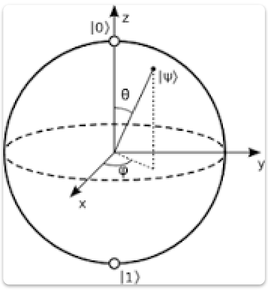
Classical vs Quantum Information Theory

#lectures

Sets vs Hilbert spaces

Classical	Quantum
Finite alphabet	Finite dimensional, complex Hilbert space
$\Sigma_A = \{1, 2, \ldots, d\}$	$\mathcal{H}_A \cong \mathbb{C}^d = \mathrm{span}(\|1\rangle, \|2\rangle, \ldots, \|d\rangle)$
Different letters $i \neq j$	Orthogonal vectors $x \perp y$

States

Classical	Quantum
Probability distributions $p \in \mathrm{P}(\Sigma_A)$	Density matrices $\rho \in \mathrm{D}(\mathcal{H}_A)$
$p \in \mathbb{R}^d$ such that $p_i \geq 0$ and $\sum_i p_i = 1$	$\rho \in \mathcal{B}(\mathcal{H}_A) \cong \mathcal{M}_d(\mathbb{C})$ such that ρ is *positive semidefinite* $\rho \geq 0$ and *unit trace* $\mathrm{Tr}\, \rho = 1$
	Classical states can be embedded as *diagonal matrices*
Extremal points: Dirac masses	Extremal points: *pure states*
$\mathrm{ext}\, \mathrm{P}(\Sigma_A) = \{\delta_i \,:\, i \in \Sigma_A\}$	$$\mathrm{ext}\, \mathrm{D}(\mathcal{H}_A) = \{\|x\rangle\langle x\| \,:\, x \in \mathcal{H}_A,\ \\|x\\| = 1\}$$
(real) dimension: $d - 1$	(real) dimension: $d^2 - 1$
In dimension $d = 2$: *bits* A segment $[0, 1]$ $(1 - p)\delta_0 + p\delta_1$	In dimension $d = 2$: *qubits* The *Bloch ball* $$\rho = \frac{1}{2}(I_2 + x\sigma_X + y\sigma_Y + z\sigma_Z) \quad \text{with} \quad \\|(x, y, z)\\| \leq 1$$  Pure states have $\\|(x, y, z)\\| = 1$
The uniform distribution $$p = \left(\frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}\right)$$	The *maximally mixed state* $$\rho = \frac{I_d}{d} = \sum_{i=1}^{d} \|i\rangle\langle i\|$$

Bipartite and multipartite systems

Classical	Quantum
Joint system: *cartesian product* $\times$	Joint system: *tensor product* $\otimes$
	Tensor product of operators

Classical	Quantum
	$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mm}B \end{bmatrix}$
$\Sigma_{AB} = \Sigma_A \times \Sigma_B = \{(x, y) \,:\, x \in \Sigma_A \text{ and } y \in \Sigma_B\}$	$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B = \mathrm{span}\{\lvert x\rangle \otimes \lvert y\rangle\}$ where x and y index bases of $\mathcal{H}_A$ and $\mathcal{H}_B$
Bivariate probability distributions	Bipartite density matrices
$p_{AB}(x, y) \geq 0, \quad \sum_{x,y} p_{AB}(x, y) = 1$	$\rho_{AB} \geq 0, \quad \mathrm{Tr}\,\rho_{AB} = 1$
Product measures	Product states
$p_{AB} = p_A \times p_B$	$\rho_{AB} = \rho_A \otimes \rho_B$
Every state is a convex combination of product states	There exist *entangled states* = states that are not convex combinations of product states
$p_{AB} = \sum_{x,y} p_{AB}(x, y) \cdot \delta_x \times \delta_y$	*separable states* $\rho_{AB} = \sum_i t_i \rho_A^{(i)} \otimes \rho_B^{(i)} \in \mathsf{SEP}(A, B)$ Pure state example $\frac{1}{\sqrt{2}}(\lvert 00\rangle + \lvert 01\rangle) = \lvert 0\rangle \otimes \frac{1}{\sqrt{2}}(\lvert 0\rangle + \lvert 1\rangle)$ Mixed state example $\rho = \frac{I_A}{\dim \mathcal{H}_A} \otimes \frac{I_B}{\dim \mathcal{H}_B}$
	The *maximally entangled state*: $\mathcal{H}_A = \mathcal{H}_B$ and $\lvert\Omega\rangle = \frac{1}{\sqrt{\dim H_A}} \sum_i \lvert i\rangle_A \otimes \lvert i\rangle_B$ $\omega = \lvert\Omega\rangle\langle\Omega\rvert = \frac{1}{d} \sum_{i,j} \lvert ii\rangle\langle jj\rvert$ In dimension $\dim \mathcal{H}_A = 2$, $\omega = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$
Marginalisation: given p_{AB} $p_A(x) := \sum_y p_{AB}(x, y)$	*Partial trace*: given ρ_{AB} $\rho_A := [\mathrm{id} \otimes \mathrm{Tr}](\rho_{AB}) = \mathrm{Tr}_B(\rho_{AB})$ $\rho_B := [\mathrm{Tr} \otimes \mathrm{id}](\rho_{AB}) = \mathrm{Tr}_A(\rho_{AB})$

Classical	Quantum
$$p_B(y) := \sum_x p_{AB}(x,y)$$	
	Example: $$\mathrm{Tr}_A(\omega) = \mathrm{Tr}_B(\omega) = \frac{I}{d}$$

Measurements

Classical	Quantum
Outcome set: Σ_A	Outcome set: finite alphabet Σ Measure operators: $M_x \in \mathcal{B}(\mathcal{H}_A)$, for $x \in \Sigma$ Axioms for *POVM*s (Positive Operator Valued Measure): $$M_x \geq 0 \text{ and } \sum_x M_x = I_A$$
$$\mathbb{P}[x] = p_A(x) \quad \forall x \in \Sigma_A$$	$$\mathbb{P}[x] = \langle M_x, \rho_A \rangle = \mathrm{Tr}(M_x \rho_A)$$
	Basis measurements: fix $\{e_i\}$ a basis of $\mathcal{H}_A$ and set $$M_i := \|e_i\rangle\langle e_i\|$$ We have then $$\mathbb{P}[i] = \langle e_i \| \rho_A \| e_i \rangle$$
	Trivial measurements: for a probability vector p, set $$M_i := p_i I_A$$ We have then $$\mathbb{P}[i] = \mathrm{Tr}(p_i I_A \rho_A) = p_i,$$ independently of ρ_A
Total variation distance	*Trace distance* (aka nuclear norm or Schatten 1-norm)
$$\\|p - q\\|_{\mathrm{TV}} = \frac{1}{2} \sum_x \|p(x) - q(x)\|$$	$$\\|\rho - \sigma\\|_1 = \mathrm{Tr}\,\|\rho - \sigma\| = \sum_i s_i(\rho - \sigma)$$
Probability measure discrimination Given a random outcome sampled from either p or q, the optimal success probability for guessing whether it came from p or q is $$P^{\mathrm{opt}}(\mathrm{success}) = \frac{1}{2}(1 + \\|p - q\\|_{\mathrm{TV}})$$	*State discrimination* [Holevo-Helstrom] Given either ρ_0 or ρ_1 (with probability 1/2), the optimal success probability of guessing "0" or "1" is $$\frac{1}{2} + \frac{1}{4} \\|\rho_0 - \rho_1\\|_1$$

Channels

Classical	Quantum
Tranformations of classical states (discrete probability measures)	Transformations of quantum states (density matrices)
Stochastic matrices aka classical channels, Markov kernels, conditional probabilities $$N : \mathbb{R}^{\Sigma_A} \to \mathbb{R}^{\Sigma_B}$$ $$\forall a, b \quad N_{b\|a} := \langle \delta_b, N(\delta_a) \rangle \geq 0$$ $$\forall a \quad \sum_b N_{b\|a} = 1$$	*Quantum channels* $$\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$$ are *completely positive* $$\forall \mathcal{H}_R, \forall \rho_{AR} \quad [\Phi \otimes \mathrm{id}_R](\rho_{AR}) \geq 0$$ and *trace preserving* $$\forall \rho_A \quad \mathrm{Tr}\, \Phi(\rho_A) = \mathrm{Tr}\, \rho_A$$
	Classical channels embed as follows: $$\Phi_N(\rho) := \sum_{x,y} N_{y\|x} \langle x\|\rho\|x\rangle \|y\rangle\langle y\|$$
For classical maps, positivity $\iff$ complete positivity	Why complete positivity? There exists *positive maps* $$P(\mathrm{D}(\mathcal{H}_A)) \subseteq \mathrm{D}(\mathcal{H}_B)$$ that are not completely positive; e.g. the *transposition map* $X \mapsto X^\top$. Such maps might not preserve positivity when they act on subsystems: $$[\mathrm{transp} \otimes \mathrm{id}](\omega) \not\geq 0$$
$$N = \sum_{a,b} N_{b\|a} \|b\rangle\langle a\|$$ $$C_{\Phi_N} = \sum_{a,b} N_{b\|a} \|b\rangle\langle b\| \otimes \|a\rangle\langle a\|$$	*Choi matrix* For a linear map $\Phi : A \to B$, define $$C_\Phi := \sum_{i,j} \|i\rangle\langle j\| \otimes \Phi(\|i\rangle\langle j\|)$$ $$= [\mathrm{id} \otimes \Phi](\dim \mathcal{H}_A \cdot \omega_A) \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$$
Uniform random walk $$N_{b\|a} = \frac{1}{\|\Sigma_B\|}$$	*Depolarizing channel* $$\Delta(X) = \mathrm{Tr}(X) \frac{I_B}{\dim \mathcal{H}_B}$$
Permutation channel: for $\pi \in \mathfrak{S}(\Sigma_A)$ $$N^{(\pi)}_{b\|a} = \mathbf{1}_{b=\pi(a)}$$	*Unitary channel*: for $U \in \mathcal{U}(\mathcal{H}_A)$ $$\mathrm{Ad}_U(X) = UXU^*$$
	Measurements as channels A POVM M can be seen as a quantum $\to$ classical channel $$\Phi_M(\rho) = \sum_x \langle M_x, \rho \rangle \|x\rangle\langle x\|$$
$N : A \to B$ is a stoch. matrix $\iff$ $$N_{b\|a} \geq 0 \text{ and } \sum_b N_{b\|a} = 1$$	*Choi theorem* $\Phi : A \to B$ is a quantum channel $\iff$ $$C_\Phi \geq 0 \text{ and } \mathrm{Tr}_B C_\Phi = I_A$$

Classical	Quantum
$\dim(\text{stoch. matrices}) = \lvert\Sigma_A\rvert(\lvert\Sigma_B\rvert - 1)$	$\dim(\text{q. channels}) = (\dim \mathcal{H}_A)^2\big((\dim \mathcal{H}_B)^2 - 1\big)$
Classical channels Φ_N have Kraus operators $$K_{a,b} = \sqrt{N_{b\|a}}\lvert b\rangle\langle a\rvert$$	*Kraus theorem* $\Phi : A \to B$ is a quantum channel $\iff$ there exist operators $$K_1, \ldots, K_r : \mathcal{H}_A \to \mathcal{H}_B$$ such that $$\Phi(X) = \sum_{i=1}^{r} K_i X K_i^*$$ and $$\sum_{i=1}^{r} K_i^* K_i = I_A$$ The minimum r in such a decomposition is called the *Kraus rank* of the channel and it is equal to $\operatorname{rank} C_\Phi$
Dilation of classical channels to deterministic functions Given a channel $N : A \to B$ there exist a deterministic channel $D : AE \to B$ and a random variable Z on E such that $$N(a) = \mathbb{E}_{z\sim Z}\Big[D(a,z)\Big]$$	*Stinespring theorem* $\Phi : A \to B$ is a quantum channel $\iff$ there exist an isometry $V : A \to BE$ such that $$\Phi(X) = \operatorname{Tr}_E(VXV^*)$$ The dimension of the auxiliary space E can be taken to be the Kraus rank of Φ
Extreme points *Deterministic matrices* $$N^{(f)}_{b\|a} = \mathbf{1}_{b=f(a)}$$ for some function $f : \Sigma_A \to \Sigma_B$	Extreme points *Choi's condition*: A quantum channel Φ with Kraus operators $\{K_i\}$ is extreme iff the set $\{K_i^* K_j\}$ is linearly independent

Positivity notions for quantum channels

Channel	Choi matrix
Positive $$\forall X \geq 0 \implies \Phi(X) \geq 0$$	Block-positive $$\forall x \in \mathcal{H}_A, y \in \mathcal{H}_B \quad \langle x \otimes y \rvert C_\Phi \lvert x \otimes y\rangle \geq 0$$
Completely positive $$\forall \mathcal{H}_R \quad Z_{AR} \geq 0 \implies [\Phi \otimes \mathrm{id}_R](Z_{AR}) \geq 0$$	Positive semidefinite $$C_\Phi \geq 0 \iff \forall z \in \mathcal{H}_A \otimes \mathcal{H}_B \; \langle z \rvert C_\Phi \lvert z\rangle \geq 0$$
Entanglement-breaking $$\forall \mathcal{H}_R \, Z_{AR} \geq 0 \implies [\Phi \otimes \mathrm{id}_R](Z_{AR}) \in \mathsf{SEP}(B,R)$$	Separable $$C_\Phi \in \mathsf{SEP}(A,B)$$

Entropy

Classical	Quantum
Shannon entropy	*von Neumann entropy*
$$H(p) = -\sum_i p_i \log p_i$$	$$S(\rho) = S(A)_\rho = -\operatorname{Tr}(\rho \log \rho) = H(\lambda_\rho)$$

Classical	Quantum
$H(p) \in [0, \log\lvert\Sigma\rvert]$	$S(\rho) \in [0, \log\dim\mathcal{H}]$
$H(p) = 0 \iff p = \delta_i$	$S(\rho) = 0 \iff \rho = \lvert x\rangle\langle x\rvert$
$H(p) = \log\lvert\Sigma\rvert \iff p = \left(\frac{1}{\lvert\Sigma\rvert}, \dots, \frac{1}{\lvert\Sigma\rvert}\right)$	$S(\rho) = \log\dim\mathcal{H} \iff \rho = \frac{I_A}{\dim\mathcal{H}_A}$
Conditional entropy $$H(A\vert B)_p = H(AB)_p - H(B)_p$$	*Conditional entropy* $$S(A\vert B)_\rho = S(AB)_\rho - S(B)_\rho$$
Is always non-negative $$\forall p \qquad H(A\vert B)_p \geq 0$$	Can be negative $$S(A\vert B)_\omega = S(\omega) - S(I/2) = 0 - \log 2 = -1$$
Kullback–Leibler divergence $$D_{\mathrm{KL}}(p\vert\vert q) = \sum_i p_i \log\frac{p_i}{q_i}$$	*Quantum relative entropy* $$D(\rho\vert\vert\sigma) = \operatorname{Tr}\rho(\log\rho - \log\sigma)$$ if $\ker\sigma \subseteq \ker\rho$ and $+\infty$ otherwise
	It is *jointly convex* and non-negative $$\forall\rho,\sigma \quad D(\rho\vert\vert\sigma) \geq 0$$
	Data-processing inequality (DPI): for all quantum channels Φ, $$D(\rho\vert\vert\sigma) \geq D(\Phi(\rho)\vert\vert\Phi(\sigma))$$
	Mutual information $$I(A:B)_\rho = S(A)_\rho + S(B)_\rho - S(AB)_\rho$$ Is always non-negative (sub-additivity) $$\forall\rho_{AB} \qquad I(A:B)_\rho \geq 0$$ Equality $\iff \rho_{AB} = \rho_A \otimes \rho_B$
	Connection to relative entropy $$I(A:B)_\rho = D(\rho_{AB}\vert\vert\rho_A \otimes \rho_B)$$
strong sub-additivity follows from $$D_{\mathrm{KL}}(p_{ABC}\vert\vert q_{ABC}) \geq 0$$ where $$q_{ABC}(a,b,c) := \frac{p_{AB}(a,b)p_{BC}(b,c)}{p_B(b)}$$	DPI $\implies$ *strong sub-additivity*: $\forall\rho_{ABC}$ $$S(ABC)_\rho + S(B)_\rho \leq S(AB)_\rho + S(BC)_\rho$$ Equivalently in terms of mutual information $$I(A:BC)_\rho \geq I(A:B)_\rho$$

Source coding

Classical	Quantum
Discrete memoryless information source A sequence of i.i.d. random variables with common distribution $p \in \mathrm{P}(\Sigma)$	*Discrete memoryless source of quantum information*

Classical	Quantum
	A sequence of tensor powers $\rho^{\otimes n}$ of a quantum state $\rho \in \mathsf{D}(\mathcal{H})$
(n, m, δ) *compression scheme* A pair of functions $E : \Sigma^n \to \{0,1\}^m$ and $D : \{0,1\}^m \to \Sigma^n$ such that $$\mathbb{P}\Big[D \circ E(X) = X\Big] \geq 1 - \delta$$ where $X = (X_1, \ldots, X_n)$ come from a discrete memoryless source with distribution p	(n, m, δ) *compression scheme* A pair of quantum channels $$E : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}\big((\mathbb{C}^2)^{\otimes m}\big)$$ $$D : \mathcal{B}\big((\mathbb{C}^2)^{\otimes m}\big) \to \mathcal{B}(\mathcal{H}^{\otimes n})$$ such that (F is the channel fidelity) $$F(D \circ E, \rho^{\otimes n}) \geq 1 - \delta$$
Achievable rate A number $R > 0$ such that for all n there exists a (n, m_n, δ_n) compression scheme such that $$\lim_{n\to\infty} \frac{m_n}{n} = R \quad \text{and} \quad \lim_{n\to\infty} \delta_n = 0$$	*Achievable rate* Same as in the classical case
Shannon's source coding theorem Consider a discrete memoryless source with distribution p. Any rate $R > H(p)$ is achievable. For any sequence of (n, m_n, δ_n) compression schemes with rate $R = \lim m_n/n < H(p)$, we have $\lim \delta_n = 1$, i.e. the probability of successful decoding converges to 0.	*Schumacher's quantum source coding theorem* Consider a discrete quantum memoryless source associated to a quantum state $\rho \in \mathsf{D}(\mathcal{H})$. Any rate $R > H(p)$ is achievable. For any sequence of (n, m_n, δ_n) compression schemes with rate $R = \lim m_n/n < H(p)$, we have $\lim \delta_n = 1$, i.e. the probability of successful decoding converges to 0.
Proof idea: encode only *typical strings* A string $x \in \Sigma^n$ is ϵ-typical if $$2^{-n(H(p)+\epsilon)} < p(x_1) \cdots p(x_n) < 2^{-n(H(p)-\epsilon)}$$	Proof idea: consider typical strings with respect to the eigenvalues of ρ. Encode only states supported on the *typical subspace*

Channel coding

Classical	Quantum
(n, m, δ) *coding scheme* for N A pair of functions $E : \{0,1\}^m \to \Sigma_A^n$ and $D : \Sigma_B^n \to \{0,1\}^m$ such that $$\forall i \in \{0,1\}^m \quad \mathbb{P}\Big[D \circ N^{\times n} \circ E(i) = i\Big] \geq 1 - \delta$$	(n, m, δ) *coding scheme* for Φ An encoding map $$E : \{0,1\}^m \to \mathcal{B}(\mathcal{H}_A^{\otimes n})$$ and a measurement map $$D : \mathcal{B}(\mathcal{H}_B^{\otimes n}) \to \{0,1\}^m$$ such that $$\forall i \in \{0,1\}^m \quad \langle D_i, \Phi^{\otimes n}\big(E(i)\big)\rangle \geq 1 - \delta$$
Achievable rate A number $R > 0$ such that for all n there exists a (n, m_n, δ_n) compression scheme such that $$\lim_{n\to\infty} \frac{m_n}{n} = R \quad \text{and} \quad \lim_{n\to\infty} \delta_n = 0$$	*Achievable rate* Same as in the classical case
	Holevo information Given an ensemble of quantum states

Classical	Quantum
	$\mathcal{E} = \{p(x), \rho_x\}_{x\in\Sigma}$ with states $\rho_x \in \mathrm{D}(\mathcal{H})$ define $$\chi(\mathcal{E}) := S\big(\sum_{x\in\Sigma} p(x)\rho_x\big) - \sum_{x\in\Sigma} p(x) S(\rho_x)$$
	Holevo capacity For a channel $\Phi : A \to B$ define $$\chi(\Phi) = \sup_{\{p(x),\rho_x\}_{x\in\Sigma}} \chi\Big(\{p(x), \Phi(\rho_x)\}_{x\in\Sigma}\Big)$$ where the supremum is over all ensembles $\{p(x), \rho_x\}_{x\in\Sigma}$ with states $\rho_x \in \mathrm{D}(\mathcal{H}_A)$ and all finite alphabets Σ
Classical capacity $C(N)$ = the maximum achievable rate for for transmitting information using N	*Classical capacity of a quantum channel* $C(\Phi)$ = the maximum achievable rate for for transmitting classical information using the quantum channel Φ
Shannon's channel coding theorem $$C(N) = \sup_{p_A \in \mathrm{P}(\Sigma_A)} I(A:B)_{p^N_{AB}}$$ where $$p^N_{AB}(a,b) = p_A(a)N(b\|a) \in \mathrm{P}(\Sigma_A \times \Sigma_B)$$	*Holevo-Schumacher-Westmoreland theorem* $$C(\Phi) = \lim_{n\to\infty} \frac{1}{n}\chi(\Phi^{\otimes n})$$