LECTURE 3
AN INTRODUCTION TO QUANTUM INFORMATION THEORY

Advanced Quantum Information and Computing

Thomas Debris-Alazard

Inria, École Polytechnique

- Source coding $\Big($**compression**$\Big)$: remove redundancy/compress as much as possible

An example: compress the language

In French, E is frequent, Z is not

$\longrightarrow$ E is compressed with fewer "symbols" than Z

- Channel coding: add redundancy to recover messages in the presence of noise

An example: spell your name over the phone, send first names!

M like Mike, O like Oscar, R like Romeo, A like Alpha, I like India and N like November

M: **message** ; Mike: **encoding**

Source and Channel coding are "dual"

Information Theory answers the following two (fundamental) questions:

- Ultimate data compression? Entropy
- Ultimate transmission rate of communication? Channel capacity

$\longrightarrow$ Information Theory is much more!

A common denominator: typical sequences/realisations!

Anecdote:

At the police station, is it easier to answer the following questions: what were you doing three Monday ago? or what were you doing a typical Monday?

$\longrightarrow$ Typical realisations: simple mean to answer hard questions!

To generalize information theory to the quantum case!

$\longrightarrow$ Typical sequences were at the core of classical information theory

- But how are defined typical sequences in the classical case and how can we use them to reach the ultimate compression rate?
- Does this concept admit a quantum analogue? Could we also use it to "compress" quantum states?

1. Typical Sequences
2. Shannon’s Compression Theorem
3. Von Neumann Entropy
4. Quantum Typical Subspace Theorem
5. Schumacher’s Compression Theorem

TYPICAL SEQUENCES

- An alphabet: $\mathcal{X}$ discrete
- An event: $\mathcal{E} \subseteq \mathcal{X}$
- Random variable: $\mathsf{X} : \Omega \to \mathcal{X}$
- Probability law / Associated distribution: $\big(\mathbb{P}(\mathsf{X} = x)\big)_{x \in \mathcal{X}}$

Abuse of notation:

$$\mathbb{P}(\mathsf{X} = x) = p(x)$$

Remark: the probability law uniquely determines the random variable

Whatever is the event $\mathcal{E}$,

$$\mathbb{P}(\mathsf{X} \in \mathcal{E}) = \sum_{x \in \mathcal{E}} p(x)$$

Important notation: i.i.d.

$X_1, \ldots, X_n$ are said Independent and Identically Distributed $\left(\text{i.d.d.}\right)$ when they are

1. independent, $\forall \mathcal{I} \subseteq \{1, \ldots, n\}, \forall (x_i)_{i\in\mathcal{I}},\ \mathbb{P}(X_i = x_i,\ i \in \mathcal{I}) = \prod_{i\in\mathcal{I}} \mathbb{P}(X_i = x_i)$
2. identically distributed: $\forall i, j, x, \quad \mathbb{P}(X_i = x) = \mathbb{P}(X_j = x)$

When $X_1, \ldots, X_n$ is i.i.d. and the X_i's are distributed according to X

$\longrightarrow$ We use the notation $X^{\otimes n}$ to denote $(X_1, \ldots, X_n)$

Weak law of large number:

Given i.i.d. random variables $X^{\otimes n} = (X_1, \ldots, X_n)$ and $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X)\right| \leq \varepsilon\right) \xrightarrow[n\to+\infty]{} 1$$

Source of information:

We will be given $X_1, \ldots, X_n : \Omega \longrightarrow \mathcal{X}$, *i.e.*, n random variables over the same space $\mathcal{X}$

- Most of the times we will consider $X_1, \ldots, X_n$ as i.i.d $\Big($to simplify our presentation$\Big)$ but our results stand for more general sources
- n is a parameter, larger it is, more accurate will be our results but $(X_1, \ldots, X_n)$ can be think as one random variable $Y : \Omega \longrightarrow \mathcal{X}^n$

Our goal:

To understand how $(X_1, \ldots, X_n) \in \mathcal{X}^n$ behaves

Most of the time $(X_1, \ldots, X_n)$ *has a "deterministic behaviour"*

$\longrightarrow$ It "always gives" a typical sequence!

$(X_1, \ldots, X_n) \in \{0,1\}^n$ be i.i.d. with $\mathbb{P}(X_i = 1) = p < 1/2$

What is the most probable sequence/realisation?

$(X_1, \ldots, X_n) \in \{0,1\}^n$ be i.i.d. with $\mathbb{P}(X_i = 1) = p < 1/2$

What is the most probable sequence/realisation?

$0 \ldots 0$ appears with probability: $(1-p)^n$

$\longrightarrow$ Most probable event!

But do you expect this realisation?

$(X_1, \ldots, X_n) \in \{0, 1\}^n$ be i.i.d. with $\mathbb{P}(X_i = 1) = p < 1/2$

What is the **most probable sequence/realisation**?

$0 \ldots 0$ appears with probability: $(1 - p)^n$

$\longrightarrow$ Most probable event!

But do you expect this realisation? **No!**

$(X_1, \ldots, X_n) \in \{0,1\}^n$ be i.i.d. with $\mathbb{P}(X_i = 1) = p < 1/2$

What is the most probable sequence/realisation?

$0\ldots0$ appears with probability: $(1-p)^n$

$\longrightarrow$ Most probable event!

But do you expect this realisation? No!

Hamming weight:

Given $\mathbf{x} = (x_1 \ldots x_n) \in \{0,1\}^n$, its Hamming weight is defined as

$$|\mathbf{x}| \stackrel{\text{def}}{=} \sharp\{i : x_i \neq 0\}$$

Chernoff's bound:

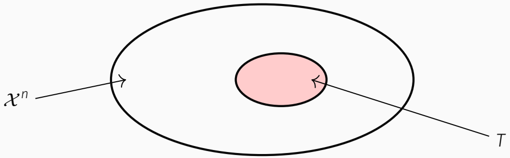
$$\forall \varepsilon > 0, \quad \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i - np\right| \leq \varepsilon n\right) \geq 1 - 2e^{-2\varepsilon^2 n}$$

Typical sequence/realisation: $\mathbf{x}$ such that $|\mathbf{x}| \approx np$, which happens with probability ≈ 1

$\longrightarrow$ Our random vector $(X_1, \ldots, X_n)$ "always" gives a vector with Hamming weight $\approx np$

Given a classical source of information $(X_1, \ldots, X_n) \in \mathcal{X}^n$

Your new motto: focus on typical sequences!



$$\begin{aligned} T &\overset{\text{def}}{=} \text{typical sequences} \\ \mathbb{P}\left((X_1, \ldots, X_n) \in T\right) &\approx 1 \end{aligned}$$

Crucial question:

How many typical sequences are there?

Entropy (informal definition):

$$\text{Entropy}\,(X_1, \ldots, X_n) \overset{\text{def}}{=} \log_2 \sharp T \iff \sharp T = 2^{\text{Entropy}(X_1, \ldots, X_n)}$$

Given a classical source of information $(X_1, \ldots, X_n) \in \mathcal{X}^n$

$$\begin{aligned}
\log_2 \mathbb{P}\Big((X_1, \ldots, X_n) \Big) &\approx \mathbb{E}\Big(\log_2 \mathbb{P}\Big((X_1, \ldots, X_n) \Big) \Big) \\
&= \sum_{(x_1,\ldots,x_n)\in\mathcal{X}^n} p(x_1, \ldots, x_n) \log_2 p(x_1, \ldots, x_n) \quad \Big(\text{transfer formula}\Big) \\
&\overset{\text{def}}{=} -H(X_1, \ldots, X_n) \quad \Big(H \text{ entropy function}\Big)
\end{aligned}$$

Conclusion (informal):

$\mathbb{P}\Big((X_1, \ldots, X_n) = (x_1, \ldots, x_n) \Big)$ is $\approx$ equal to $2^{-H(x_1,\ldots,x_n)}$ $\Big($for typical sequences$\Big)$

or it is $\approx$ equal to 0 $\Big($for non-typical sequences$\Big)$

$\longrightarrow$ There are $2^{H(x_1,\ldots,x_n)}$ “typical sequences” $\Big($by using that $\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) = 1\Big)$

Entropy:

Given $\mathsf{Y} : \Omega \to \mathcal{Y}$, its entropy is defined as

$$H(\mathsf{Y}) \stackrel{\text{def}}{=} \sum_{y \in \mathcal{Y}} p(y) \ \log_2 \frac{1}{p(y)} \ \Big(= \mathbb{E}\left(-\log_2 \mathbb{P}(\mathsf{Y})\right) \Big)$$

with the convention that $0 \times \log_2 \frac{1}{0} = 0$

Some example:

Given Y being uniform over $\mathcal{Y}$,

$$H(\mathsf{Y}) = \sum_{y \in \mathcal{Y}} \frac{1}{\sharp\mathcal{Y}} \log_2 \sharp\mathcal{Y} = \log_2 \sharp\mathcal{Y}$$

Is this computation consistent with our discussion so far?

Entropy:

Given $\mathsf{Y} : \Omega \to \mathcal{Y}$, its entropy is defined as

$$H(\mathsf{Y}) \stackrel{\text{def}}{=} \sum_{y \in \mathcal{Y}} p(y) \ \log_2 \frac{1}{p(y)} \ \left(= \mathbb{E}\left(-\log_2 \mathbb{P}(\mathsf{Y})\right) \right)$$

with the convention that $0 \times \log_2 \frac{1}{0} = 0$

Some example:

Given Y being uniform over $\mathcal{Y}$,

$$H(\mathsf{Y}) = \sum_{y \in \mathcal{Y}} \frac{1}{\sharp\mathcal{Y}} \log_2 \sharp\mathcal{Y} = \log_2 \sharp\mathcal{Y}$$

Is this computation consistent with our discussion so far?

$\longrightarrow$ Yes! The above computation shows that we expect $2^{\log_2 \sharp\mathcal{Y}} = \sharp\mathcal{Y}$ typical sequences when being uniform over $\mathcal{Y}$

For a uniform random variable Y all sequences are typical, no subset is preferred to another

$$\left(\text{given } \mathcal{Z} \subseteq \mathcal{Y}, \ \ \mathbb{P}\left(\mathsf{Y} \in \mathcal{Z}\right) = \frac{\sharp\mathcal{Z}}{\sharp\mathcal{Y}} \ll 1\right)$$

ENTROPY OF INDEPENDENT IDENTICALLY DISTRIBUTED SOURCES

Given an i.i.d source $(X_1, \dots, X_n)$ where the X_i's are distributed according to X

$$
\begin{aligned}
H(X_1, \dots, X_n) &= -\sum_{(x_1,\dots,x_n)} p(x_1, \dots, x_n) \log_2 p(x_1, \dots, x_n) \\
&= -\sum_{(x_1,\dots,x_n)} p(x_1) \cdots p(x_n) \log_2 p(x_1) \cdots p(x_n) \quad \big(\text{By indep. assumption}\big) \\
&= -\sum_{(x_1,\dots,x_n)} p(x_1) \cdots p(x_n) \Big(\log_2 (p(x_1)) + \cdots + \log_2 (p(x_n)) \Big) \\
&= -\sum_{i=1}^{n} \sum_{x_i} p(x_i) \log_2 p(x_i) \sum_{(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)} p(x_1) \cdots p(x_{i-1}) \cdot p(x_{i+1}) \cdots p(x_n) \\
&= -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \quad \big(\text{Probabilities sum to } 1\big) \\
&= \sum_{i=1}^{n} H(X_i) \\
&= nH(X) \quad \big(\text{The } X_i\text{'s are equi-distributed as } X\big)
\end{aligned}
$$

Conclusion:

Given an i.i.d source $X^{\otimes n} = (X_1, \dots, X_n)$:

$$H(X_1, \dots, X_n) = nH(X)$$

We expect $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$ *to be equal to* $2^{-H(X_1, \ldots, X_n)}$ *or* 0

$\longrightarrow$ Typical sequences $(x_1, \ldots, x_n)$ are those for which $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \approx 2^{-H(X_1, \ldots, X_n)}$

We expect $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$ *to be equal to* $2^{-H(X_1,\ldots,X_n)}$ *or* 0

$\longrightarrow$ Typical sequences $(x_1, \ldots, x_n)$ are those for which $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \approx 2^{-H(X_1,\ldots,X_n)}$

Typical set of i.i.d sources

Given $\varepsilon > 0$, n and an i.i.d. source $X^{\otimes n} = (X_1, \ldots, X_n)$, its typical set is defined as:

$$T_\varepsilon^{(n)} \overset{\text{def}}{=} \left\{ x \in \mathcal{X}^n : \left| \frac{1}{n} \log_2 \frac{1}{\mathbb{P}(X^{\otimes n} = x)} - H(X) \right| < \varepsilon \right\}$$
$$= \left\{ x \in \mathcal{X}^n : 2^{-n(H(X)+\varepsilon)} < \mathbb{P}\left(X^{\otimes n} = x\right) < 2^{-n(H(X)-\varepsilon)} \right\}$$

$\Big($in the above definition we implicitly used that $H\left(X^{\otimes n}\right) = nH(X)\Big)$

Theorem:

Given an i.i.d. source $X^{\otimes n} = (X_1, \ldots, X_n)$:

1. $\mathbb{P}\left((X_i)_{1 \leq i \leq n} \in T_\varepsilon^{(n)}\right) \geq 1 - \varepsilon$ for n being sufficiently large
2. $\sharp T_\varepsilon^{(n)} \leq 2^{n(H(X)+\varepsilon)}$
3. $\sharp T_\varepsilon^{(n)} \geq (1-\varepsilon) 2^{n(H(X)-\varepsilon)}$ for n being sufficiently large

Proof:

1. First, by independence assumption and the fact that log maps products into sums,

$$- \log_2 \mathbb{P}(X_1, \ldots, X_n) = - \sum_{i=1}^{n} \log_2 \mathbb{P}(X_i)$$

Notice now by i.i.d. assumption, the $-\log_2 \mathbb{P}(X_i)$ are i.i.d. with expectation $H(X)$ $\Big($transfer formula$\Big)$. Therefore, by the weak law of large number,

$$\mathbb{P}\left(\frac{-\log_2 \mathbb{P}(X_1, \ldots, X_n)}{n} \in [H(X) - \varepsilon, H(X) + \varepsilon]\right) \xrightarrow[n \to +\infty]{} 1$$

But, by definition,

$$\mathbb{P}\left(\frac{-\log_2 \mathbb{P}(X_1, \ldots, X_n)}{n} \in [H(X) - \varepsilon, H(X) + \varepsilon]\right) = \mathbb{P}\Big((X_i)_{1 \le i \le n} \in T_\varepsilon^{(n)}\Big)$$

2. We have the following computation,

$$1 = \sum_{x} p(x) \geq \sum_{x \in T_\varepsilon^{(n)}} p(x) \geq \sum_{x \in T_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)}$$

where we used the definition of typical sequences. It concludes the proof

3. Same reasoning but starting from $1 - \varepsilon \leq \mathbb{P}\Big((X_i)_{1 \le i \le n} \in T_\varepsilon^{(n)}\Big)$ instead of $1 = \sum_x p(x)$

Above we defined the typical set $T_\varepsilon^{(n)}$ and we have shown that $\mathbb{P}\Big((X_i)_{1\leq i\leq n} \in T_\varepsilon^{(n)}\Big) \approx 1$

$\Big($*for n large enough*$\Big)$

$\longrightarrow$ We crucially rely on the independence and equi-distributed assumption!

Do the concept of typical set also hold for more general random variables?

$\longrightarrow$ Yes and it is an extremely general concept

$\Big($not an easy task to find random variables for which there are no typical sets$\Big)$

- See the information theory course
- The “bible” of information theory: *Elements of Information Theory*, T.M. Cover, J. A. Thomas
- A very nice book with a computer scientist approach: *Information Theory, Inference, and Learning Algorithms*, D. J. C. MacKay.

SHANNON'S COMPRESSION THEOREM

Given a classical source of information $(X_1, \ldots, X_n) \in \mathcal{X}^n$

What is the minimum number of bits required to represent outputs of this source of information?

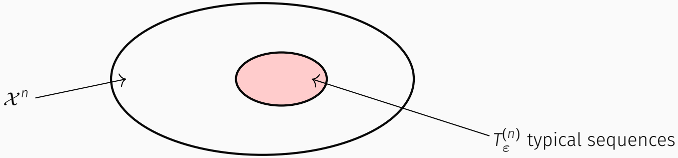
$\Big($optimal compression$\Big)$

$\longrightarrow$ It asks a priori $n \cdot \log_2 \sharp\mathcal{X}$ bits... $\left(\sharp\mathcal{X}^n = 2^{n \log_2 \sharp\mathcal{X}}\right)$

We can do much better by allowing ourselves an exponentially small probability of failure!

$\Big($some outputs $(x_1, \ldots, x_n)$ are not compressed$\Big)$

Given an i.i.d. source $X^{\otimes n} = (X_1, \ldots, X_n)$



$$\mathbb{P}\left(X^{\otimes n} \in T_\varepsilon^{(n)}\right) \approx 1$$

Shannon's compression algorithm

1. Describe elements of $T_\varepsilon^{(n)}$ with bits: it requires $\approx nH(X)$ bits as $\sharp T_\varepsilon^{(n)} \approx 2^{nH(X)}$
2. Given a realisation $\mathbf{x}$: if $\mathbf{x} \in T_\varepsilon^{(n)}$ describe it with bits, otherwise output fail $\perp$

The compression works with probability ≈ 1 and to decompress we just inverse the bit description of elements in $T_\varepsilon^{(n)}$

Conclusion:

We can compress $X^{\otimes n}$ with $n \cdot H(X) \ll n \cdot \log_2 \sharp\mathcal{X}$ bits with a success probability ≈ 1

Furthermore, if we compress with $< nH(X)$ bits, then our failure probability will tend to 1

A non-ambiguous coding is a mapping $\varphi : \mathcal{X}^n \longrightarrow \{0,1\}^+$

Given a source of information $(\mathsf{X}_1, \dots, \mathsf{X}_n)$, the average length of φ is defined as

$$L(\varphi) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{(x_1,\dots,x_n)\in\mathcal{X}^n} p(x_1,\dots,x_n)\ell\left(\varphi(x_1,\dots,x_n)\right) \quad \text{where} \quad \ell(\cdot) \text{ length in number of bits}$$

Shannon's compression theorem:

Given an i.i.d. source $\mathsf{X}^{\otimes n} = (\mathsf{X}_1, \dots, \mathsf{X}_n)$:

1. For all ε and n large enough, It exists a non-ambiguous coding φ such that $L(\varphi) \leq H(\mathsf{X}) + \varepsilon$
2. All non-ambiguous coding verifies $L(\varphi) \geq H(\mathsf{X})$

Exercise:

Given $\mathsf{X}^{\otimes n} \in \mathcal{X}^n$, show that $H\left(\mathsf{X}^{\otimes n}\right)$ cannot be larger than $n \cdot \log_2 \sharp\mathcal{X}$

Entropy is a fundamental concept coming from the size of the typical set

$\longrightarrow$ Entropy quantifies how many bits are required to write non-ambiguously realisations of random variables (Shannon's compression theorem)

Entropy is not some vague concept linked to some property of "Nature"...

We want to generalize this discussion to the case of a quantum source

$\longrightarrow$ A classical source outputs $j \in \mathcal{X}$ $\Big($discrete set$\Big)$ with some probability p_j,

a quantum source will output some quantum state $|\psi_j\rangle \in \mathcal{H}$ $\Big($Hilbert space$\Big)$ with probability p_j

Our approach to led the foundations of quantum information theory:

To investigate the question of compressing a quantum source to highlight what would be a "good" definition of the quantum entropy

VON NEUMANN ENTROPY

An i.i.d. quantum source is is simply repeating n times independently the drawing of a quantum state according to some fixed probability distribution

i.i.d quantum source:

It is defined as $\rho^{\otimes n}$ where ρ is a density operator over some Hilbert space $\mathcal{H}$

$\longrightarrow$ Given an i.i.d. quantum source, the underlying description of ρ is known,

i.e., the knowledge of quantum states $|\psi_j\rangle \in \mathcal{H}$ with associated probabilities p_j

$$\left(\rho = \sum_j p_j \, |\psi_j\rangle\langle\psi_j| \right)$$

In order to introduce a meaningful description of a "quantum entropy" we need to understand the minimum number of qubits to represent $\rho^{\otimes n}$

Trivial approach:

For a density operator $\rho^{\otimes n}$ living in $\left(\mathbb{C}^2\right)^{\otimes n}$ which has dimension 2^n

$\longrightarrow$ it requires n qubits

Could we use less qubits?

COMPRESSION: FIRST EXAMPLE (I)

$\mathcal{Q}$. Source:
- → $|0\rangle$ with probability $1 - p$
- → $|1\rangle$ with probability p

Given a quantum i.i.d source $\rho^{\otimes n} \stackrel{\text{def}}{=} \Big((1-p)\,|0\rangle\langle 0| + p\,|1\rangle\langle 1|\Big)^{\otimes n}$

Can we use less than n qubits to represent this quantum source?

COMPRESSION: FIRST EXAMPLE (I)

$\mathcal{Q}$. Source: $\to |0\rangle$ with probability $1-p$; $\to |1\rangle$ with probability p

Given a quantum i.i.d source $\rho^{\otimes n} \stackrel{\text{def}}{=} \Big((1-p)\,|0\rangle\langle 0| + p\,|1\rangle\langle 1| \Big)^{\otimes n}$

Can we use less than n qubits to represent this quantum source?

Fundamental remark: Given $\mathsf{x} \in \{0,1\}^n$, and $\mathsf{X}^{\otimes n}$ where $\begin{cases} \mathbb{P}(\mathsf{X}=0) = 1-p \\ \mathbb{P}(\mathsf{X}=1) = p \end{cases}$,

$$\mathrm{tr}\left(|\mathsf{x}\rangle\langle\mathsf{x}|\,\rho^{\otimes n}\right) = \langle\mathsf{x}|\,\rho^{\otimes n}\,|\mathsf{x}\rangle = \mathbb{P}\left(\mathsf{X}^{\otimes n} = \mathsf{x}\right) = p^{|\mathsf{x}|}(1-p)^{n-|\mathsf{x}|}$$

But $\mathsf{X}^{\otimes n}$ concentrates over words of Hamming weight $\approx np$ (see Chernoff's bound)

$$\rho^{\otimes n} \approx \sum_{\substack{\mathsf{x}\in\{0,1\}^n \\ |\mathsf{x}|\approx np}} p^{|\mathsf{x}|}(1-p)^{n-|\mathsf{x}|}\,|\mathsf{x}\rangle\langle\mathsf{x}|$$

Conclusion:

$\rho^{\otimes n}$ concentrates over the span of $|\mathsf{x}\rangle$ where $\mathsf{x} \in \{0,1\}^n$ are typical sequences for $\mathsf{X}^{\otimes n}$!

There are $nH(\mathsf{X})$ typical sequences: we can approximate (very well) $\rho^{\otimes n}$ with $n \cdot H(\mathsf{X}) \ll n$ qubits

Given the quantum i.i.d source $\rho^{\otimes n} \overset{\text{def}}{=} \Big((1-p)\,|0\rangle\langle 0| + p\,|1\rangle\langle 1|\Big)^{\otimes n}$

$\longrightarrow$ We can use $n \cdot H(\mathsf{X})$ qubits to represent this quantum source!

This is not surprising: this quantum source can be seen as the classical source "1" with probability p and "0" with probability $1-p$

We can perfectly distinguish outputs by the source using measurement $P_0 = |0\rangle\langle 0|$ and $P_1 = |1\rangle\langle 1|$

But what happens if the quantum states of the source are not orthogonal?

Given a quantum i.i.d source $\rho^{\otimes n} \stackrel{\text{def}}{=} \Big((1-p)\,|0\rangle\langle 0| + p\,|+\rangle\langle +|\Big)^{\otimes n}$

Fundamental remark: $\rho^{\otimes n}$ will consist of $\approx n(1-p)$ copies of $|0\rangle$ and np copies of $|+\rangle$ $\Big($by using law of large numbers$\Big)$,

$$|0\rangle^{\otimes n(1-p)}\,|+\rangle^{\otimes np} = |0\rangle^{\otimes n(1-p)}\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)^{\otimes n(1-p)}$$

But $\left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right)^{\otimes n(1-p)}$ is itself $\approx |0\rangle^{\frac{n(1-p)}{2}}\,|1\rangle^{\frac{n(1-p)}{2}}$ $\Big($law of large number once again$\Big)$. Therefore,

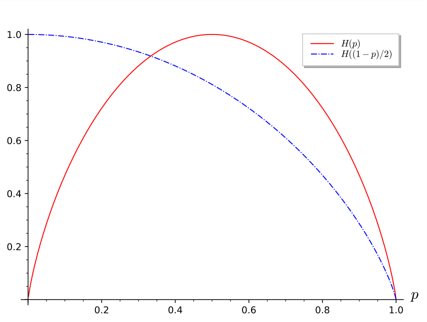
$$|0\rangle^{\otimes n(1-p)}\,|+\rangle^{\otimes np} \approx |0\rangle^{\otimes \frac{n(1+p)}{2}}\,|1\rangle^{\otimes \frac{n(1-p)}{2}}$$

Conclusion:

$\rho^{\otimes n}$ concentrates over the span of the $|\mathbf{x}\rangle$'s with $\mathbf{x} \in \{0,1\}^n$ and $|\mathbf{x}| \approx \frac{n(1-p)}{2}$

For $p \geq \frac{1}{3}$ we can compress more efficiently the source $\Big((1-p)\,|0\rangle\langle 0| + p\,|+\rangle\langle +|\Big)^{\otimes n}$

than the "classical source" $\Big((1-p)\,|0\rangle\langle 0| + p\,|1\rangle\langle 1|\Big)^{\otimes n}$



Intuitively, we reach a better compression rate with $|0\rangle$ and $|+\rangle$ as they share a component in $|0\rangle$

$\big($the condition $p \leq 1/3$ is an artefact of our reasoning$\big)$

Fundamental remark:

In the case of the compression of $\left((1-p)\,|0\rangle\langle 0| + p\,|1\rangle\langle 1|\right)^{\otimes n}$ to a number of qubits given by the classical entropy we used the fact that $|0\rangle$ and $|1\rangle$ are orthogonal quantum states

Each output of the source can be interpreted as “0” or ”1” appearing with probability $1-p$ and p (via a non-destructive measurement)

$\longrightarrow$ If we could reach a smaller compression rate for these kind of source it will contradict Shannon’s theorem stating that we cannot (reliably) compress with fewer bits than the entropy!

Problem (as highlighted by the second example):

It would be far too restrictive to assume that the i.i.d. quantum source only outputs orthogonal quantum states (we would reduce to the classical case)

Problem $\left(\text{as highlighted by the second example}\right)$:

It would be far too restrictive to assume that the i.i.d. quantum source only outputs orthogonal quantum states $\left(\text{we would reduce to the classical case}\right)$

Our diabolic remark:

Any density operator ρ is Hermitian. By the spectral decomposition theorem, it exists an orthonormal basis $\left|\psi_j\right\rangle$ such that

$$\rho = \sum_j p_j \left|\psi_j\middle\rangle\middle\langle\psi_j\right|$$

But ρ is also positive and has trace one! Therefore the p_j form a probability distribution

$\longrightarrow$ We can define the quantum entropy of ρ as the entropy of the "classical source" $\left|\psi_j\right\rangle$ with probability p_j

It seems that our definition of quantum entropy is basis dependent... but notice:

1. Our definition is nothing else than the trace of $-\rho \log_2 \rho$ when decomposing ρ in a spectral basis
2. But the trace is basis independent!

Von Neumann entropy:

Given a density operator ρ, its Von Neumann entropy is defined as:

$$S(\rho) \stackrel{\text{def}}{=} -\operatorname{tr}(\rho \log_2 \rho)$$

Proposition: basis properties of von Neumann entropy

1. The entropy is non-negative. The entropy is zero if and only if the state is pure
2. In a d-dimensional Hilbert space the entropy is at most $\log_2 d$. The entropy is equal to $\log_2 d$ if and only if the system is in the completely mixed state Id/d
3. $S\left(\rho^{\otimes n}\right) = nS(\rho)$
4. Suppose a composite system AB is in a pure state. Then $S(A) = S(B)$
5. Suppose that p_i are probabilities, and the states ρ_i have support on orthogonal subspaces. Then,
$$S\left(\sum_i p_i \rho_i\right) = H\left((p_i)_i\right) + \sum_i p_i S(\rho_i)$$
6. **Joint entropy**: suppose p_i are probabilities, $|i\rangle$ are orthogonal states for a system A, and ρ_i is any set of density operators for another system, B. Then
$$S\left(\sum_i p_i |i\rangle\langle i| \otimes \rho_i\right) = H\left((p_i)_i\right) + \sum_i p_i S(\rho_i)$$

Proof:

See Exercise Session

Some properties of the Shannon entropy fail to hold for the von Neumann entropy

- We always have $H(X) \leq H(X, Y)$ as we always need more bits to compress (X, Y) than X
- For instance given $\rho = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$, its von Neumann entropy is 0, while its von Neumann entropy over its first and second qubits is 1 $\left(\mathrm{tr}_1\,\rho = \mathrm{tr}_2\,\rho = \mathrm{Id}/2\right)$

QUANTUM TYPICAL SUBSPACE

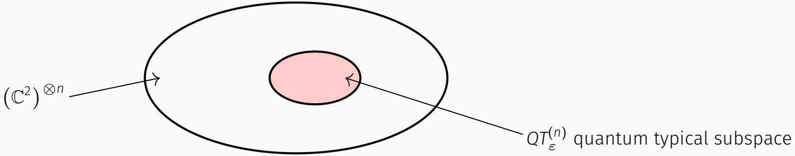
To compress $X^{\otimes n} \in \mathcal{X}^{\otimes n}$ in the classical case: we use two facts

- $X^{\otimes n}$ "always" lives in a smaller set than $\mathcal{X}^n$: the typical set which has size $2^{H(X^{\otimes n})}$
- There is no smaller S than the typical set such that $\mathbb{P}(X^{\otimes n} \in S) \approx 1$

Quantum case:

We will use the same reasoning instead that this times $\rho^{\otimes n}$ concentrates all its mass in a subspace, *i.e.*, projecting $\rho^{\otimes n}$ over some subspace $\left(\text{typical subspace}\right)$ does not change it at much

Given a quantum i.i.d. source $\rho^{\otimes n}$ over the Hilbert space $(\mathbb{C}^2)^{\otimes n}$



$$\Pi_{QT_\varepsilon^{(n)}} \rho^{\otimes n} \approx \rho^{\otimes n} \quad \text{where} \quad \Pi_{QT_\varepsilon^{(n)}} \text{ orthogonal projector onto } QT_\varepsilon^{(n)}$$

$\longrightarrow$ We can then write $\approx$ non-ambiguously $\rho^{\otimes n}$ with $\mathbf{dim}\, QT_\varepsilon^{(n)} \approx S\left(\rho^{\otimes n}\right) \ll n$ qubits!

But how to define the quantum typical subspace $QT_\varepsilon^{(n)}$?

The quantum typical subspace is defined relatively to the spectral decomposition of the density operator

Quantum typical subspace of quantum i.i.d sources:

Let n and a quantum i.i.d. source $\rho^{\otimes n}$. We write ρ as

$$\rho = \sum_{j \in \mathcal{J}} p\left(|\psi_j\rangle\right) |\psi_j\rangle\langle\psi_j|$$

where the $\left\{|\psi_j\rangle : j \in \mathcal{J}\right\}$ are orthogonal quantum states and $p(\cdot)$ a distribution

Given $\varepsilon > 0$, the quantum typical subspace associated to ρ is defined as:

$$QT_{\varepsilon}^{(n)} \stackrel{\text{def}}{=} \text{Span}\left(\left|\varphi^{(1)}\right\rangle \otimes \cdots \otimes \left|\varphi^{(n)}\right\rangle \in \left\{ |\psi_j\rangle : j \in \mathcal{J} \right\}^{\otimes n} : \left| \frac{1}{n} \log_2 \frac{1}{p\left(|\varphi^{(1)}\rangle\right) \cdots p\left(|\varphi^{(n)}\rangle\right)} - S(\rho) \right| < \varepsilon \right)$$

Remark:

In the above definition, $p\left(\left|\varphi^{(1)}\right\rangle\right) \cdots p\left(\left|\varphi^{(n)}\right\rangle\right)$ is nothing else than the probability that the quantum source outputs $\left|\varphi^{(1)}\right\rangle, \dots, \left|\varphi^{(n)}\right\rangle$ after n uses

- Quantum source $\rho^{\otimes n}$ where,

$$\rho = \sum_{j \in \mathcal{J}} p\left(\left|\psi_j\right\rangle\right) \left|\psi_j\right\rangle\!\left\langle\psi_j\right| \text{ where } \left\{\left|\psi_j\right\rangle :\ j \in \mathcal{J}\right\} \text{ is a set of orthogonal quant. states}$$

- Associated typical subspace $QT_\varepsilon^{(n)}$ where,

$$QT_\varepsilon^{(n)} = \mathrm{Span}\Big(\left|\varphi^{(1)}\right\rangle \otimes \cdots \otimes \left|\varphi^{(n)}\right\rangle \in \left\{ \left|\psi_j\right\rangle :\ j \in \mathcal{J} \right\}^{\otimes n} : \left| \frac{1}{n} \log_2 \frac{1}{p\left(\left|\varphi^{(1)}\right\rangle\right) \cdots p\left(\left|\varphi^{(n)}\right\rangle\right)} - S(\rho) \right| < \varepsilon \Big)$$

Orthogonal projector onto $QT_\varepsilon^{(n)}$:

$$\Pi_{QT_\varepsilon^{(n)}} = \sum_{\substack{\left|\varphi^{(1)}\right\rangle, \dots, \left|\varphi^{(n)}\right\rangle \in \left\{\left|\psi_j\right\rangle\right\} \\ \left|\varphi^{(1)}\right\rangle \otimes \cdots \otimes \left|\varphi^{(n)}\right\rangle \in QT_\varepsilon^{(n)}}} \left|\varphi^{(1)}\right\rangle\!\left\langle\varphi^{(1)}\right| \otimes \cdots \otimes \left|\varphi^{(n)}\right\rangle\!\left\langle\varphi^{(n)}\right|$$

Theorem:

Given a quantum i.i.d. source $\rho^{\otimes n}$, for n being sufficiently large,

1. $\mathbf{tr}\left(\Pi_{QT_\varepsilon^{(n)}}\rho^{\otimes n}\right) \geq 1-\varepsilon$
2. $(1-\varepsilon)2^{n(S(\rho)-\varepsilon)} \leq \dim QT_\varepsilon^{(n)} = \mathbf{tr}\left(\Pi_{QT_\varepsilon^{(n)}}\right) \leq 2^{n(S(\rho)+\varepsilon)}$
3. Let Π be a projector onto any subspace of $\mathcal{H}^{\otimes n}$ with dimension $\leq 2^{nR}$ where $R < S(\rho) - 2\varepsilon$ is fixed. Then,
$$\mathbf{tr}\left(\Pi\rho^{\otimes n}\right) \leq \varepsilon + 2^{-\varepsilon n}$$

(Item 3 is a negative result, it will enable to show that we cannot write $\rho^{\otimes n}$ with $< nS(\rho)$ qubits)

Exercise:

Show that $\mathbf{tr}\left(\Pi_F\rho^{\otimes n}\right) \geq 1-\varepsilon$ when F is a subspace containing $QT_\varepsilon^{(n)}$

Lemma 1:

Let Π be a projector over some space of dimension N and A be an Hermitian operator with eigenvalues $\leq \lambda$. Then,

$$\mathrm{tr}(\Pi \mathsf{A}) \leq N \cdot \lambda$$

Proof:

Let $(|x_i\rangle)_i$ be a spectral basis of A. We have,

$$\mathrm{tr}(\Pi \mathsf{A}) = \sum_i \langle x_i| \Pi \mathsf{A} |x_i\rangle = \sum_i \lambda_i \langle x_i| \Pi |x_i\rangle \leq \lambda \cdot \sum_i \langle x_i| \Pi |x_i\rangle = \lambda \cdot \mathrm{tr}(\Pi)$$

where in the first inequality we used our assumption over A and the fact that Π is a positive operator. To conclude the proof, all we have to do is use the fact that the trace of a projective operator is nothing other than the dimension of the space onto which the projection is made

Lemma 2:

Let Π be a projector and A be a positive operator. We have

$$\mathrm{tr}(\Pi A) \leq \mathrm{tr}(A)$$

Proof:

Let $U \overset{\perp}{\oplus} V$ be the space decomposition according to the projector Π. Let $(|u_i\rangle)_i$ and $(|v_i\rangle)_i$ be an orthonormal basis according to this decomposition. We have

$$\mathrm{tr}(\Pi A) = \sum_i \langle u_i| \Pi A |u_i\rangle = \sum_i \langle u_i| A |u_i\rangle$$

where we used that Π is Hermitian and $\Pi |u_i\rangle = |u_i\rangle$ while $\Pi |v_i\rangle = 0$. Notice now that $\langle v_j| A |v_j\rangle \geq 0$ as A is supposed to be positive. We deduce that,

$$\mathrm{tr}(\Pi A) \leq \sum_i \langle u_i| A |u_i\rangle + \sum_j \langle v_j| A |v_j\rangle = \mathrm{tr}(A)$$

where in the last equality we used that $(|u_i\rangle)_i$, $(|v_j\rangle)_j$ is an orthonormal basis

Proof:

1,2. It directly follows from Theorem about the typical set (see Slide 14).

Indeed, by decomposing ρ as $\sum_j p(|\psi_j\rangle) |\psi_j\rangle$ we interpret this quantum i.i.d. source as a classical source outputting "j" with probability $p(|\psi_j\rangle)$.

3. By linearity of the trace,

$$\mathrm{tr}\left(\Pi\rho^{\otimes n}\right) = \mathrm{tr}\left(\Pi\rho^{\otimes n}\Pi_{QT_\varepsilon^{(n)}}\right) + \mathrm{tr}\left(\Pi\rho^{\otimes n}\left(\mathrm{Id} - \Pi_{QT_\varepsilon^{(n)}}\right)\right)$$

Let us look at each component separately. By definition, eigenvectors of ρ are the $|\psi_j\rangle$ and therefore, eigenvectors of $\rho^{\otimes n}$ are given by the $\left|\psi_{j_1}\right\rangle \otimes \cdots \otimes |\psi_{j_n}\rangle$. Notice now that the only eigenvectors of $\rho^{\otimes n}\Pi_{QT_\varepsilon^{(n)}}$ belong to $QT_\varepsilon^{(n)}$. But this subspace if spanned by eigenvectors of $\rho^{\otimes}$ with by definition eigenvalues $\leq 2^{-n(S(\rho)-\varepsilon)}$. Furthermore, $\rho^{\otimes n}\Pi_{QT_\varepsilon^{(n)}}$ is Hermitian as both operators are Hermitian and commute. Therefore, by the using previous Lemma 1:

$$\mathrm{tr}\left(\Pi\rho^{\otimes n}\Pi_{QT_\varepsilon^{(n)}}\right) \leq 2^{nR} \cdot 2^{-n(S(\rho)-\varepsilon)} \leq 2^{-\varepsilon n}$$

Proof:

3.
$$\mathrm{tr}\left(\Pi\rho^{\otimes n}\right) \leq 2^{-\varepsilon n} + \mathrm{tr}\left(\Pi\rho^{\otimes n}\left(\mathrm{Id} - \Pi_{QT_\varepsilon^{(n)}}\right)\right)$$

Notice that Π is a projective operator. Let us show that $\rho^{\otimes n}\left(\mathrm{Id} - \Pi_{QT_\varepsilon^{(n)}}\right)$ is a positive operator. First, it is clearly an Hermitian operator. Let $|u_i\rangle$ be a basis according to the decomposition as space onto which $\Pi_{QT_\varepsilon^{(n)}}$ projects. We have $\Pi_{QT_\varepsilon^{(n)}} |u_i\rangle = |u_i\rangle$ or 0. We deduce that $\langle u_i| \left(\rho^{\otimes n}\left(\mathrm{Id} - \Pi_{QT_\varepsilon^{(n)}}\right)\right) |u_i\rangle$ is either 0 or $\langle u_i| \rho^{\otimes} |u_i\rangle \geq 0$.

We deduce according to previous Lemma 2 that,

$$\begin{aligned}
\mathrm{tr}\left(\Pi\rho^{\otimes n}\left(\mathrm{Id} - \Pi_{QT_\varepsilon^{(n)}}\right)\right) &\leq \mathrm{tr}\left(\rho^{\otimes n}\left(\mathrm{Id} - \Pi_{QT_\varepsilon^{(n)}}\right)\right) \\
&= \mathrm{tr}\left(\rho^{\otimes n}\right) - \mathrm{tr}\left(\rho^{\otimes n}\Pi^{(n)}_{QT_\varepsilon}\right) \\
&= 1 - \mathrm{tr}\left(\rho^{\otimes n}\Pi^{(n)}_{QT_\varepsilon}\right) \\
&\leq 1 - (1 - \varepsilon) \\
&= \varepsilon
\end{aligned}$$

which concludes the proof

We are now almost ready to show that von Neumann entropy is the ultimate quantum data compression rate!

Proof idea:

Project $\rho^{\otimes n}$ onto the quantum typical subspace

$\longrightarrow$ If the source emits eigenstates of ρ, *i.e.*, orthogonal quantum states, then the projection almost does not change ρ and we can decompress $\big($just do nothing$\big)$

Be careful:

Eigenstates of ρ are not necessarily states emitted by the source, and we have no guarantee that any of them actually will be projected within orthogonal $QT_{\varepsilon}^{(n)}$ and the projection can distort it. We need to compute carefully how much it is distorted

SCHUMACHER'S COMPRESSION THEOREM

In all this section we suppose that our i.i.d. quantum source $\rho^{\otimes n}$ *is*

$$\rho = \sum_{j=1}^{d} q_j \left|x_j\right\rangle\!\left\langle x_j\right| \; \left(\text{the source emits } \left|x_j\right\rangle \text{ with probability } q_j\right)$$

where the $\left|x_j\right\rangle$ ***are not necessarily orthogonal*** and belong to $\mathcal{H}$ with $d \stackrel{\text{def}}{=} \dim \mathcal{H}$

$$\left(\dim \mathcal{H}^{\otimes n} = d^n = 2^{n \log_2 d}\right)$$

Our goals:

- To describe some process enabling to store ρ with $nS(\rho) \ll n \log_2 d$ qubits such that, after the storing phase we can reliably recover the quantum state emitted by the source
- Showing that we cannot reliably recover the quantum state emitted by the source if we use $< nS(\rho)$ qubits during the storing phase

Remark:

Don't confuse this decomposition of ρ with its spectral decomposition involving its spectral decomposition, *i.e.*, orthogonal quantum states $\left|\psi_j\right\rangle$ with associated distribution $p\left(\left|\psi_j\right\rangle\right)$

The quantum i.i.d source emits $|x_j\rangle \in \mathcal{H}$ *with probability* q_j*, where* $d \overset{def}{=} \dim \mathcal{H}$*, i.e.,*

$$\rho = \sum_{j=1}^{d} q_j \, |x_j\rangle\langle x_j|$$

Notation:

If after n uses the source emits $\left|x_{j_1}\right\rangle \otimes \cdots \otimes \left|x_{j_n}\right\rangle$, then for $\mathbf{j} = (j_1, \ldots, j_n) \in [1, d]^n$,

$$|\mathbf{j}\rangle \overset{\text{def}}{=} \left|x_{j_1}\right\rangle \otimes \cdots \otimes \left|x_{j_n}\right\rangle$$

Furthermore, as the source is i.i.d. it emits $|\mathbf{j}\rangle$ with probability

$$q_{\mathbf{j}} \overset{\text{def}}{=} q_{j_1} \cdots q_{j_n}$$

COMPRESSION SCHEME

Given our quantum i.i.d. source $\rho^{\otimes n}$ living in the Hilbert space $\mathcal{H}^{\otimes n}$ with $d \overset{\text{def}}{=} \dim \mathcal{H}$

$$\left(\dim \mathcal{H}^{\otimes n} = d^n = 2^{n \log_2 d} \right)$$

Compression scheme:

A compression scheme with rate $R \in (0,1)$ is defined as follows where $\mathbf{U}$ is some unitary,

$$\underset{\mathcal{H}^{\otimes n}}{\boxed{\rho^{\otimes n}}} \xrightarrow{\ \mathsf{U}\ } \underset{(\mathbb{C}^2)^{\otimes nR} \otimes (\mathbb{C}^2)^{\otimes (n(\log_2(d)-R))}}{\boxed{\rho_{\mathsf{U}}}} \xrightarrow{\text{discard 2nd comp.}} \underset{(\mathbb{C}^2)^{nR}}{\boxed{\rho_{\mathsf{Cp}}}} \xrightarrow{\text{adding a 2nd comp.}} \underset{(\mathbb{C}^2)^{\otimes nR} \otimes (\mathbb{C}^2)^{\otimes (n(\log_2(d)-R))}}{\boxed{\rho_{\mathsf{Cp}} \otimes \sigma}} \xrightarrow{\ \mathsf{U}^\dagger\ } \boxed{\rho'}$$

- **Compression phase:** from ρ to ρ_{Cp} which uses only nR qubits
- **Decompression phase:** from ρ_{Cp} to ρ'

But how to measure the reliability of our compression scheme?

Notation:

w_j denotes the quantum state at the end of the process when $|j\rangle$ is emitted

At the end of the process we measure according to the $\left\{ |j\rangle\langle j| \, : \, j \in [1,d]^n \right\}, \mathbf{Id} - \sum_j |j\rangle\langle j|$

$\longrightarrow$ Supposing that $|j\rangle$ was emitted, we recover it with probability

$$\mathbf{tr}\left(|j\rangle\langle j| \, w_j \right)$$

$$p_{\text{succ}} = \sum_{j \in [1,d]^n} q_j \, \mathbf{tr}\left(|j\rangle\langle j| \, w_j \right)$$

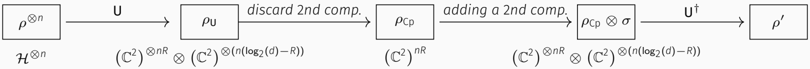
$\longrightarrow$ Probability of being successful (via the law of total numbers)

Reliable compression scheme:

A compression scheme with rate R and with unitary $\mathbf{U}$ is said to be reliable if

$$p_{\text{succ}} \xrightarrow[n \to +\infty]{} 1$$

SOME LEMMA



- Notice that discarding the second component E amounts to tracing out ρ_{U} according to the last $n\,(\log_2(d) - R)$ qubits, *i.e.*, $\rho_{\mathsf{Cp}} = \mathrm{tr}_E\left(\rho_{\mathsf{U}}\right)$

Lemma:

Suppose that we add a fixed pure quantum state $|0\rangle\langle 0|$ to ρ_{Cp} before applying $\mathsf{U}^\dagger$. Let w_{j} be the quantum state just before applying $\mathsf{U}^\dagger$ if $|\mathsf{j}\rangle$ was emitted. Let ρ_{j} be the state ρ_{U} if $|\mathsf{j}\rangle$ was emitted. We have,

$$p_{\mathrm{succ}} = \sum_{\mathsf{j}} q_{\mathsf{j}}\,\mathrm{tr}\left(\rho_{\mathsf{j}} w_{\mathsf{j}}\right)$$

Proof:

Suppose that $|\mathbf{j}\rangle$ was emitted by the source. We have $\rho_{\mathbf{j}} = \mathbf{U}\,|\mathbf{j}\rangle\langle\mathbf{j}|\,\mathbf{U}^\dagger$. Furthermore, $w_{\mathbf{j}}$ is equal to $\mathbf{tr}_E\left(\rho_{\mathbf{j}}\right) \otimes |0\rangle\langle 0|$ (where E denotes the span of the last $n\,(\log_2(d) - R)$ qubits). The final quantum state is $w'_{\mathbf{j}} = \mathbf{U}^\dagger w_{\mathbf{j}}\mathbf{U}$ and we have,

$$p_{\text{succ}} = \sum_{\mathbf{j}} q_{\mathbf{j}}\,\mathbf{tr}\left(|\mathbf{j}\rangle\langle\mathbf{j}|\,w'_{\mathbf{j}}\right) = \sum_{\mathbf{j}} q_{\mathbf{j}}\,\mathbf{tr}\left(\mathbf{U}^\dagger\rho_{\mathbf{j}}\mathbf{U}\mathbf{U}^\dagger w_{\mathbf{j}}\mathbf{U}\right) = \sum_{\mathbf{j}} q_{\mathbf{j}}\,\mathbf{tr}\left(\rho_{\mathbf{j}} w_{\mathbf{j}}\right)$$

Theorem:

Let $\rho^{\otimes n}$ be a quantum i.i.d. source.

1. $\Big($positive part$\Big)$ If $R > S(\rho)$ then there exists a reliable compression scheme of rate R for the quantum source
2. $\Big($negative part$\Big)$ If $R < S(\rho)$ then any compression scheme of rate R is not reliable

$$\boxed{\rho^{\otimes n}} \xrightarrow{\mathsf{U}} \boxed{\rho_{\mathsf{U}}} \xrightarrow{\textit{discard 2nd comp.}} \boxed{\rho_{\mathsf{Cp}}} \xrightarrow{\textit{adding a 2nd comp.}} \boxed{\rho_{\mathsf{Cp}} \otimes \sigma} \xrightarrow{\mathsf{U}^\dagger} \boxed{\rho'}$$

$$\mathcal{H}^{\otimes n} \qquad (\mathbb{C}^2)^{\otimes nR} \otimes (\mathbb{C}^2)^{\otimes(n(\log_2(d)-R))} \qquad (\mathbb{C}^2)^{nR} \qquad (\mathbb{C}^2)^{\otimes nR} \otimes (\mathbb{C}^2)^{\otimes(n(\log_2(d)-R))}$$

The operation from ρ to ρ' is a projection from $\mathcal{H}^{\otimes n}$ to a subspace with dimension nR

$\longrightarrow$ We can use the quantum typical subspace theorem!

Proof:

First, remark that $\rho_{\mathrm{Cp}} \otimes \sigma$ lives in a space of dimension nR as we always add the same component to ρ_{Cp}. Therefore, if the source emits $|\mathbf{j}\rangle$, then $w_{\mathbf{j}}$ $\Big($the state at the end of the process, before measuring$\Big)$ belongs to a space of dimension nR which is independent of the emitted quantum state.

Let $|\gamma_k\rangle$ be an orthonormal basis of this space which diagonalizes $w_{\mathbf{j}}$. Let Π be the projection on this space. We have,

$$\Pi = \sum_{k=1}^{nR} |\gamma_k\rangle\langle\gamma_k| \quad \text{and} \quad w_{\mathbf{j}} = \sum_{k=1}^{nR} \lambda_k |\gamma_k\rangle\langle\gamma_k|$$

Notice that $\lambda_k \in [0,1]$. We have the following computation,

$$\begin{aligned} \mathbf{tr}\left(|\mathbf{j}\rangle\langle\mathbf{j}|\, w_{\mathbf{j}} \right) &= \sum_{k=1}^{nR} \lambda_k\, \mathbf{tr}\left(|\mathbf{j}\rangle\langle\mathbf{j}|\, |\gamma_k\rangle\langle\gamma_k| \right) \\ &\leq \sum_{k=1}^{nR} \mathbf{tr}\left(|\mathbf{j}\rangle\langle\mathbf{j}|\, |\gamma_k\rangle\langle\gamma_k| \right) \\ &= \mathbf{tr}\left(|\mathbf{j}\rangle\langle\mathbf{j}|\, \Pi \right) \end{aligned}$$

But,

$$p_{\mathrm{succ}} = \sum_{\mathbf{j}\in[1,d]^n} q_{\mathbf{j}}\, \mathbf{tr}\left(|\mathbf{j}\rangle\langle\mathbf{j}|\, w_{\mathbf{j}} \right) \leq \sum_{\mathbf{j}\in[1,d]^n} q_{\mathbf{j}}\, \mathbf{tr}\left(|\mathbf{j}\rangle\langle\mathbf{j}|\, \Pi \right) = \mathbf{tr}\left(\rho^{\otimes n} \Pi \right)$$

Proof:

We have shown that it exists a projection Π on a space of dimension nR such that

$$p_{\text{succ}} \leq \mathrm{tr}\left(\rho^{\otimes n}\Pi\right)$$

But $R < S(\rho)$. Let $\varepsilon > 0$ such that $R < S(\rho) - 2\varepsilon$. Therefore, according to the quantum typical subspace theorem $\left(\text{see Slide 41}\right)$, we have for n large enough,

$$p_{\text{succ}} \leq \varepsilon + 2^{-\varepsilon n}$$

showing that with rate $R < S(\rho)$ we cannot reliably compress

Proof:

Suppose $R > S(\rho)$ and let $\varepsilon > 0$ such that $R > S(\rho) + \varepsilon$.

The idea of the proof is to choose $\mathbf{U}$ such that ρ' always lives in the quantum typical subspace $QT_\varepsilon^{(n)}$. We have $\Big($see Slide 41$\Big)$,

$$\dim QT_\varepsilon^{(n)} = \mathrm{tr}\left(\Pi_{QT_\varepsilon^{(n)}}\right) \leq 2^{n(S(\rho)+\varepsilon)} \leq 2^{nR}$$

Let us choose an orthonormal basis $\Big(\,|a\rangle\,\Big)_{1\leq a\leq 2^{n\log_2 d}}$ of $\mathcal{H}$ such that $\mathcal{C}$ **be the span of its** $\leq 2^{nR}$ **first elements contains** $QT_\varepsilon^{(n)}$.

Given $1 \leq a \leq 2^{n\log_2 d}$, let $(x_a|y_a)$ be the encoding of a as bits where x_a consists of the first nR bits. Notice that,

$$\mathbf{y}_a = \mathbf{0} \;\text{ if }\; a \leq 2^{nR} \quad \text{and} \quad \mathbf{y}_a \neq \mathbf{0} \;\text{ if }\; a > 2^{nR}$$

Our unitary is as follows:

$$\mathbf{U} : |a\rangle \mapsto \begin{cases} |x_a, \mathbf{0}\rangle \text{ if } a \leq 2^{nR} \\ |x_a, y_a\rangle \text{ otherwise} \end{cases}$$

Notice that given $|\mathbf{x}_a, \mathbf{y}_a\rangle$, we have $|\mathbf{x}_a\rangle \in \left(\mathbb{C}^2\right)^{\otimes nR}$ while $|\mathbf{y}_a\rangle \in \left(\mathbb{C}^2\right)^{\otimes n\left(\log_2(d)-R\right)}$

Proof:

Suppose that $|j\rangle$ is emitted by the source. We write in the decomposition given by $\mathcal{C}$,

$$|j\rangle = \alpha_j \underbrace{|\alpha(j)\rangle}_{\in\mathcal{C}} + \beta_j \underbrace{|\beta(j)\rangle}_{\in\mathcal{C}^\perp} \quad \text{where } |\alpha_j|^2 = \mathrm{tr}\left(\Pi_{\mathcal{C}} |j\rangle\langle j|\right) \text{ with } \Pi_{\mathcal{C}} \text{ orthogonal projection onto } \mathcal{C}$$

Therefore, according to our notation,

$$|j\rangle = \alpha_j \left(\sum_{a=1}^{2^{nR}} \langle \alpha(j)|a\rangle |a\rangle\right) + \beta_j \left(\sum_{a=2^{nR}+1}^{2^{n\log_2 d}} \langle \mu(j)|a\rangle |a\rangle\right)$$

After applying U we obtain,

$$\begin{aligned} U|j\rangle &= \alpha_j \left(\sum_{x_a\in\{0,1\}^{2nR}} \langle \alpha(j)|a\rangle |x_a, 0\rangle\right) + \beta_j \left(\sum_{\substack{(x_a,y_a)\in\{0,1\}^{n\log_2 d} \\ y_a\neq 0}} \langle \mu(j)|a\rangle |x_a, y_a\rangle\right) \\ &= \alpha_j |\lambda(j)\rangle |0\rangle + \beta_j \sum_{y_a\neq 0} |\gamma(j, y_a)\rangle |y_a\rangle \end{aligned}$$

This pure state $\rho_j \overset{\text{def}}{=} U|j\rangle\langle j| U^\dagger$ $\left(\rho_U \text{ when } |j\rangle \text{ is emitted}\right)$ is given by:

$$\rho_j = |\alpha_j|^2 |\lambda(j), 0\rangle\langle\lambda(j), 0| + \alpha_j\overline{\beta_j} \sum_{y_a\neq 0} |\lambda(j), 0\rangle\langle\gamma(j, y_a), y_a| + \overline{\alpha_j}\beta_j \sum_{y_a\neq 0} |\gamma(j, y_a), y_a\rangle\langle\lambda(j), 0|$$
$$|\beta_j|^2 \sum_{y_a\neq 0} \langle\gamma(j, y_a), y_a|\gamma(j, y_a), y_a\rangle$$

Proof:

$$\rho_{\mathrm{j}} = |\alpha_{\mathrm{j}}|^2 \, |\lambda(\mathrm{j}), 0\rangle \, \langle\lambda(\mathrm{j}), 0| + \alpha_{\mathrm{j}}\overline{\beta_{\mathrm{j}}} \sum_{\mathbf{y}_a \neq \mathbf{0}} |\lambda(\mathrm{j}), 0\rangle \, \langle\gamma(\mathrm{j}, \mathbf{y}_a), \mathbf{y}_a| + \overline{\alpha_{\mathrm{j}}}\beta_{\mathrm{j}} \sum_{\mathbf{y}_a \neq \mathbf{0}} |\gamma(\mathrm{j}, \mathbf{y}_a), \mathbf{y}_a\rangle \, \langle\lambda(\mathrm{j}), 0|$$
$$|\beta_{\mathrm{j}}|^2 \sum_{\mathbf{y}_a \neq \mathbf{0}} \langle\gamma(\mathrm{j}, \mathbf{y}_a), \mathbf{y}_a | \gamma(\mathrm{j}, \mathbf{y}_a), \mathbf{y}_a\rangle$$

But after applying the unitary $\mathbf{U}$, we remove the second component over the last $n(\log_2(d) - R)$ qubits, *i.e.*, we apply the partial trace tr_E $\Big($where E denotes the space of the last $n(\log_2(d) - R)$ qubits$\Big)$. Therefore, as all the $\mathbf{y}_a \neq \mathbf{0}$, we have

$$\mathrm{tr}_E \, \rho_{\mathrm{j}} = |\alpha_{\mathrm{j}}|^2 \, |\lambda(\mathrm{j})\rangle \, \langle\lambda(\mathrm{j})| + |\beta_{\mathrm{j}}|^2 \sum_{\mathbf{y}_a \neq \mathbf{0}} |\gamma(\mathrm{j}, \mathbf{y}_a)\rangle\langle\gamma(\mathrm{j}, \mathbf{y}_a)|$$

Suppose now that during the decompression we add $|0\rangle$ as a second component $\Big($we wrote σ in our definition$\Big)$. It gives before applying $\mathbf{U}^\dagger$,

$$|\alpha_{\mathrm{j}}|^2 \, |\lambda(\mathrm{j}), 0\rangle \, \langle\lambda(\mathrm{j}), 0| + |\beta_j|^2 \sum_{\mathbf{y}_a \neq \mathbf{0}} |\gamma(\mathrm{j}, \mathbf{y}_a), 0\rangle\langle\gamma(\mathrm{j}, \mathbf{y}_a), 0|$$

Proof:

We have before applying $\mathsf{U}^\dagger$,

$$w_{\mathsf{j}} = |\alpha_{\mathsf{j}}|^2 \, |\lambda(\mathsf{j}), 0\rangle \langle \lambda(\mathsf{j}), 0| + |\beta_j|^2 \sum_{\mathsf{y}_a \neq 0} |\gamma(\mathsf{j}, \mathsf{y}_a), 0\rangle\langle\gamma(\mathsf{j}, \mathsf{y}_a), 0|$$

Therefore according to the lemma given in Slide 53, $p_{\text{succ}} = \sum_{\mathsf{j}} q_{\mathsf{j}} \, \mathbf{tr}\left(\rho_{\mathsf{j}} w_{\mathsf{j}}\right)$,

$$\begin{aligned}
\mathbf{tr}\left(\rho_{\mathsf{j}} w_{\mathsf{j}}\right) &= |\alpha_{\mathsf{j}}|^4 + |\alpha_{\mathsf{j}}|^2 |\beta_{\mathsf{j}}|^2 \sum_{\mathsf{y}_a \neq 0} | \langle \gamma(\mathsf{j}, \mathsf{y}_a) | \gamma(\mathsf{j}, \mathsf{y}_a), 0\rangle |^2 \\
&\geq |\alpha_{\mathsf{j}}|^4 \\
&= \left(1 - |\beta_{\mathsf{j}}|^2\right)^2 \\
&\geq 1 - 2|\beta_{\mathsf{j}}|^2 \\
&= 2|\alpha_{\mathsf{j}}|^2 - 1
\end{aligned}$$

Therefore,

$$p_{\text{succ}} \geq 2 \sum_{\mathsf{j}} q_{\mathsf{j}} |\alpha_{\mathsf{j}}|^2 - 1 = 2 \sum_{\mathsf{j}} \mathbf{tr}\left(\pi_{\mathcal{C}} q_j \, |\mathsf{j}\rangle\langle\mathsf{j}| \right) - 1 = 2 \, \mathbf{tr}\left(\pi_{\mathcal{C}} \rho^{\otimes n}\right) - 1$$

But according to the typical quantum subspace theorem (see Slide 41) $\mathbf{tr}\left(\pi_{\mathcal{C}} \rho^{\otimes n}\right)$ which concludes the proof

Conclusion:

We have proved the quantum analogue of Shannon's noiseless coding theorem which involves von Neumann entropy

But we have not investigated the second important topic of information theory:

capacity of noisy channels

$\longrightarrow$ The study of the quantum analogue classical noisy channels and their capacity is a hard task

If you interested by investigating the question of quantum channels:

- *Achieving the Holevo Capacity of a Pure State Classical-Quantum Channel via Unambiguous State Discrimination*, M. Takeoka, H. Krovi and S. Guha

EXERCISE SESSION