Treating the $x \rightarrow y$ process as a classical channel, the capacity is $I_{acc}$ of the ensemble.

We can do better given the Q box — not because of Alice but because of Bob! We can do even better given a channel —

1. How much can Alice learn to Bob? (She decides what $x$ instead of drawing $x \sim p(x)$)
2. She can use $Q$, many times? (But notice incoherence)
3. Same question, but a channel $N$ (instead of $Q$) is available instead?

Elaborating the notations:

\[ \Lambda = \sum_{i=1}^{k} \lambda_i \mathbf{p}_i \]  
\[ \Lambda^{-1} \]  
\[ \Lambda^{-1} \text{ performed only on supp}(\Lambda) \]

Note that the $\text{PM}$ is still well defined if $\mathbf{f} \neq 0$, but otherwise in constant $(x \sim \mathbf{f} \mathbf{p} \mathbf{f}^T)$.

Intuitively, the measurement has an error if the state is $\mathbf{f}$, but the outcome is $\mathbf{f} \mathbf{f}^T$.

How good is the "pretty good" mean?

- If $\mathbf{f}$'s are orthogonal, it is perfect.
- Upon "repeating" many times on $\text{PM}$ being optimal in specific applications.
- For pure states $\mathbf{f} = |\mathbf{x}\rangle \langle \mathbf{x}|$, given equivalently, are prob error $\leq \frac{1}{k} \mathbf{f} \mathbf{f}^T$.

We'll use packing lemma to add error prob.

Pretty good measurement


Gentle measurement lemma (Winter ...):

\[ \text{Tr}(\rho) < 1, 0 \leq \epsilon \leq 1 \]

If $\text{Tr}(\rho M) > 1 - \epsilon$

then $\exists U$ s.t. $\text{Tr}(U^{1/2} \rho U^{1/2} + \epsilon 1/2) \geq \chi^2_i$

Best possible past error state for $\rho$ and come over to $C$.

Interpretations:

Even a state $\rho$, if it mess yields, I am sure why then the state is basically changed by the means.
Def: for a set of states $S = \{ \rho_1, \rho_2, \ldots, \rho_s \}$, let distinguishability error of $S$ be $d(S) = \min_{\pi_i, \pi_j} \Pr(\text{outcome } i \neq j | \text{state } \rho_i)$.

NB: an upper bound $d(S)$ by considering specific measurements.

The packing lemma:

Let $p(S)$ be random, $S_x$ states, $S = \{ \rho_1, \rho_2, \ldots, \rho_s \}$ if projectors $\pi_i, \pi_j$ exist s.t. $\rho_i, \rho_j$:

1. $\Tr(\rho_i \pi_j) \geq 1 - \delta$
2. $\Tr(\rho_i \pi_i) \geq 1 - \delta$
3. $\Tr(\rho_j \pi_j) \leq \frac{1}{d_0}$
4. $\Tr(\rho_i \pi_j) \leq \frac{1}{d_0}$

Then $S_x \leq 2 (1 + \frac{1}{d_0}) + 4 \delta$ (achieved for $d_0$)

What does this lemma mean?

Conditions 1 & 2 say each $\rho_i$ lives in some $d_i$-dim space (up to $\delta$ approx) defined by $\pi_i$. Condition 3 says all $\rho_i$ lives in a space defined by $\pi_j$.

In general, for 2 rank pure states $|\psi\rangle \& |\phi\rangle$,

$
\chi(\langle \psi | \psi \rangle + \langle \phi | \phi \rangle) = 2 \sqrt{\lambda_{\text{min}}(\langle \psi | \phi \rangle)}
$

$\chi(\langle \psi | \phi \rangle)$ is large when $|\psi\rangle \& |\phi\rangle$ have high overlap.

In the current problem: $S = \{ \rho_1, \rho_2, \ldots, \rho_s \}$

We expect $\chi(\langle \rho_i | \rho_j \rangle)$ to be small if the $\rho_i$'s are distinguishable but having mixed state $\rho_i$ complicates things.

Condition 4: $\rho_i, \rho_j$ close s.t. $\rho_i \approx \rho_j$

Since $\chi(\langle \rho_i | \rho_j \rangle) \approx 1 - \delta$, $\lambda_{\text{min}}(\langle \rho_i | \rho_j \rangle)$

$\chi(\langle \rho_i | \rho_j \rangle)$ bounds the max eigenvalue of $S_x$ to be no more than $\frac{1}{d_0}$, or $d_0 = \chi(S_x)$

Where $\chi(\langle \psi | \phi \rangle)$ denotes the max eigenvalue of $\langle \psi | \phi \rangle$.
The failure lemma tells us, if we're communicating using quantum states &
and we know little about them except each lives under the $d$-dim ch, then we can
send $\frac{k}{d}$ of a message into more errors.

\[ \frac{k}{d} \] is the fraction of space we're willing to leave blank.
This comes from how distinguishing $Y - X$ is.
(Also do $\text{rank}(X)$; $d$ is size of $Xx$; \( \frac{d}{\lambda} \), sounds right).
Theorem: $C(C) = \max_{P_X} X(\{P_X, f_X\})$

Again, need a direct coding proof & a converse proof.

Converse: arbitrary

Construc the state $\sum_{x_1, x_2, \ldots, x_n} p(x_1, x_2, x_3) \otimes P_{x_1} \otimes \otimes P_{x_n}$

System labels $X_1, X_2, X_n, B_1, \ldots, B_n$

Then for $R \in I(X_1, \ldots, X_n; B_1, \ldots, B_n)$, arbitrary $y_1, \ldots, y_n$

$$= \sum_{x_1, x_2, \ldots, x_n} s(x_1, x_2, \ldots, x_n) - s(x_1, \ldots, x_n, y_1) - s(y_1, y_2, \ldots, y_n)$$

$\leq \sum_{i=1}^{n} s(y_i, \ldots, y_n) - s(y_i, y_{i+1}, \ldots, y_{i+n})$

Next, let $i \in [n]$, $x_1, \ldots, x_{i-1}, y_{i-1}, x_{i+1}, \ldots, x_n$

$$= \sum_{i=1}^{n} s(y_i, \ldots, y_n) - s(y_i, y_{i+1}, \ldots, y_{i+n}) \leq I(X_i; B_i)$$

Direct Coding:

Recall in Shannon's noisy coding theorem

For any $x_1, x_2, \ldots, x_n$, let $x_i \sim P_X$ iid.

These are typical sequences if $x_i \sim X_i$.

Here, we demand $C$ be defined randomly among strongly typical sequences (next page) and to send message $\sum_{i=1}^{n} x_i$ into the $n$ uses of $\mathcal{C}$.

States to be distinguished by Bob:

$Y_n = f_{x_1} \otimes f_{x_2} \otimes \ldots \otimes f_{x_n}$

Strongly Typical Sequence:

For all $x_1, x_2, \ldots, x_n$, $\xi_i \sim P_x$ is a strongly typical sequence if the empirical distribution of $X$ observed in $\xi_i$ is $\mathcal{C}$-close to $P_x$.

For $x_1, y_1 \in \mathcal{C}, p(y_1 | x_1) = \frac{1}{n} \sum_{i=1}^{n} p(y_{i1} | x_{i1})$ and $\| \frac{1}{n} - p \|_{\infty} = \frac{1}{n} \sum_{i=1}^{n} | p(x_{i1}) - p(y_{i1}) | \leq \varepsilon$

HLL: $\xi$ strongly typical sequences are $\xi'$ typical.

Example: Let $\mathcal{C} = \{1, 2, 3, 4\}$, $p(x) = \frac{1}{4}$, $n = 20$

$\{1, 3, 1, 3, 2, 1, 2, 3, 1, 2, 4, 1, 2, 3, 2, 1, 2, 3, 1\}

$p(1) = \frac{1}{20}$, $p(3) = \frac{3}{20}$, $p(2) = \frac{6}{20}$, $p(4) = \frac{8}{20}$

$p(1) = \frac{1}{4}$, $p(3) = \frac{3}{4}$, $p(2) = \frac{1}{4}$, $p(4) = \frac{3}{4}$

$\| \frac{1}{n} - p \|_{\infty} = 0.1$. So $\mathcal{C}$ is $\varepsilon$-strongly typical.

MB we focus on $n \gg 1$.

Given $C: X_1, \ldots, X_n \in \mathcal{C}$ strongly typical, what do we know about $\mathcal{C}\mathcal{C}x_1, \ldots, x_n$? $X_1, \ldots, X_n$ could be ordered.

Up to reordering, we need a system for $Y \circ \mathcal{C}$

For each $X$, by discussion leading to quantum data compression, there exists a projector $\Pi_X$ s.t.

$\text{Tr} (\Pi_X \otimes \mathcal{C}) = 1 - \delta_1$,

$\dim \Pi_X \leq 2^{-nK_{\mathcal{C}} (2^{\mathcal{C}} - 1)}$.

Given $C$, let $\mathcal{C}_C = \otimes_{X} \mathcal{C}$ acting on the

$K_{\mathcal{C}} = \mathcal{C} \circ \mathcal{C}_C$ systems.

$\text{Tr} (\Pi_X \otimes \mathcal{C}_C) = 1 - \delta_2$, $\dim \Pi_X \leq 2^{-nK_{\mathcal{C}} (2^{\mathcal{C}} - 1)}$.
More precisely, for each $n$, $\gamma_n$ for $C$, $\gamma$, $T$, $\tilde{T}$ defined earlier, and let $\gamma = \frac{1}{2} f_\delta f_\varepsilon$. $T$ projects onto the unique $\hat{f}$:

\begin{align*}
\text{and so on...}
\end{align*}

Finally, back to direct using proof for $C(2)$:

The 6 states are encoded as: $x_1, x_2, \ldots, x_6$.

where $x_1 = y_1 \bullet y_2 \bullet y_3 \bullet y_4 \bullet y_5 \bullet y_6$.

for $C = X_1, X_2, \ldots, X_6$.

A random check strongly supports our argument.

The hacking lemma now applies.

$S_k$, being $f_\gamma$, is strongly typical.

$T_k$, being $T$, is $\delta$-random.

$\mathcal{D}_1$,... $\mathcal{D}_m$, $\mathcal{D}_n = N$

such that $S_k = Y_1 \cdots Y_n$.

The following lemma holds.

$$
\mathcal{D}_n = N [S_k = Y_1 \cdots Y_n] = 2^{p_k (p_k - 1) - \delta}.
$$

Since $S_k$ emits $N$ symbols, applying what we have, $C(2)$ is $\max \gamma (p_k, p_k)$.

\begin{align*}
\text{and so on...}
\end{align*}

What should Alice send to Bob?

Consider all strings of 12 strings of length $n$.

\begin{align*}
\text{with } 50\% 1's, 50\% 2's, 50\% 3's.
\end{align*}

$$
\mathcal{C}_1 = \{0, 1, 2, 3\}.
$$

There are $2^n (111111) = 1571$

Alice is drawing 2 strings from the pool (with replacement), if $d = 0.02$, then are 20=4 messages.

Say $C_1 = 211231212123$.

$$
\text{and so on...}
$$

\begin{align*}
\text{and so on...}
\end{align*}
Bob's measurement: \[ \Lambda_i = \prod_{k \neq i} \Lambda_k, \quad \Lambda = \sum_{i} \Lambda_i, \quad \mathcal{M}_i = \Lambda_i^* \Lambda_i, \quad \mathcal{M}_{120} = \sum \text{ of the } 3 \text{ projections} \]

Tr or rank of \( \Lambda_{120} \approx (1.25)^2 \times 3 \approx 4.5 \)

Theoretically, should be \( \approx \frac{10 \log_{10}(2.5)}{15} \approx 1.5 \) also

It projection onto typical space \( \prod \left( \mathcal{M}_1 + \mathcal{M}_2 + \mathcal{M}_3 \right) \)

which is the full space, \( d_0 \approx 10 \)

Hence and \( \approx \frac{10(1 - \alpha^2 - \alpha^3)}{10 - 2\xi} \approx 120 \) messages.

So \( \prod \mathcal{M}_i \approx \mathcal{M}_{120} \Lambda \).

Same for \( \prod \mathcal{M}_i \Lambda \).

A more collective

\( \Lambda = \prod_{i=1}^{120} \mathcal{M}_i \Lambda \).

\( \left\{ \mathcal{M}_i : \Lambda^2 \right\} \) defines a highly collective measurement.

What is the Issac of the ensemble?

\( \text{E.g., } \left( \mathcal{M}_1 + \mathcal{M}_2 \right) \text{ in the above example?} \)

That of the \( \bigotimes \) three states \( \approx 0.585 \)

Issac for the current ensemble should be even lower.

Now, classical capacity of a quantum channel.

Basic use: Issac is the new "Q" box.

\[ [N] \quad \rightarrow \quad [N'] \]

\[ \quad \rightarrow \quad \mathbb{C} \quad \rightarrow \quad \mathbb{R} \]

Web can also use \( \mathcal{N}^{\text{RR}} \) corresponding Q box.

\[ \left\{ \mathcal{N}^{\text{RR}} \right\} \quad \rightarrow \quad \mathbb{N}^{\text{RR}}(p) \]

It follows from the capacity discussions the Q box that

\[ C(N) \geq \frac{\gamma}{\delta} \quad \mathcal{X}(N^{\delta'}) \]

and defined as

\[ \leq \max_{\mathcal{X}(N^{\delta'})} \frac{\gamma}{\delta} \]

for \( \delta \ll 1 \)

about \( n \) channels.

This is called the

Protocol 1:

\[ x_i \rightarrow (X) \rightarrow (Y) \rightarrow y_i \]

Alice chooses \( p(x) \) to maximize \( I(x; y) \)

\( I(x; y) = \max_{p(x)} \mathbb{E}\left[ \log_2 \frac{1}{P(x,y)} \right] \)

The common capacity of No.

Protocol 2:

\[ x_1 \rightarrow (X) \rightarrow (Y) \rightarrow y_1 \]

Alice chooses \( p(x) \) to maximize \( I(x; y) \)

\( I(x; y) = \max_{p(x)} \mathbb{E}\left[ \log_2 \frac{1}{P(x,y)} \right] \)

This is the capacity \( C(X) \)


\[ x_1 \rightarrow (X) \rightarrow (Y) \rightarrow y_1 \]

Maximize \( I(x; y) \)

Capacity: \( C(X) = \max_{p(x)} \mathbb{E}\left[ \log_2 \frac{1}{P(x,y)} \right] \)

Note that the capacity expression has an optimization over \( p(x) \). It is often called a "repeated" expression in contrast to the capacity expression for a classical channel with \( d \) in which the receiver only uses one of the \( d \) resources (i.e., it is a single resource expression).

The step \( S_i \) is

\[ S_i = S_i \cup S_i \cup \ldots \cup S_i \]

In the converse of capacity of Q-box breaks down when output state \( N^{d}(1) \) is entangled in the channel setting.

To do:
- Problems like not having upper Ando to capacity of
- Some examples: AO Channel
- Adversely effects vs an additivity (this really goes well with the problems you have)