4. Entanglement conversion & quantification

a) Introduction and setup

Entanglement = those properties of a system which cannot be changed using Local Operations and Classical Communication (LOCC).

(Note: This is typically the definition of entanglement.)

Question: When – and how – can we convert entangled states into another by LOCC?

Relevance:

- Different protocols might require different – "cheaper" or "more expensive" – entangled states.
- We can easily produce some entangled state but need a different one.
- Can be used to quantify entanglement via conversion rate, e.g. to some "gold standard"

– for instance, how many "e-bits" $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ are contained in a state? (And is there a meaning to "$|\chi\rangle$ contains 0.7 e-bits?")

Let us first consider pure states!

Known from Chapter II:

Same Schmidt coefficients $\Longleftrightarrow$ States can be converted by local unitaries (subclass of LOCC!)

Question: What if Schmidt coefficients are different?

Example:
$$|\chi\rangle = \sqrt{\tfrac{2}{3}}\,|00\rangle + \sqrt{\tfrac{1}{3}}\,|11\rangle$$
$$|\phi^+\rangle = \sqrt{\tfrac{1}{2}}\,|00\rangle + \sqrt{\tfrac{1}{2}}\,|11\rangle$$

Question 1: Can we do $|\phi^+\rangle \xrightarrow{\text{LOCC}} |\chi\rangle$?

convert using LOCC.

A does POVM $\{M_0, M_1\}$,

$$M_0 = \begin{pmatrix} \sqrt{2/3} & \\ & \sqrt{1/3} \end{pmatrix}; \quad M_1 = \begin{pmatrix} \sqrt{1/3} & \\ & \sqrt{2/3} \end{pmatrix}.$$

Post-measurement states $|\tilde{\psi}_k\rangle = (M_k \otimes I)|\phi^+\rangle$:

$$|\tilde{\psi}_0\rangle = \frac{1}{\sqrt{2}}\left(\sqrt{\frac{2}{3}}|00\rangle + \sqrt{\frac{1}{3}}|11\rangle\right)$$

$$|\tilde{\psi}_1\rangle = \frac{1}{\sqrt{2}}\left(\sqrt{\frac{1}{3}}|00\rangle + \sqrt{\frac{2}{3}}|11\rangle\right)$$

$\Rightarrow$
- $p_0 = 1/2$: $|\psi_0\rangle = \sqrt{\frac{2}{3}}|00\rangle + \sqrt{\frac{1}{3}}|11\rangle = |\chi\rangle$ ✓ SUCCESS!
- $p_1 = 1/2$: $|\psi_1\rangle = \sqrt{\frac{1}{3}}|00\rangle + \sqrt{\frac{2}{3}}|11\rangle$

$\rightarrow$ wrong state, but right Schmidt coefficients $\rightarrow$ can be fixed by local unitaries! (specifically, A & B need to apply $X \otimes X$).

Protocol for conversion $|\phi^+\rangle \xrightarrow{\text{LOCC}} |\chi\rangle$:

① A does POVM $\{\Pi_0, \Pi_1\}$

② A sends her outcome to B

③ If result is 1, both apply X.

$\Longrightarrow$ Success probability $p = p_0 + p_1 = 1$!

Note: This is the best possible, since POVM cannot increase Schmidt rank $\longrightarrow$ we cannot get more than one copy of $|\phi^+\rangle$ with any prob.!

Question 2: Can we do $|\chi\rangle \xrightarrow{\text{LOCC}} |\phi^+\rangle$?

A does POVM $\{\Pi_0, \Pi_1\}$,

$$\Pi_0 = \begin{pmatrix} \sqrt{1/2} & \\ & 1 \end{pmatrix}, \quad \Pi_1 = \begin{pmatrix} \sqrt{1/2} & \\ & 0 \end{pmatrix}$$

$\longrightarrow$ $|\tilde{\psi}_0\rangle = \sqrt{1/3}\,|00\rangle + \sqrt{1/3}\,|11\rangle$

$|\tilde{\psi}_1\rangle = \sqrt{1/3}\,|00\rangle$.

Outcome:

- $p_0 = \frac{2}{3}$: $|\psi_0\rangle = |\phi^+\rangle$ ✓ success
- $p_1 = \frac{1}{3}$: $|\psi_1\rangle = |00\rangle$ ✗ FAILURE

(no ent. left → cannot salvage state!)

Protocol for $|\chi\rangle \xrightarrow{LOCC} |\phi^+\rangle$

1. A does POVM $\{\Pi_0, \Pi_1\}$.
2. A sends result to B.
3. If result is 0, they have obtained $|\phi^+\rangle$. Otherwise, they declare failure.

$\Longrightarrow$ Success probability $\underline{p = p_0 = \frac{2}{3}}$.

Note: Will see: this is indeed the best conversion rate possible.

Big drawback: Conversion is not reversible — we cannot use it to attach one number to an entangled state (i.e. quantify entanglement).

b) The most general LOCC protocol

What is the most general LOCC protocol?

time

A

B

POVM $\{A_i\}$

send result i

POVM $\{B_j^i\}$ dep. on i

send j

POVM $\{C_k^{ij}\}$
dep. on i, j

k

$\{D_e^{ijk}\}$

⋮

… etc …

this can go any number of rounds!

Total map implemented:

$$\rho \mapsto \sum \left(\ldots \cdot C_k^{ij} \cdot A_i \right) \otimes \left(\ldots \cdot D_e^{ijk} B_j^i \right) \rho \, (\,)^\dagger \otimes (\,)^\dagger$$

Very complicated structure!

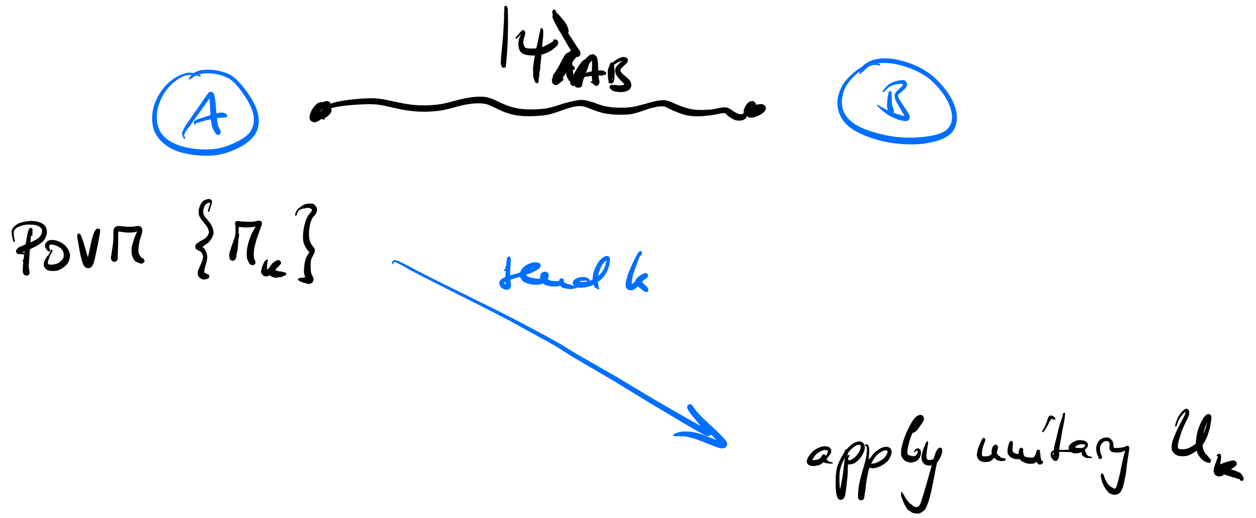
(Limit/closure over $N \to \infty$ rounds, ...)

Can we simplify this?

For general mixed states ρ_{AB} — no!

But: For LOCC protocols acting on pure states $|\psi\rangle_{AB}$, a significant simplification is possible:

Theorem: For a pure state $|\psi\rangle$, any LOCC protocol above can be replaced by the following protocol:



$$|\psi\rangle \longmapsto \Pi_k \otimes U_k \, |\psi\rangle,$$

where $\{\Pi_k\}$ is a POVM and the U_k are unitary.

That is: Alice does a POVM $\{\Pi_k\}$, sends the result to Bob, and Bob applies a unitary U_k.

This only requires one round, and only one-way classical communication!

Proof: $\rightarrow$ Homework.

(Idea: Any state $|\chi\rangle$ can be written as

$$|\chi\rangle = (M\otimes I)|\phi^+\rangle = (I\otimes N)|\phi^+\rangle$$

$\rightarrow$ A can use this to "simulate" any meas. of B through a dif. meas. on her side – if state is known!)

c) Single-copy conversion & majorization

Can we characterize the optimal single-copy conversion protocols for pure states?

Have seen: most general protocol:

$$|\psi\rangle \longmapsto |\tilde{\psi}_k\rangle = (M_k \otimes U_k)|\psi\rangle,$$

$$p_k = \| |\tilde{\psi}_k\rangle \|^2$$

with $\{M_k\}$ POVM, U_k unitary.

Converts $|\psi\rangle \longmapsto \{p_k, |\psi_k\rangle = \frac{1}{\sqrt{p_k}}|\tilde{\psi}_k\rangle\}$,

i.e. state $|\psi_k\rangle$ w/ prob. p_k.

To characterize conversion power of LOCC:

- Only <u>Schmidt coefficients</u> of initial state $|\psi\rangle$ and final states $|\psi_k\rangle$ relevant – anything else can be changed at any time by local unitaries.

- Schmidt coefficients = eigenvalues of reduced density matrices

$$\rho_A = \mathrm{tr}_B\, |\psi\rangle\langle\psi|$$

$$\rho_{A,k} = \mathrm{tr}_B\, |\psi_k\rangle\langle\psi_k|$$

- LOCC protocol acts instead as

$$\rho_A \longmapsto \{p_k, \rho_{A,k}\}$$

$$\text{with} \quad p_k \rho_{A,k} = \Pi_k \rho_A \Pi_k^+ , \quad \sum \Pi_k^+ \Pi_k = I,$$

Thus: To characterize the power of LOCC protocols, we can – and will – equivalently study the following question:

Question: Given ρ_A and $\{p_k, \rho_{A,k}\}$, under which conditions does there exist a POVM Π_k, $\sum \Pi_k^+ \Pi_k = I$, such that $p_k \rho_{A,k} = \Pi_k \rho_A \Pi_k^+$, or more generally $\{\Pi_{k,i_k}\}$, $\sum \Pi_{k,i_k}^{\dagger} \Pi_{k,i_k} = I$,

such that $\quad q_k \rho_{A,k} = \sum_{i_k} \Pi_{k,i_k} \rho_A \Pi_{k,i_k}^{\dagger}$?

This allows for grouping of outcomes, such as e.g. for the example $|\phi^+\rangle \longrightarrow |\chi\rangle$

Definition: For $\lambda \in \mathbb{R}^d_{\geq 0}$, let $\lambda^{\downarrow} = (\lambda_1^{\downarrow}, \dots, \lambda_d^{\downarrow})$, $\lambda_1^{\downarrow} \geq \lambda_2^{\downarrow} \geq \dots \geq 0$ denote the **ordered version** of λ.

Definition (Majorization): We say that **λ is majorized by μ**, or **μ majorizes λ**, denoted as

$$\lambda \prec \mu,$$

iff $\displaystyle\sum_{i=1}^{k} \lambda_i^{\downarrow} \leq \sum_{i=1}^{k} \mu_i^{\downarrow} \quad \forall\, k = 1, \dots, d,$

with equality for $k = d$.

<u>Theorem:</u> The following are equivalent

(i) $\lambda \prec \mu$

(ii) there exist permutations P_i and probabilities q_i, $\sum q_i = 1$, such that

$$\lambda = \sum q_i P_i \mu$$

(iii) there exists a doubly stochastic matrix Q

(i.e.: $Q_{ij} \geq 0$, $\sum_i Q_{ij} = 1 \ \forall j$, $\sum_j Q_{ij} = 1 \ \forall i$,

that is, Q describes a random process with the fully random distribution $\left(\frac{1}{d}, \dots, \frac{1}{d}\right)$ as a fixed point)

such that $\lambda = Q\mu$.

Intuitively, this is saying that when regarded as prob. distributions, <u>λ is "more random" than μ</u>, in the sense that it can be obtained from μ by adding in randomness.

Proof:

(ii) $\Rightarrow$ (iii): Let $Q = \sum q_i P_i$. ☒

(iii) $\Rightarrow$ (ii): This is known as Birkhoff's Theorem:

Every doubly stochastic Q is of the form $Q = \sum q_i P_i$. ☒

(ii) $\Rightarrow$ (i): First, note that for any λ,

$$\underbrace{\sum_{i=1}^{k} \lambda_i^{\downarrow}}_{\text{sum of } k \text{ largest terms}} \geq \underbrace{\sum_{i=1}^{k} \lambda_{\pi(i)}}_{\text{sum of any } k \text{ terms}} \quad \text{for any perm. } \pi.$$

Now let $\lambda := \sum q_s P_s \mu$. Then,

$$\sum_{i=1}^{k} \lambda_i^{\downarrow} = \sum_{i=1}^{k} \Big[\sum_s q_s \mu_{P_s(i)} \Big]^{\downarrow}$$

ordering = some perm. π

$$= \sum_s q_s \underbrace{\sum_{i=1}^{k} \mu_{\pi(P_s(i))}}_{\leq \sum_{i=1}^{k} \mu_i^{\downarrow}} \leq \sum_{i=1}^{k} \mu_i^{\downarrow}$$

☒

<u>(i) $\Rightarrow$ (ii):</u> Homework.

<u>Idea:</u> Find a permutation P s.th.

$$\lambda_1^{\downarrow} = \Big[\overbrace{\big(qI + (1-q)P\big)}^{Q}\mu^{\downarrow}\Big]_1 ,$$

then consider $(\lambda_2^{\downarrow}, \dots, \lambda_d^{\downarrow}) \prec \big((Q\mu^{\downarrow})_2, \dots, (Q\mu^{\downarrow})_d\big)$,

& proceed by induction.

$\blacksquare$

<u>Remarks:</u>

- Majorization defines a <u>partial order</u> on the space of prob. distributions.
- $\lambda \prec \mu$: λ "more disordered" than μ – in part., it has larger entropy! (Made rigorous by notion of "Schur concavity/convexity": for a concave/convex $f(x)$,

  $F(\lambda) = \sum f(\lambda_i)$ fulfills

  $$\lambda \prec \mu \;\Rightarrow\; F(\lambda) \underset{(\leq)}{\geq} F(\mu)$$

  concave   convex

Majorization can be generalized to positive operators:

Definition: For $A, B \geq 0$, we define

$$A \prec B \;:\Longleftrightarrow\; \lambda^{\downarrow}(A) \prec \lambda^{\downarrow}(B),$$

with $\lambda^{\downarrow}(X)$ the ordered eigenvalues of X.

Theorem (Ky-Fan maximum principle):

For A hermitian,

$$\sum_{j=1}^{k} \lambda_j^{\downarrow}(A) = \max_{P} \operatorname{tr}(AP),$$

where the maximum is over all projectors P with rank k.

Proof:

"$\leq$": Let $A = \sum_{j=1}^{d} \lambda_j^{\downarrow}(A) |q_j\rangle\langle q_j|$. Choose $\tilde{P} = \sum_{j=1}^{k} |q_j\rangle\langle q_j|$.

Then, $$\sum_{j=1}^{k} \lambda_j^{\downarrow}(A) = \operatorname{tr}(A\tilde{P}) \leq \max_{P} \operatorname{tr}(AP).$$

"$\geq$": Write $P = \sum_{i=1}^{k} |p_i\rangle\langle p_i|$, with $\{|p_i\rangle\}_{i=1}^{d}$ ONB.

Then, $\langle p_i | A | p_i \rangle = \sum \underbrace{|\langle p_i | a_j \rangle|^2}_{=: u_{ij}} \lambda_j^\downarrow(A)$

Since $\sum_j u_{ij} \overline{u_{kj}} = \sum_j \langle p_i | q_j \rangle \langle q_j | p_k \rangle = \langle p_i | p_k \rangle = \delta_{ik}$

$\Rightarrow (u_{ij})$ unitary $\Rightarrow \sum_i |u_{ij}|^2 = \sum_j |u_{ij}|^2 = 1$

$\Leftrightarrow (|u_{ij}|^2)$ doubly stochastic.

$\Rightarrow (\langle p_i | A | p_i \rangle)_i \prec \lambda^\downarrow(A)$

$\Leftrightarrow \operatorname{tr}(AP) = \sum_{i=1}^k \langle p_i | A | p_i \rangle \leq \sum_{i=1}^k \lambda_i^\downarrow(A)$

$\boxed{}$

Corollary: $\lambda^\downarrow(A+B) \prec \lambda^\downarrow(A) + \lambda^\downarrow(B)$.

Proof: $\sum_{i=1}^k \lambda^\downarrow(A+B) = \max_{P:\, \operatorname{rank} P = k} \operatorname{tr}(P(A+B))$

$\leq \max_P \operatorname{tr}(PA) + \max_P \operatorname{tr}(PB)$

$= \sum_{i=1}^k \lambda_i^\downarrow(A) + \sum_{i=1}^k \lambda_i^\downarrow(B)$.

$\boxed{}$

Theorem (single-copy entanglement conversion):

We can convert

$$|\psi\rangle \xrightarrow{\text{LOCC}} \{p_k, |\psi_k\rangle\}$$

by LOCC if and only if

$$\lambda^{\downarrow}(\rho) \prec \sum_{k=1}^{K} p_k \lambda^{\downarrow}(\rho_k),$$

where $\rho = \mathrm{tr}_A |\psi\rangle\langle\psi|$, $\rho_k = \mathrm{tr}_A |\psi_k\rangle\langle\psi_k|$.

Eigenvalues of $\mathrm{tr}_A |\psi\rangle\langle\psi|$ and $\mathrm{tr}_B |\psi\rangle\langle\psi|$ are the same. But since Alice does POVM, tracing her will facilitate the proof.

Proof:

"$\Rightarrow$": A does some POVM $\{\Pi_k\}$. Then,

$$\sum_{k=1}^{K} p_k \lambda^{\downarrow}(\rho_k) = \sum_{k=1}^{K} \lambda^{\downarrow}(p_k \rho_k)$$

$$= \sum_{k=1}^{K} \lambda^{\downarrow}\left(\mathrm{tr}_A\left[(\Pi_k \otimes I)|\psi\rangle\langle\psi|(\Pi_k^{\dagger} \otimes I)\right]\right)$$

... use cyclicity of tr_A in A ...

$$\overset{\text{Corollary}}{\succ} \lambda^{\downarrow}\left(\mathrm{tr}_A\left[\sum_{k=1}^{k}(\Pi_k^{+}\Pi_k\otimes I)\,|\psi\rangle\langle\psi|\right]\right)$$

$$= \lambda^{\downarrow}(\rho).$$

Note: This also works if Alice loses outcomes – i.e. performs $\{\Pi_{k,i_k}\}$, where for a given k, all i_k, together with Bob's U_{k,i_k}, give the same $|\psi_k\rangle$ – since $\lambda^{\downarrow}(\rho_{k,i_k})$ is the same for all i_k (fixed k).

"$\Leftarrow$": $\lambda^{\downarrow}(\rho) \prec \sum_k p_k \lambda^{\downarrow}(\rho_k)$

$$\Longrightarrow \exists P_j, q_j \text{ s.th. } \lambda^{\downarrow}(\rho) = \sum_{k,j} q_j P_j p_k \lambda^{\downarrow}(\rho_k)$$

W.l.o.g., we can assume ρ, ρ_i are diagonal (i.e. $\rho = \mathrm{diag}(\lambda^{\downarrow}(\rho))$, etc.), since only the eigenvalues matter. – The corresponding POVMs $\{\Pi_k\}$ are related by unitaries, $\Pi_k \rightsquigarrow V_k \Pi_k W$.

Furthermore, if $\lambda^{\downarrow}(\rho)$ has zeros (i.e. zero eigenvalues), $\sum_k p_k \lambda^{\downarrow}(\rho_k)$ and thus each $\lambda^{\downarrow}(\rho_k)$ must have at least the same number of zeros (since $\lambda^{\downarrow}(\rho) < \sum p_k \lambda^{\downarrow}(\rho_k)$).

We can then discard those zeros (i.e. POVM acts trivially on those).

Thus w.l.o.g.: ρ has full rank (invertible!).

Define POVM $\{M_{kj}\}$ via $M_{kj}\sqrt{\rho} = \sqrt{p_k q_j}\sqrt{\rho_k}\, P_j^{+}$.

Then, $\sqrt{\rho}\left(\sum_{kj} M_{kj}^{+} M_{kj}\right)\sqrt{\rho} = \sum_{kj} p_k q_j P_j \rho_k P_j^{+} = \rho$

$P_j \rho_k = \mathrm{diag}(\lambda^{\downarrow}(\rho,\rho_k))$

ρ full rk.
$\Longrightarrow \sum_{kj} M_{kj}^{+} M_{kj} = I$.

Further, $M_{kj}\rho M_{kj}^{+} = p_k q_j \rho_k$

$$\Longrightarrow \sum_j M_{kj}\rho M_{kj}^{+} = p_k \rho_k$$

$\Longrightarrow$ by binning all j-outcomes for fixed k, we obtain POVM for $\rho \longmapsto \{p_k, \rho_k\}$, i.e. $|\psi\rangle \mapsto \{p_k, |\psi_k\rangle\}$.

Example:

Optimal rate for $\left(\frac{1}{2}, \frac{1}{2}\right) \longleftrightarrow \left(\frac{2}{3}, \frac{1}{3}\right)$

vectors with Schmidt coeffs.

- $\left(\frac{1}{2}, \frac{1}{2}\right) \prec \left(\frac{2}{3}, \frac{1}{3}\right)$: prob $= 1$ ✓

- $\left(\frac{2}{3}, \frac{1}{3}\right) \prec \underbrace{\frac{2}{3}}_{\text{max. possible value.}} \left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{3}(1,0) = \left(\frac{2}{3}, \frac{1}{3}\right)$

  prob $= 2/3$ ✓

- we also see that Schmidt rank cannot be increased, e.g.: Can we do

$$|\phi^+\rangle \longrightarrow \left(\sqrt{\tfrac{2}{3}}|00\rangle + \sqrt{\tfrac{1}{3}}|11\rangle\right)^{\otimes 2} \; ?$$

$$\left(\tfrac{1}{2}, \tfrac{1}{2}, 0, 0\right) \not\prec p\left(\tfrac{4}{9}, \tfrac{2}{9}, \tfrac{2}{9}, \tfrac{1}{9}\right) + (1-p)\cdot \text{Stuff}$$

impossible!

d) Asymptotic protocols

Single-copy entanglement conversion:
<u>not reversible!</u>

E.g.: $|\phi^+\rangle \rightleftarrows |\chi\rangle = \sqrt{\frac{2}{3}}\,|00\rangle + \sqrt{\frac{1}{3}}\,|11\rangle$

need 1 ebit for "$\longrightarrow$",
get 2/3 ebit from "$\longleftarrow$".

→ Entanglement is <u>lost</u>

→ need at least two numbers to quantify entanglement:

ebits needed to <u>build</u> state

ebits <u>extractable</u> from state

Can we do better if we work with <u>more copies</u>?

$$|\chi\rangle^{\otimes 2} \longrightarrow p_2\, |\phi^+\rangle^{\otimes 2} + p_1\, |\phi^+\rangle^{\otimes 1} \ ?$$

$$|\chi\rangle^{\otimes 3} \longrightarrow p_3\, |\phi^+\rangle^{\otimes 3} + p_2\, |\phi^+\rangle^{\otimes 2} + \dots \ ?$$

Average yield of max. ent. states $|\phi^+\rangle$

= "# of ebits per copy of $|\chi\rangle$":

$$\bar{p} = \frac{p_1 + 2p_2 + 3p_3 + \dots}{k} \leftarrow \text{\# of copies.}$$

E.g. for $|\chi\rangle^{\otimes 3}$:

$$\left(\frac{8}{27}, \frac{4}{27}, \frac{4}{27}, \frac{4}{27}, \frac{2}{27}, \frac{2}{27}, \frac{2}{27}, \frac{1}{27}\right) \prec$$

$$p_3\left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right) +$$

$$p_2\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0\right) +$$

$$p_1\left(\frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, 0\right) +$$

$$p_0\left(1, 0, 0, 0, 0, 0, 0, 0\right)$$

$$p_0 = 1-(p_1+p_2+p_3)$$

$p_3 = 8/27, \ p_2 = 16/27, \ p_1 = 0, \ p_0 = 3/27,$

gives a valid solution.

$$\bar{p} = \frac{3 \cdot \frac{8}{27} + 2 \cdot \frac{16}{27}}{3} = \frac{56}{81} > 2/3 !$$

$\longrightarrow$ Improvement over single-copy protocols!

How good can we get by using $N \to \infty$ copies?

Requirements for asymptotic protocols

- convert $|\phi^+\rangle^{\otimes M} \rightleftarrows |\chi\rangle^{\otimes N}$ with a rate $R = \lim \frac{M}{N} > 0$ as $N, M \to \infty$

- Need not be deterministic:
  Success probability $p \to 1$ as $N, M \to \infty$

- Need not be perfect:

  Require that distance δ from correct state goes to zero, $\delta \to 0$, as $N, n \to \infty$.

(Note: We can afford these imperfections since asymptotically, they vanish: Not meaningful in finite-copy scenario.)

How can we measure error δ?

Definition (Fidelity): For two states $|\psi\rangle$ and $|\phi\rangle$,

$F = |\langle\psi|\phi\rangle|^2$ is called the fidelity of $|\psi\rangle$ and $|\phi\rangle$.

Lemma: $\delta := 1 - F$ bounds the error on any expectation value,

$$|\langle\psi|O|\psi\rangle - \langle\phi|O|\phi\rangle| \leq 2\sqrt{\delta}\, \|O\|_\infty,$$

with $\|O\|_\infty = \sup \frac{\|O|\psi\rangle\|}{\||\psi\rangle\|}$ the operator norm.

That is: δ bounds how well we can (or cannot) distinguish $|\phi\rangle$ and $|\psi\rangle$ by <u>any physical test</u>!

<u>Proof</u>: $\rightarrow$ Homework.

We use $\delta = 1 - F$ to measure the error, and we require a good <u>global</u> distance, i.e.

$$|\phi^+\rangle^{\otimes M} \longrightarrow |\Theta_N\rangle \underset{\uparrow}{\approx} |\chi\rangle^{\otimes N}$$

$$\left|\langle\Theta_N|\left(|\chi\rangle^{\otimes N}\right)\right|^2 \longrightarrow 1$$

(and vice versa).

Let us now consider some state

$$|\chi\rangle = \sum_{x=1}^{d} \sqrt{p(x)}\ |x\rangle_A |x\rangle_B .$$

Then, $$|\chi\rangle^{\otimes N} = \sum_{x_1,\dots,x_N} \sqrt{p(x_1)\cdot\ldots\cdot p(x_N)}\ |x_1,\dots,x_N\rangle_A |x_1,\dots,x_N\rangle_B ,$$

i.e., the probabilities $p(x_1, \dots, x_N)$ associated to a configuration $x_1, \dots, x_N$ describe independently & identically distributed (iid) random variables with prob. $p(x_i)$.

Theorem (Law of large numbers (LLN))

Let X be a random variable with prob. $p(x)$. Then,

$\forall \varepsilon > 0 \ \forall \delta > 0 \ \exists N_0 \ \forall N \geq N_0$

$$\mathrm{prob}\left(\left| \frac{1}{N} \sum_{i=1}^{N} x_i - E(x) \right| \geq \varepsilon \right) \leq \delta$$

where $E(x) = \sum_x p(x) x$.

(Proof → probability theory.)

How are the outcomes of an i.i.d. source distributed?

What is a <u>typical outcome</u> of an i.i.d. source?

Example: $x = 0, 1$; $p_0 = p$; $p_1 = 1-p$.

k times outcome 0, $N-k$ times outcome 1:

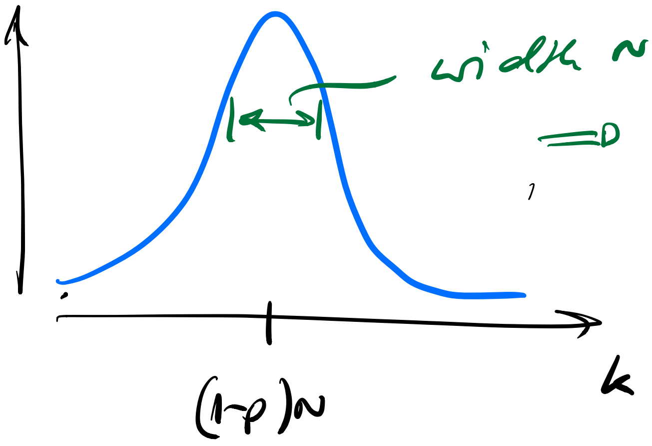
$$\text{Prob.} = p^k (1-p)^{N-k}$$

$$\#\ \text{possibilities}: \binom{N}{k}$$

$\Longrightarrow$ Binomial distribution:

total prob. for $k \times 0$, $(N-k) \times 1$:

$$p^k (1-p)^{N-k} \binom{N}{k}$$



width $\sim \sqrt{N}$

$\Longrightarrow$ rel. width $\sim \frac{1}{\sqrt{N}}$

– i.e., for large N, we expect that $\frac{k}{N} \approx p$ with high probability $\to 1$ as $N \to \infty$.

General case:

"Typical" output: Expect output x $N \cdot p(x)$ times.

$$\Rightarrow p(x_1,\dots,x_N) = p(x_1)\cdot\ldots\cdot p(x_N) \approx p(1)^{Np(1)}\cdot\ldots\cdot p(d)^{Np(d)}$$

$$\Rightarrow -\log p(x_1,\dots,x_N) \approx N\cdot\underbrace{\left(-\sum_{x=1}^{d} p(x)\log p(x)\right)}_{=:\,H(p)}$$

$=: H(p)$: Shannon entropy of p.

our logs are base 2!

$\Rightarrow$ we typically expect to see $x_1,\dots,x_N$ with

$$p(x_1,\dots,x_N) \approx 2^{-NH(p)},$$

and there are $\approx 2^{NH(p)}$ such typical sequences.

More formally, this is defined as follows:

Definition (ε-typical sequences):

For an iid-variable X, we say that $x_1, \dots, x_N$ is an **ε-typical sequence** if

$$2^{-N(H(p)+\varepsilon)} \leq p(x_1, \dots, x_N) \leq 2^{-N(H(p)-\varepsilon)}$$

We denote the set of ε-typical sequences by $T(N,\varepsilon)$.

Theorem:

$\forall \varepsilon > 0 \ \forall \delta > 0 \ \exists N_0 \ \forall N \geq N_0$ such that

1) a random sequence $x_1, \dots, x_N$ of length N is ε-typical w/ prob. $\geq 1-\delta$.

2) $(1-\delta)\, 2^{N(H(p)-\varepsilon)} \leq |T(N,\varepsilon)| \leq 2^{N(H(p)+\varepsilon)}$

<u>Proof:</u> 1) $-\log p(x_i)$ is iid variable (w/ prob. $p(x_i)$)

$$\xRightarrow{LLN} \forall \varepsilon, \delta\ \exists N_0\ \forall N \geq N_0 :$$

$$\text{prob}\Big(\Big| \underbrace{\frac{1}{N} \sum_{i=1}^{N} -\log p(x_i)}_{= -\log p(x_1, \dots, x_N)} - \underbrace{E\big(-\log p(x)\big)}_{\substack{= \sum_x p(x)(-\log p(x)) \\ = H(p)}} \Big| \geq \varepsilon \Big) \leq \delta$$

$$\Rightarrow \text{prob}\Big(\Big| -\frac{1}{N} \log p(x_1, \dots, x_N) - H(p) \Big| \geq \varepsilon \Big) \leq \delta$$

$\Rightarrow$ w. prob. $\geq 1-\delta$,

$$-N(H(p)+\varepsilon) \leq \log p(x_1, \dots, x_N) \leq -N(H(p)-\varepsilon) \quad \blacksquare$$

2)
$$1 \geq \sum_{x_1, \dots, x_N \in T(N,\varepsilon)} p(x_1, \dots, x_N)$$

$$\overset{\text{Def. } T(N,\varepsilon)}{\geq} \sum_{T(N,\varepsilon)} 2^{-N(H(p)+\varepsilon)}$$

$$= |T(N,\varepsilon)| \cdot 2^{-N(H(p)+\varepsilon)}$$

$$1-\delta \leq \sum_{T(N,\varepsilon)} p(x_1,\dots,x_N) \leq |T(N,\varepsilon)| \cdot 2^{-N(H(p)-\varepsilon)}$$

$\blacksquare$

<u>In words:</u>

ε-typical sequence $:\Leftrightarrow$ $\dfrac{\log p(x_1,\dots,x_N)}{N}$ ε-close to $H(p)$

Asymptotically, a sequence is <u>ε-typical w/ prob $\to 1$</u>,

and there are <u>$\sim 2^{NH(p)}$ ε-typical sequences.</u>

<u>Note:</u> Typical sequences are an important concept in <u>classical</u> information theory ($\to$ data compression etc.!)

<u>Application of typicality to entanglement conversion:</u>

$$|\chi\rangle = \sum_x \sqrt{p(x)}\, |x\rangle_A |x\rangle_B$$

$$\Rightarrow |\chi\rangle^{\otimes N} = \sum_{x_1,\dots,x_N} \sqrt{p(x_1)\cdot\dots\cdot p(x_N)}\; |x_1,\dots,x_N\rangle_A |x_1\dots x_N\rangle_B$$

Fix some $\varepsilon > 0$ and $\delta > 0$, and a matching N_0. Let

$$|\tilde{\vartheta}_N\rangle := \sum_{x_1,\dots,x_N \in T(N,\varepsilon)} \sqrt{p(x_1)\cdot\dots\cdot p(x_N)}\, |x_1,\dots,x_N\rangle |x_1,\dots,x_N\rangle,$$

$$|\vartheta_N\rangle := \frac{|\tilde{\vartheta}_N\rangle}{\sqrt{\langle\tilde{\vartheta}_N|\tilde{\vartheta}_N\rangle}}$$

We have (for $N \geq N_0$):

$$\langle\vartheta_N|\left(|\chi\rangle^{\otimes N}\right) = \frac{\sum_{T(N,\varepsilon)} p(x_1,\dots,x_N)}{\underbrace{\sqrt{\sum_{T(N,\varepsilon)} p(x_1,\dots,x_N)}}_{\geq 1-\delta \text{ for } N \geq N_0}} \geq \sqrt{1-\delta} \geq 1-\delta$$

and $|T(N,\varepsilon)| \leq 2^{N(H(p)+\varepsilon)}$.

$\Rightarrow$ That is: Instead of converting $|\phi^+\rangle^{\otimes n} \rightleftarrows |\chi\rangle^{\otimes N}$,

we can instead convert $|\phi^+\rangle^{\otimes n} \rightleftarrows |\vartheta_N\rangle$,

as the error δ can be taken to zero.

Protocol for $|\phi^+\rangle^{\otimes M} \longrightarrow |\vartheta_N\rangle \approx |\chi\rangle^{\otimes N}$:

A prepares $|\hat{\vartheta}_N\rangle$ locally & teleports Bob's part to Bob

$\Longrightarrow$ uses $M = \log |T(N,\varepsilon)| \leq N(H(p)+\varepsilon)$ copies of $|\phi^+\rangle$.

$\lim \frac{M}{N} = H(p)+\varepsilon$ "entanglement dilution rate"

Since any $\varepsilon > 0$ admissible: Can achieve

asymptotic rate $R_{dilute} = H(p)$ for ent. dilution.

(Alternatively: use majorization, same result).

Protocol for $|\chi\rangle^{\otimes N} \approx |\vartheta_N\rangle \longrightarrow |\phi^+\rangle^{\otimes M}$:

- Can consider $|\vartheta_N\rangle$ as fidelity $\to 1$.
- $|\tilde{\vartheta}_N\rangle$: $|T(N,\varepsilon)|$ Schmidt coefficients,

  largest Schmidt coef. $\leq 2^{-N(H(p)-\varepsilon)}$

- $\Rightarrow |\vartheta_N\rangle$: largest Schmidt coef.

  $\leq \frac{1}{1-\delta} 2^{-N(H(p)-\varepsilon)}$

Choose M s.t. $\frac{1}{1-\delta} 2^{-N(H(\rho)-\varepsilon)} \leq 2^{-M}$

$\Rightarrow (\underbrace{2^{-M}, 2^{-M}, \ldots, 2^{-M}}_{2^M \text{ times}}, 0, \ldots 0) \succ$ (Schmidt coef. of $|\vartheta_N\rangle$)

$\Rightarrow$ can convert $|\vartheta_N\rangle$ to $|\phi^+\rangle^{\otimes M}$ by LOCC.

<u>Protocol</u>: i) A projects onto ε-typ. subspace

$\longrightarrow$ obtains $|\vartheta_N\rangle$.

(i.e.: POVM $\{\Pi_0 = \Pi_{\varepsilon\text{-typ}}, \Pi_1 = \mathbb{1} - \Pi_0\}$)

ii) A & B convert $|\vartheta_N\rangle$ to $|\phi^+\rangle^{\otimes M}$.

Protocol works for any $M \leq N(H(\rho)-\varepsilon) + \log(1-\delta)$

$\Rightarrow$ Rate $R = \lim \frac{M}{N} \longrightarrow H(\rho) - \varepsilon \quad \forall \varepsilon > 0.$

$\Rightarrow$ Asymptotic rate $R_{dist} = H(\rho)$ for entanglement distillation.

<u>We find:</u> Asymptotically (i.e., $N, M \to \infty$),

$$\boxed{\text{distillation rate} = \text{dilution rate} = H(\rho)}$$

Is this optimal?

→ Yes! Otherwise, we could "go in circles" and increase # of $|\phi^+\rangle$ in every iteration by LOCC! ↯

Note: Instead of the Shannon entropy $H(p)$, we typically use the

$$\boxed{\text{von Neumann entropy} \quad S(\rho) = -\mathrm{tr}(\rho \log \rho)}$$

– since $\log \rho$ is defined in the eigenbasis, we have

$-\mathrm{tr}\,\rho \log \rho = -\sum p_i \log p_i$ with $(p_i) = \mathrm{eig}(\rho)$,

i.e., $H(p) = S(\mathrm{tr}_B |\psi\rangle\langle\psi|) = S(\mathrm{tr}_A |\psi\rangle\langle\psi|)$.

Note: This protocol allows us to go between any two states $|\psi\rangle^{\otimes N}$ and $|\phi\rangle^{\otimes M}$ asymptotically reversibly, provided

$$\boxed{N \cdot S(\mathrm{tr}_B |\psi\rangle\langle\psi|) = M \cdot S(\mathrm{tr}_B |\phi\rangle\langle\phi|)}$$

(by going via $|\phi^+\rangle^{\otimes K}$)

$S(\mathrm{tr}_B|\psi\rangle\langle\psi|) =: E(|\psi\rangle)$ can thus be seen as a unique measure of entanglement (in an asymptotic LOCC setting).

Key result: The entropy of entanglement

$$E(|\psi\rangle) = S(\mathrm{tr}_A|\psi\rangle\langle\psi|) = S(\mathrm{tr}_B|\psi\rangle\langle\psi|)$$

uniquely quantifies the amount of entanglement in a pure bipartite state in an asymptotic setting.