \mathcal{N}^{\mathsf{B}_i \leftarrow \mathsf{A}_i} \underline{\mu}(x)_{\mathsf{A}_i}$ while the "alternative hypothesis" state is $\bigotimes_{i=1}^n \sigma_{\tilde{\mathsf{X}}_i} \otimes \sigma_{\mathsf{B}_i}$. Using Theorem 16.17 and the Quantum Stein's lemma, we have, for any $\epsilon > 0$

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} \lfloor \log \left(k_\epsilon(\mathcal{N}^{\otimes n}) \right) \rfloor &\geq \lim_{n \to \infty} \frac{1}{n} \left(\log \frac{\epsilon}{32} - \log \beta_{\epsilon/4} \left(\bigotimes_{i=1}^{n} \sigma_{\tilde{\mathsf{X}}_i \mathsf{B}_i} \middle\| \bigotimes_{i=1}^{n} \sigma_{\tilde{\mathsf{X}}_i} \otimes \sigma_{\mathsf{B}_i} \right) \right) \\ &= D\left(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}} \right) = I(\tilde{\mathsf{X}} : \mathsf{B})_\sigma, \end{aligned}$$

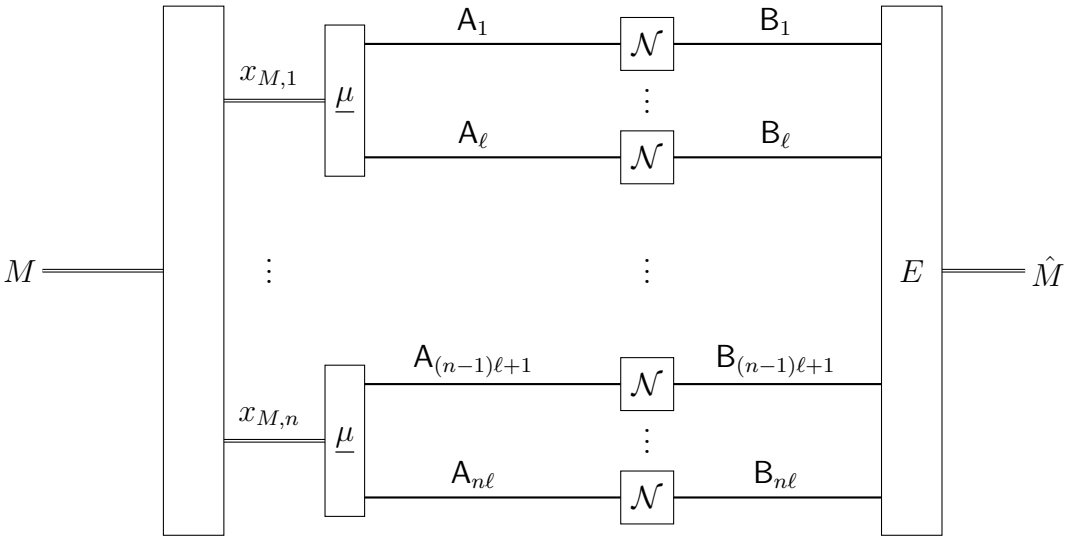


Figure 16.4: There exist codes of this form which, in the large n limit, achieve the rate $\frac{1}{\ell}\chi(\mathcal{N}^{\otimes\ell})$.

and therefore

$$C(\mathcal{N}) \geq I(\tilde{\mathsf{X}} : \mathsf{B})_\sigma.$$

Maximising this lower bound over choice of $\mathcal{A}_X$, P_X, and $\underline{\mu}$ we obtain

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

Now, $C(\mathcal{N}) = \frac{1}{\ell}C(\mathcal{N}^{\otimes\ell})$ so we also have, for all $\ell \in \mathbb{N}$,

$$C(\mathcal{N}) = \frac{1}{\ell}C(\mathcal{N}^{\otimes\ell}) \geq \frac{1}{\ell}\chi(\mathcal{N}^{\otimes\ell}).$$

Since the Holevo information is superadditive, we obtain the highest lower bound by taking the large ℓ limit, and this proves the "achievability" part of the HSW theorem. □

Note that, to achieve the lower bound $\frac{1}{\ell}\chi(\mathcal{N}^{\otimes\ell})$ on capacity we optimise over maps $\underline{\mu}$ which take some set $\mathcal{A}_X$ to states of ℓ input system, and then consider random codes for $\mathcal{N}^{\otimes\ell n}$ whose codewords X_i are strings $(X_{i,1}, \ldots X_{i,n})$ taking values in $\mathcal{A}_X^n$ with distribution P_X^n. This is the same as saying that each *symbol* $X_{i,j}$ in each string X_i is chosen i.i.d. with distribution P_X. The inputs to the channel which are states of the form

$$\underline{\mu}(x_{m,1})_{\mathsf{A}_1\cdots\mathsf{A}_\ell} \otimes \underline{\mu}(x_{m,2})_{\mathsf{A}_{\ell+1}\cdots\mathsf{A}_{2\ell}} \otimes \cdots \underline{\mu}(x_{m,n})_{\mathsf{A}_{(n-1)\ell+1}\cdots\mathsf{A}_{n\ell}}.$$

The possibility of having entanglement within the blocks of ℓ input systems does, for certain $\mathcal{N}$, allow us to do better as we make ℓ larger.