From one-shot to asymptotic quantum information theory

Marco Tomamichel

Department of Electrical & Computer Engineering
Faculty of Engineering

QIP 2022, Pasadena

Information processing with finite resources

- Information theory traditionally deals with asymptotic limits where resources are unrestricted (e.g., a channel can be used arbitrarily many times to transmit a packet of information).
- An example is Shannon's channel capacity formula,

$$C = \max_{P_X} I(X : Y), \qquad \text{(in bits/channel use)}$$

  which gives the maximum rate at which we can transmit information over a memoryless channel in the limit of infinite channel uses.
- This is often a good approximation as classical computers can pre- and post-process large amounts of data quite efficiently.

Information processing with finite resources

- Such an asymptotic approximation fails in some regimes (e.g., when time-sharing is used in modern wireless networks and the exchanged packets become small). This has led to renewed interest in corrections to the asymptotic behaviour.
- Noisy intermediate scale quantum (NISQ) era quantum information processors can only cope with a small amount of data coherently for the foreseeable future.
- Although certainly of great conceptual interest, asymptotic results are thus often not practically relevant. We need to take into account corrections to the asymptotic bounds to check feasibility of protocols under NISQ.
- In cryptography, assumptions like that a channel is memoryless are undesirable.

From one-shot to asymptotic quantum information theory

Today's Tutorial

Introducing the toolkit allowing us to study quantum information processing tasks with finite resources.

Focus is on the toolkit and not on individual results. Even if you do not care too much about Shannon theory, its tools might still be of great use in your research.

Examples of results using tools from Shannon theory here at QIP:

- communication complexity and quantum key distribution[1]

[1] Jain, Kundu: A direct product theorem for quantum communication complexity with applications to device-independent QKD.

Examples of results using tools from Shannon theory here at QIP:

- statistical physics[2] [3] [4]
- quantum machine learning[5]

[2] Bluhm, Capel, Harnandez: Exponential decay of mutual information for Gibbs states of local Hamiltonians.

[3] Kuwahara, Saito: Exponential clustering of bipartite quantum entanglement at arbitrary temperatures.

[4] Bravyi, Chowdhury, Gosset, Wocjan: On the complexity of quantum partition functions.

[5] Huang, Küng, Torlai, Albert, Preskill: Provably efficient machine learning for quantum many-body systems.

From one-shot to asymptotic quantum information theory

What we will cover today:

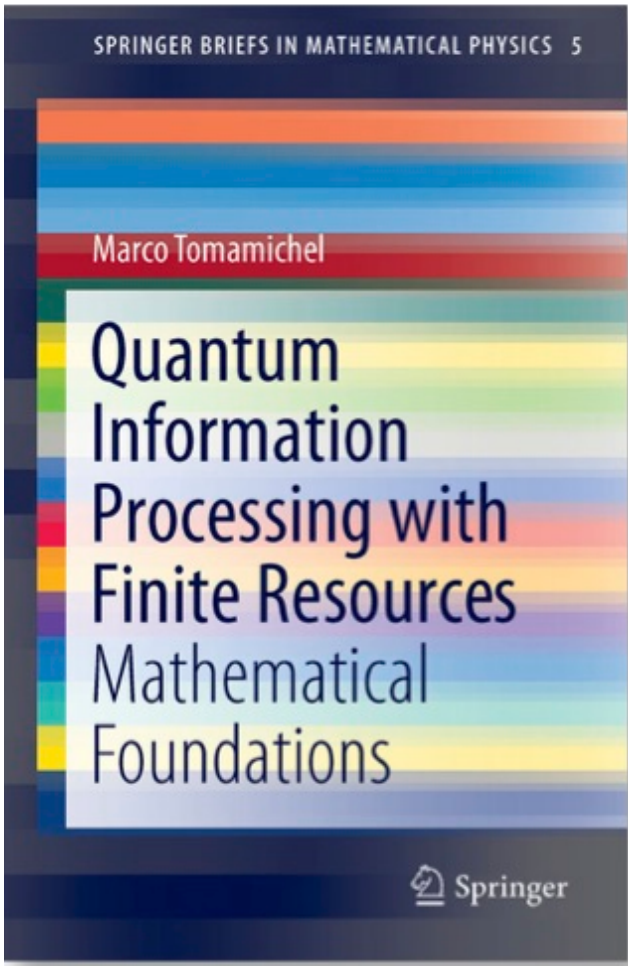
Part I: The structure of quantum information
- for classical and quantum information
- for dealing with additive and multiplicative errors

Part II: The central task: quantum hypothesis testing
- different asymptotic regimes: small, moderate, and large deviation
- a little bit beyond i.i.d.

Part III: Applications to channel coding and randomness extraction
- finite block-length analysis
- error exponents

Finally we will discuss some open problems.

From one-shot to asymptotic quantum information theory



Sources for further reading will be given when possible, except when content is based on my book.

Do not hesitate to contact me with questions, or to point out typos and missing citations.

A permanent home for these slides, with corrections, will be established at `www.marcotom.info/files/qip2022.pdf`

Part I

The structure of quantum information

(a guided tour through the entropy zoo[6])

[6] see also Philippe Faist's entropy zoo: `https://phfaist.com/entropyzoo`

Classical information — Surprisal

- Quantifying information is conceptually non-trivial. It is fruitful to interpret it as the lack of surprise about the outcome of a random experiment.
- Consider a random variable X taking values on a set $\mathcal{X}$ and a probability mass function (pmf)

$$P : \mathcal{X} \to [0, 1], \qquad \sum_{x \in \mathcal{X}} P(x) = 1 \,.$$

- The surprisal when observing some $x \in \mathcal{X}$, denoted $s(x)$, should be monotonically decreasing in $P_X(x)$. We also want it to be additive for independent events.

$$\implies \qquad s(x) := \log \frac{1}{P(x)}$$

Entropies

- Shannon entropy is the average surprisal of X:

$$H(X) := \mathbb{E}[s(X)] = \sum_x P(x) \log \frac{1}{P(x)}$$

- The minimal surprisal of X is often relevant in cryptography:

$$H_{\min}(X) := \min_x s(x) = \log \frac{1}{\max_x P(x)}$$

- The full distribution of the surprisal can be characterised via the cumulant generating function K of $s(X)$:

$$H_\alpha(X) = \frac{K(1-\alpha)}{1-\alpha} = \frac{1}{1-\alpha} \log \left(\sum_x P(x)^\alpha \right)$$

  These are the Rényi entropies of order $\alpha \in (0,1) \cup (1,\infty)$.
- The limits $\alpha \to \{0, 1, \infty\}$ lead to $H_{\max}(X) := \log \operatorname{supp}\{P\}$, $H(X)$ and $H_{\min}(X)$, respectively.

Shannon entropy is not sufficient

- Consider two random variables, X with pmf $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and Y with pmf $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$. We have

$$H(X) = 2, \qquad H_{\min}(X) = 2, \qquad H_{\max}(X) = 2$$
$$H(Y) = 2, \qquad H_{\min}(Y) = 1, \qquad H_{\max}(Y) = \log 5$$

- Asymptotically they behave the same since $H(X) = H(Y)$. For example, we can optimally compress memoryless sources producing X and Y down to 2 bits per symbol.
- But if we want to store a single instantiation of X or Y?
- Which random variable is easier to guess? Which one would you use to power your casino?

Log-likelihood ratio and relative entropies

- Often we are interested in the difference of surprisals under two pmfs P and Q. This is called the log-likelihood ratio:

$$\log \frac{P(x)}{Q(x)}$$

- Its expectation under P is the Kullback-Leibler (KL) divergence:

$$D(P\|Q) := \begin{cases} \sum_x P(x) \log \frac{P(x)}{Q(x)} & \text{if } P \ll Q \\ +\infty & \text{otherwise} \end{cases}$$

  This is well defined even if Q is not normalised.
- And similarly we can define Rényi divergence:

$$D_\alpha(P\|Q) := \frac{1}{\alpha - 1} \log \sum_x P(x)^\alpha Q(x)^{1-\alpha},$$

  where D is again recovered in the limit $\alpha \to 1$.

Divergence as a parent quantity

- We can conveniently express all important entropic quantities in terms of the KL divergence. Consider a joint pmf P_{XYZ}.

$$\begin{aligned}
H(X) &= \log|X| - D(P_X \| U_X) \\
H(X|Y) &= \log|X| - D(P_{XY} \| U_X \times P_Y) \\
&= H(XY) - H(Y) \\
I(X:Y) &= D(P_{XY} \| P_X \times P_Y) \\
&= H(X) - H(X|Y) \\
I(X:Y|Z) &= D(P_{XYZ} \| \, P_Z \times P_{X|Z} \times P_{Y|Z}) \\
&= H(X|Z) - H(X|YZ)
\end{aligned}$$

- The parent quantity concept is extremely useful when considering quantum generalisations.[7] It tells us that we only need to understand divergences and get the rest for free.

[7] This viewpoint in the quantum realm is due to Datta: IEEE T-IT 55, 2009.

Operational interpretations

Why are these the right definitions?

- Because they have nice mathematical properties!
  - $D(P\|Q) \geqslant 0$ for any two pmfs P and Q.
  - $H(X|YZ) \leqslant H(X|Z)$
  - $I(X:Y) \geqslant I(X:Z)$ for $X \leftrightarrow Y \leftrightarrow Z$ a Markov chain
- Because they have operational meaning!
  - $H(X)$ is the minimum amount of memory per symbol needed to store a memoryless source distributed with P_X.
  - $H_{\min}(X) = -\log p_{\text{guess}}(X)$, the probability of correctly guessing X using the optimal strategy.
- These two reasons are tightly connected: the derivation of operational results requires some of these mathematical properties, and conversely the tasks themselves have nice properties, which the quantities characterising them inherit.[8]

[8] Exercise: Derive $H(XY) \leqslant H(X) + H(Y)$ using the above operational characterisation of entropy.

Quantum extensions are not unique

- Recall the definition of the KL divergence:

$$D(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} .$$

- We expect this to be a special case of the quantum divergence when states are diagonal.
- But many expressions have this property, e.g.

$$\begin{aligned} D(\rho\|\sigma) &= \operatorname{tr} \rho(\log \rho - \log \sigma) , \\ \widehat{D}(\rho\|\sigma) &= \operatorname{tr} \rho \log \left(\rho^{\frac{1}{2}} \sigma^{-1} \rho^{\frac{1}{2}} \right) , \\ D^{\text{do-not-use}}(\rho\|\sigma) &= \operatorname{tr} \rho \log \left(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} \right) . \end{aligned}$$

- So which one is the right one?

Data-processing inequality

- The most important property of any divergence, from which most properties of entropies follow, is the data-processing inequality (DPI):

$$D_\alpha(P\|Q) \geqslant D_\alpha(W(P)\|W(Q))$$

  for any channel (stochastic map) W.
- As an example, we expect the entropy $H_\alpha(X)$ to increase when we apply a mixing operation $X \to Y$.
- This is data-processing with a bistochastic map B that preserved the uniform distribution:

$$\begin{aligned} H(X) &= \log|X| - D(P_X\|U_X) \\ &\leqslant \log|X| - D(B(P_X)\|U_X) = H(Y)\,. \end{aligned}$$

Quantum extensions preserving data-processing

- Quantum generalisations of these quantities should maintain DPI, but now for quantum channels (completely positive trace-preserving maps).
- The minimal and maximal extensions of Rényi divergence:[9]

$$\check{D}_\alpha(\rho\|\sigma) := \sup_{\mathcal{M}} D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)), \quad \mathcal{M} \text{ a measurement}$$

$$\widehat{D}_\alpha(\rho\|\sigma) := \inf_{P,Q,\mathcal{P}} D_\alpha(P\|Q), \quad \mathcal{P}(P) = \rho \text{ and } \mathcal{P}(Q) = \sigma$$

- Those quantities satisfy DPI by construction. Moreover, any quantum generalisation satisfying DPI lies between them.

[9] For a general theory of such extensions, see Gour, T: PRA 102, 2020

Maximal extension and geometric Rényi divergence

- The maximal extensions does not have operational use, but a closed form for $\alpha \in (0,1) \cup (1,2]$:[10]

$$\widehat{D}_\alpha(\rho\|\sigma) = \frac{1}{1-\alpha} \log \operatorname{tr} \sigma \left(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}\right)^\alpha$$

- Beyond DPI, the expression is additive for tensor product states and inherits various desirable properties from its close relation to matrix geometric means.
- The limit $\alpha \to 1$ yields the Belavkin-Staszewski divergence:

$$\widehat{D}(\rho\|\sigma) = \operatorname{tr} \rho \log \left(\rho^{\frac{1}{2}} \sigma^{-1} \rho^{\frac{1}{2}}\right) .$$

[10] For larger α a closed form is not known to the best of my knowledge.

Minimal extension and sandwiched Rényi divergence

- The minimal extension is well-known under the name measured divergence.
- It is not additive, but for $\alpha \in [\frac{1}{2}, 1) \cup (1, \infty)$ we can bound

$$\begin{aligned}\check{D}_\alpha(\rho\|\sigma) &\leqslant \lim_{n\to\infty} \frac{1}{n}\check{D}_\alpha\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) \\ &= \frac{1}{\alpha-1}\log \operatorname{tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha =: \widetilde{D}_\alpha(\rho\|\sigma),\end{aligned}$$

  the sandwiched Rényi divegence, the smallest Rényi divergence that satisfies DPI and is also additive.[11]
- It was first used to show the strong converse property for entanglement-breaking channels.[12] We will encounter it again.

[11] Müller-Lennert, Dupuis, Szehr, Fehr, T: JMP 54, 2013
[12] Wilde, Winter, Yang: CMP 331(2), 2014

Sandwiched Rényi divergence

- This limit is highly non-trivial. We know that

$$\lim_{n\to\infty} \frac{1}{n} \check{D}_\alpha \left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) \leqslant \widetilde{D}_\alpha(\rho\|\sigma)$$

  because the r.h.s. is additive and satisfies DPI.[13]

- For the other direction we use the pinching measurement, $\mathcal{P}_{\sigma^{\otimes n}}(\cdot) = \sum_\lambda P_\lambda \cdot P_\lambda$, with P_λ projectors on the eigenspaces of $\sigma^{\otimes n}$. Importantly, it is not too destructive on $\rho^{\otimes n}$:

$$\begin{aligned}\check{D}_\alpha \left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) &\geqslant \widetilde{D}_\alpha \left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \middle\| \sigma^{\otimes n}\right) \\ &\geqslant \widetilde{D}_\alpha \left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) - O(\log n)\,.\end{aligned}$$

  This (essentially) follows from the pinching inequality:

$$\rho^{\otimes n} \leqslant \mathrm{poly}(n)\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\,.$$

[13] Beigi: JMP 54(12), 2013; Frank, Lieb: JMP 54(12), 2013.

Fidelity and other special cases

- Sandwiched Rényi divergences include important special cases.
- For $\alpha = \frac{1}{2}$, it evalutes to

$$\begin{aligned}\widetilde{D}_{\frac{1}{2}}(\rho\|\sigma) &= -2\log \operatorname{tr}\left(\sigma^{\frac{1}{2}}\rho\sigma^{\frac{1}{2}}\right)^{\frac{1}{2}} \\ &= -\log F(\rho,\sigma)\,,\end{aligned}$$

  where F is the Uhlmann fidelity.
- For $\alpha \to \infty$, it limits to

$$\widetilde{D}_{\infty}(\rho\|\sigma) = \inf\{\lambda : \rho \leqslant 2^{\lambda}\sigma\} =: D_{\max}(\rho\|\sigma)\,.$$

  This operator inequality is used to characterise robustness, and log-robustness can be expressed in terms of $D_{\max}$.

Petz Rényi divergence

- We will need yet another family of Rényi divergences (sorry). It was actually the first one properly investigated.
- The Petz Rényi divergence is given for $\alpha \in (0,1) \cup (1,2]$ as

$$\bar{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \operatorname{tr} \rho^\alpha \sigma^{1-\alpha} .$$

- It has operational meaning in hypothesis testing (we will see).
- For $\alpha = 0$ it evaluates to $\bar{D}_0(\rho\|\sigma) = -\log \operatorname{tr} \sigma \Pi_\rho$ where Π_ρ is a projector onto the support of ρ.
- But it does not satisfy DPI for $\alpha > 2$ and is thus not suitable for some applications.
- Today we know whole tribes of Rényi entropy families that all have desirable mathematical properties, but no complete characterisation of them.

Divergence and divergence variance

- All the families are monotonically increasing in α.
- Importantly, the latter two families (sandwiched and Petz) agree in the limit $\alpha \to 1$:

$$\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to 1} \bar{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma)$$

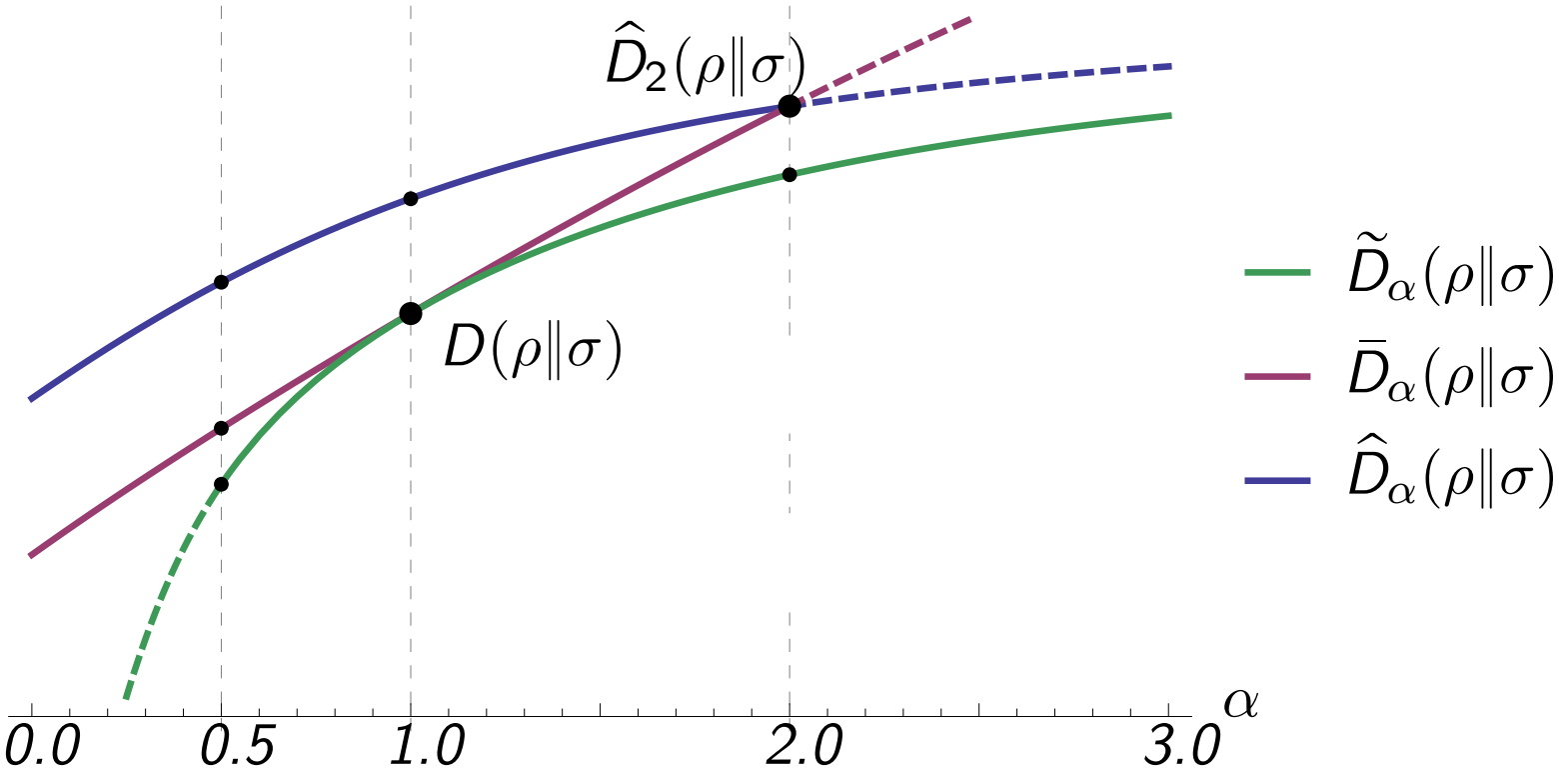
  with $D(\rho\|\sigma) := \operatorname{tr}\rho(\log\rho - \log\sigma)$.
- Moreover, they are in fact tangential at $\alpha = 1$.

$$\lim_{\alpha\to 1} \frac{\mathrm{d}}{\mathrm{d}\alpha}\widetilde{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to 1} \frac{\mathrm{d}}{\mathrm{d}\alpha}\bar{D}_\alpha(\rho\|\sigma) = \frac{V(\rho\|\sigma)}{2\log e}$$

  with $V(\rho\|\sigma) := \operatorname{tr}\rho\,(\log\rho - \log\sigma)^2 - D(\rho\|\sigma)^2$.
- The latter quantity is called the divergence variance.[14]

[14] Classically it corresponds to the variance of the log-likelihood ratio.

Rényi divergence overview



- We will see that both $\bar{D}_\alpha$ and $\widetilde{D}_\alpha$ have operational meaning.

Entropy and conditional entropy

- In contrast, entropy has unique[15] quantum generalisations. For any state ρ_A on a Hilbert space A, we have

$$H(A) := -\operatorname{tr} \rho_A \log \rho_A \quad \text{and} \quad H_\alpha(A) = \frac{1}{1-\alpha} \log \operatorname{tr} \rho_A^\alpha \,.$$

- It vanishes for pure states, is maximal for fully mixed states.
- Conditional entropy is more interesting since ρ_{AB} and its marginals ρ_A and ρ_B generally do not commute.

$$\begin{aligned} H(A|B) &:= -D(\rho_{AB} \| 1_A \otimes \rho_B) \\ &= \max_{\sigma_B} -D(\rho_{AB} \| 1_A \otimes \sigma_B) \,. \end{aligned}$$

- Here 1_A is the identity operator (not a state).

[15] If, indeed, we require it to be non-decreasing under mixing.

Negativity of conditional entropy

- The conditional entropy decomposes as

$$\begin{aligned} H(A|B) &= -D(\rho_{AB} \| 1_A \otimes \rho_B) \\ &= -\operatorname{tr} \rho_{AB} \left(\log \rho_{AB} - \log(1_A \otimes \rho_B) \right) \\ &= -\operatorname{tr} \rho_{AB} \log \rho_{AB} + \operatorname{tr} \rho_{AB} \log(1_A \otimes \rho_B) \\ &= -\operatorname{tr} \rho_{AB} \log \rho_{AB} + \operatorname{tr} \rho_B \log \rho_B \\ &= H(AB) - H(B) \,. \end{aligned}$$

- It can thus be negative, e.g. if ρ_{AB} is pure.
- This is quintessentially quantum. The conditional entropy is always non-negative if the state is separable.

Rényi conditional entropy

- Quantum conditional entropies measure uncertainty from the perspective of an observer with access to a quantum memory.
- To define Rényi conditional entropy things are less clear: we have a choice of divergence and whether we minimise or not.
- For the Petz Rényi relative entropy:

$$\bar{H}^{\downarrow}_{\alpha}(A|B)_{\rho} := -D_{\alpha}(\rho_{AB}\|1_A \otimes \rho_B),$$
$$\bar{H}^{\uparrow}_{\alpha}(A|B)_{\rho} := \max_{\sigma_B} -D_{\alpha}(\rho_{AB}\|1_A \otimes \sigma_B).$$

- For the sandwiched Rényi divergence:

$$\widetilde{H}^{\downarrow}_{\alpha}(A|B)_{\rho} := -\widetilde{D}_{\alpha}(\rho_{AB}\|1_A \otimes \rho_B),$$
$$\widetilde{H}^{\uparrow}_{\alpha}(A|B)_{\rho} := \max_{\sigma_B} -\widetilde{D}_{\alpha}(\rho_{AB}\|1_A \otimes \sigma_B).$$

- They all have their uses - we will see some of them.

Duality relations

- For any pure ρ_{ABC}, the entropy duality relation states

$$H(A|B) + H(A|C) = 0\,.$$

- It is a quantitative expression of monogamy of entanglement.
- For Rényi conditional entropy we get similar relations:[16]

$$\bar{H}^{\downarrow}_{\alpha}(A|B)_\rho + \bar{H}^{\downarrow}_{\beta}(A|C)_\rho = 0 \quad \text{for} \quad \alpha,\beta \in [0,2],\ \alpha+\beta = 2,$$

$$\widetilde{H}^{\uparrow}_{\alpha}(A|B)_\rho + \widetilde{H}^{\uparrow}_{\beta}(A|C)_\rho = 0 \quad \text{for} \quad \alpha,\beta \in \left[\frac{1}{2},\infty\right],\ \frac{1}{\alpha}+\frac{1}{\beta} = 2,$$

$$\bar{H}^{\uparrow}_{\alpha}(A|B)_\rho + \widetilde{H}^{\downarrow}_{\beta}(A|C)_\rho = 0 \quad \text{for} \quad \alpha,\beta \in [0,\infty],\ \alpha\cdot\beta = 1.$$

- They can for example be used to derive entropic uncertainty relations, which are widely used in quantum cryptography.[17]

[16] These relations hint at a deeper geometric structure connecting sandwiched and Petz Rényi divergence: I would love to better understand it!

[17] See, e.g., Coles, Berta, T, Wehner: RMP 89(1), 2017.

Mutual information

- Quantum mutual information measures correlation between two quantum memories.
- There are even more ways to define Rényi mutual information:

$$I(A:B) = D(\rho_{AB} \| \rho_A \otimes \rho_B) = \min_{\sigma_B} D(\rho_{AB} \| \rho_A \otimes \sigma_B)$$
$$= \min_{\sigma_A, \sigma_B} D(\rho_{AB} \| \sigma_A \otimes \sigma_B)$$

- We will just mention two of them:[18]

$$\bar{I}_\alpha^\downarrow(A;B) := \min_{\sigma_B} \bar{D}_\alpha(\rho_{AB} \| \rho_A \times \sigma_B)$$
$$\tilde{I}_\alpha^\downarrow(A;B) := \min_{\sigma_B} \widetilde{D}_\alpha(\rho_{AB} \| \rho_A \times \sigma_B)$$

- I expect them to have operational significance in quantum channel coding (at least for classical-quantum channels)

[18] Gupta, Wilde: CMP 334(2), 2015

Smooth Min-Entropy

- Smooth min-entropy deserves a special mention because it is widely used in quantum cryptography.
- The conditional min-entropy is a special case of the sandwiched Rényi entropy:[19]

$$H_{\min}(A|B)_\rho = \widetilde{H}^{\uparrow}_{\infty}(A|B)\,.$$

- The smooth min-entropy is found by optimising it over a ball of close (sub-normalised) states:

$$H^{\varepsilon}_{\min}(A|B) := \sup_{\tilde{\rho}_{AB} \approx^{\varepsilon} \rho_{AB}} H_{\min}(A|B)_{\tilde{\rho}}\,.$$

- The metric used is the purified distance, based on the fidelity.[20] There is a whole calculus built on it, but this is beyond the scope of this tutorial.

[19] Renner, Ph.D. thesis, ETH Zurich, 2005

[20] T, Colbeck, Renner: IEEE T-IT 56, 2010

Part II

Quantum hypothesis testing

(a central task underlying most of quantum Shannon theory)

Binary quantum hypothesis testing

We first consider the case when only a single copy is available.

Null Hypothesis: The unknown state is ρ.

Alternate Hypothesis: The unknown state is σ.

- Given a copy of the unknown state, perform a test $\{T, 1-T\}$ with $0 \leqslant T \leqslant 1$. Let T indicate the null hypothesis.
- We can define two types of errors:

$$\alpha(T) = \operatorname{tr} \rho(1-T)\,, \qquad \text{(first kind)}$$
$$\beta(T) = \operatorname{tr} \sigma T\,. \qquad \text{(second kind)}$$

- The goal is to find tests that minimise these errors.

Symmetric vs. asymmetric case

- If we treat the two errors on equal footing, the optimisation is

$$\min_T \frac{1}{2}\left(\alpha(T) + \beta(T)\right) = \frac{1}{2}\left(1 - \|\rho - \sigma\|_{\mathsf{tr}}\right)$$

  due to Helstrom.
- But often these errors have different significance. In that case we care about the region of achievable pairs

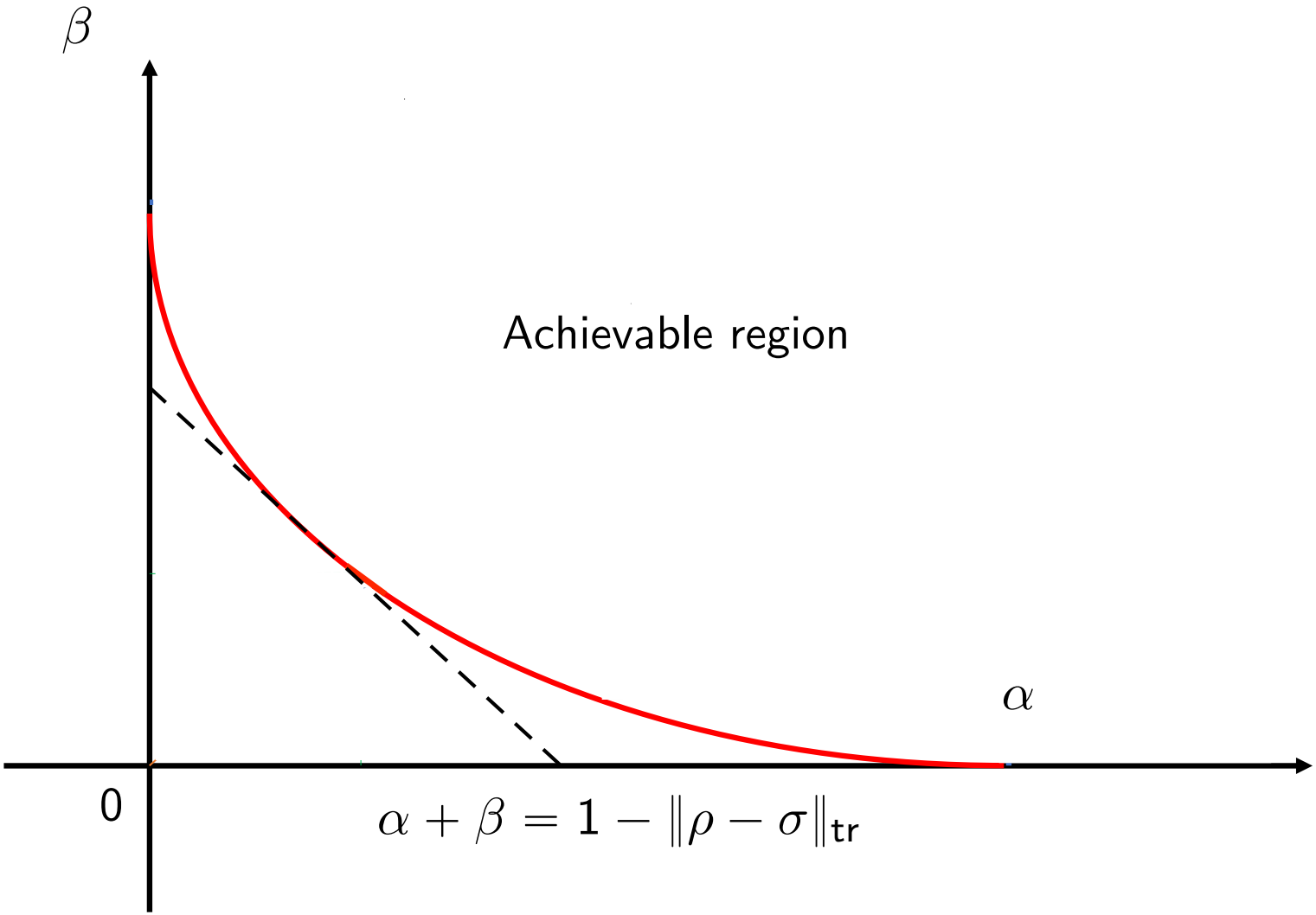
$$\big(\alpha(T), \beta(T)\big)$$

- Or, alternatively, the boundary of the region:

$$\beta^*(\varepsilon) := \min\left\{\beta(T) : 0 \leqslant T \leqslant 1 \wedge \alpha(T) \leqslant \varepsilon\right\},$$

  for $\varepsilon \in [0, 1]$. This is a semi-definite program.
- This tradeoff is often expressed as a divergence-like quantity: $D_h^\varepsilon(\rho\|\sigma) := -\log \beta^*(\varepsilon)$ where β^* is evaluated for ρ and σ.

Achievable error region

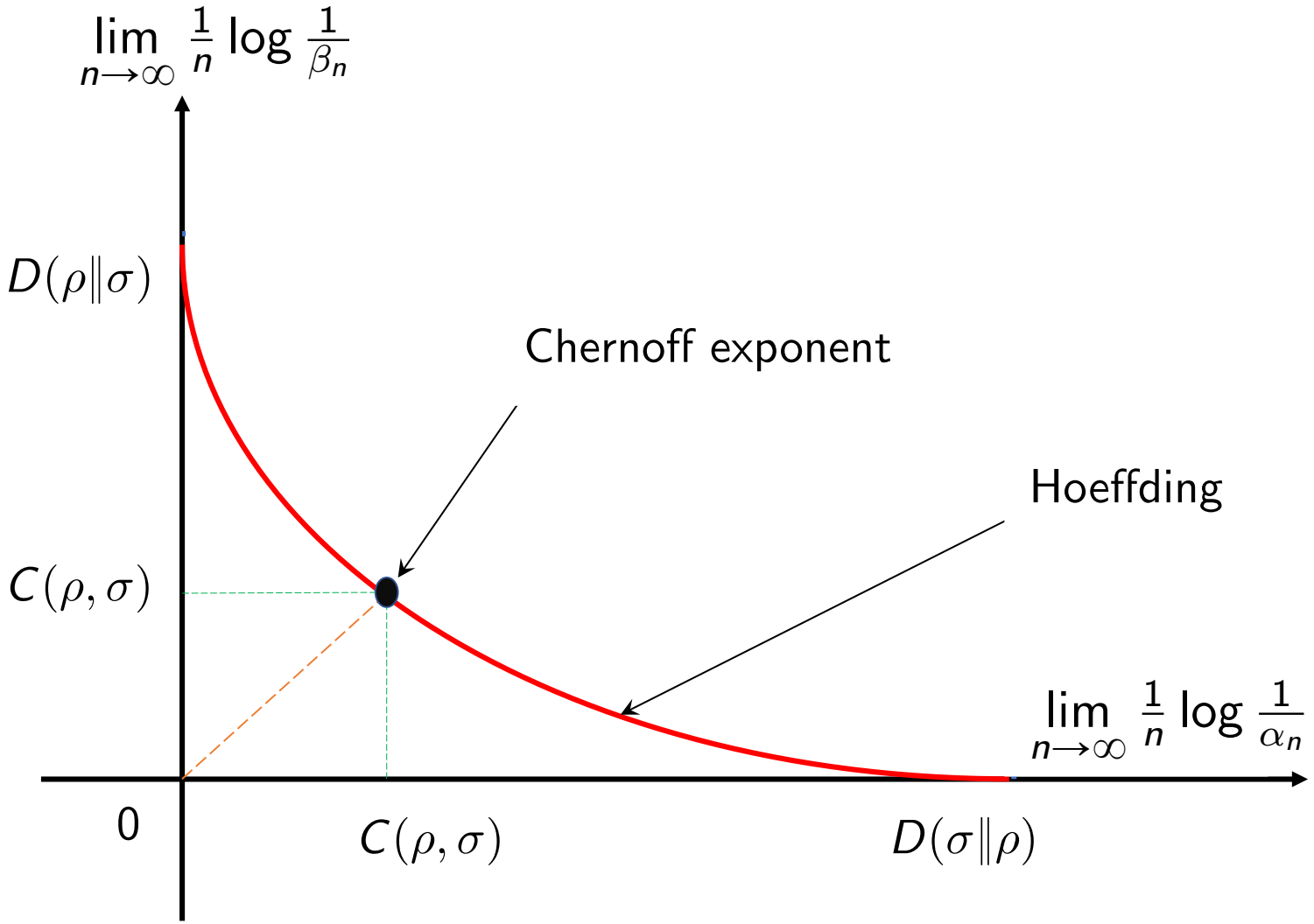


Asymptotic binary quantum hypothesis testing

- Assume now that we can prepare the unknown state multiple times and perform tests T_n jointly on n copies.
- Since $\|\rho^{\otimes n} - \sigma^{\otimes n}\|_{\mathrm{tr}} \to 1$ as $n \to \infty$ it is possible to distinguish between the two hypotheses perfectly in the asymptotic limit.
- The question is only how fast the errors $\alpha(T_n)$ and $\beta(T_n)$ vanish .
- If both errors are to vanish exponentially fast in n, we can consider the region of achievable pairs

$$\left(\lim_{n\to\infty} -\frac{1}{n} \log \alpha(T_n), \ \lim_{n\to\infty} -\frac{1}{n} \log \beta(T_n) \right)$$

Achievable error exponent region



Symmetric case: quantum Chernoff bound

- We are interested in the exponent

$$\sup_{\{T_n\}} \left\{ \lim_{n\to\infty} -\frac{1}{n} \log \left(\frac{1}{2}(\alpha(T_n) + \beta(T_n)) \right) \right\}$$

- A simple bound can be found by noting that

$$\inf_{T_n} \{\alpha(T_n) + \beta(T_n)\} = 1 - \|\rho^{\otimes n} - \sigma^{\otimes n}\|_{\mathrm{tr}} \leqslant \sqrt{F(\rho\|\sigma)}^n,$$

  which leads to a lower bound on the exponent of $\frac{1}{2}\widetilde{D}_{\frac{1}{2}}(\rho\|\sigma)$.

- But the optimal exponent is in fact given by[21]

$$C(\rho,\sigma) := \max_{0\leqslant\alpha\leqslant 1} (1-\alpha)\bar{D}_\alpha(\rho\|\sigma) \geqslant \frac{1}{2}\bar{D}_{\frac{1}{2}}(\rho\|\sigma),$$

  in terms of the Petz Rényi relative entropy.

[21] Audenaert et al: PRL 98, 2007; Nussbaum, Szkoła: Ann. Stat. 37, 2009

Asymmetric case: quantum Stein's lemma

- Stein's lemma considers the case where we keep one error constant and want to know how fast the other error can vanish. That is, the quantity

$$\sup_{\{T_n\}} \left\{ \lim_{n\to\infty} -\frac{1}{n} \log \beta(T_n) : \lim_{n\to\infty} \alpha(T_n) \leqslant \varepsilon \right\} = \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n^*(\varepsilon)}$$

- This limit is given by the Umegaki quantum divergence,

$$D(\rho\|\sigma) = \operatorname{tr} \rho(\log \rho - \log \sigma) \,.$$

- In fact, this result first established this particular generalisation of KL divergence as the correct quantum divergence.[22]

[22] Hiai, Petz: CMP 143(1), 1991

Quantum Hoeffding bound

- Quantum Stein's lemma gives two boundary points, but the region in between is limited by Hoeffding's bound. Here both errors vanish exponentially.
- Assume that $R < D(\rho\|\sigma)$. We want to compute

$$\sup_{\{T_n\}} \left\{ \lim_{n\to\infty} -\frac{1}{n}\log\alpha(T_n) : \lim_{n\to\infty} -\frac{1}{n}\log\beta(T_n) \geqslant R \right\}$$

- The quantum Hoeffing bound establishes this exponent as[23]

$$\sup_{0<\alpha\leqslant 1} \frac{1-\alpha}{\alpha}\left(\bar{D}_\alpha(\rho\|\sigma) - R\right)$$

- If $R \leqslant \bar{D}_0(\rho\|\sigma)$ we can achieve zero error of the first kind by projecting on the kernel of $\rho^{\otimes n}$, so this formula still works.

[23] Hayashi: PRA 76, 062301, 2007; Nagaoka: quant-ph/0611289, 2006

Strong converse exponents

- But what happens if we take $R > D(\rho\|\sigma)$ instead?
- In this case the error will not vanish but converge to 1 instead. And we can again ask for the exponent

$$\inf_{\{T_n\}} \left\{ \lim_{n\to\infty} -\frac{1}{n} \log(1-\alpha(T_n)) : \lim_{n\to\infty} -\frac{1}{n} \log \beta(T_n) \geqslant R \right\}$$

- This is called the strong converse exponent and it is given as[24]

$$\sup_{\alpha>1} \frac{\alpha-1}{\alpha} \left(R - \widetilde{D}_\alpha(\rho\|\sigma) \right) .$$

- This gives an operational interpretation of the family of sandwiched Rényi divergences.

[24] Mosonyi, Hiai: CMP 334(3), 2015

Second-order corrections to Stein's lemma

- From Stein's lemma we learned that

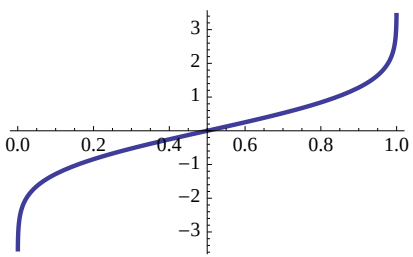
$$\log \frac{1}{\beta^*(\varepsilon)} = nD(\rho\|\sigma) + o(n)$$

- One might ask if we can expand this further. Indeed, we can[25]

$$\log \frac{1}{\beta^*(\varepsilon)} = nD(\rho\|\sigma) + \sqrt{nV(\rho\|\sigma)}\Phi^{-1}(\varepsilon) + O(\log n)\,,$$

where $V(\rho\|\sigma) = \operatorname{tr}\rho(\log\rho - \log\sigma)^2 - D(\rho\|\sigma)^2$ is the divergence variance and Φ is the cumulative normal distribution, i.e.

$$\Phi^{-1}(\varepsilon) =$$



[25] Li: Ann. Stat. 42(1), 2014; T, Hayashi: IEEE T-IT 59(11), 2013

Subexponential decay and moderate deviations

- We can have the cake and eat it too!
- We want the error of the first kind to vanish fast (but not exponentially) while having the error of the second kind decaying at the optimal exponential rate.
- Consider x_n such that $x_n \to 0$ and $\sqrt{n}x_n \to \infty$ as $n \to \infty$. Define a sub-exponential sequence

$$\varepsilon_n := \exp(-nx_n^2)\,.$$

- We get the following moderate deviation expansion:[26]

$$\log \frac{1}{\beta^*(\varepsilon_n)} = nD(\rho\|\sigma) - \sqrt{2V(\rho\|\sigma)}nx_n + o\left(nx_n\right)\,.$$

[26] Cheng, Hsieh: IEEE T-IT 64(2), 2018; Chubb, T, Tan: CMP 355(3), 2017

Beyond simple hypotheses

- There remain many open questions even here, once we change the question slightly.
- Assume the hyptheses are composite, that is

  Null Hypothesis: The unknown state in a set $\mathcal{S}_0$.
  Alternate Hypothesis: The unknown state is in a set $\mathcal{S}_1$.
- For commuting (classical) states and finite sets the exponents are understood. E.g., Stein's lemma holds with the rate

$$\min_{\rho\in\mathcal{S}_0,\sigma\in\mathcal{S}_1} D(\rho\|\sigma)\,.$$

- But the exponents for general quantum states (and even classical states with non-compact sets[27]) do not allow such a simple form.[28]

[27] Mosonyi, Szilágyi, Weiner: On the error exponents of binary state discrimination with composite hypotheses

[28] For other recent progress, see Berta, Brandao, Hirche: CMP 385, 2021.

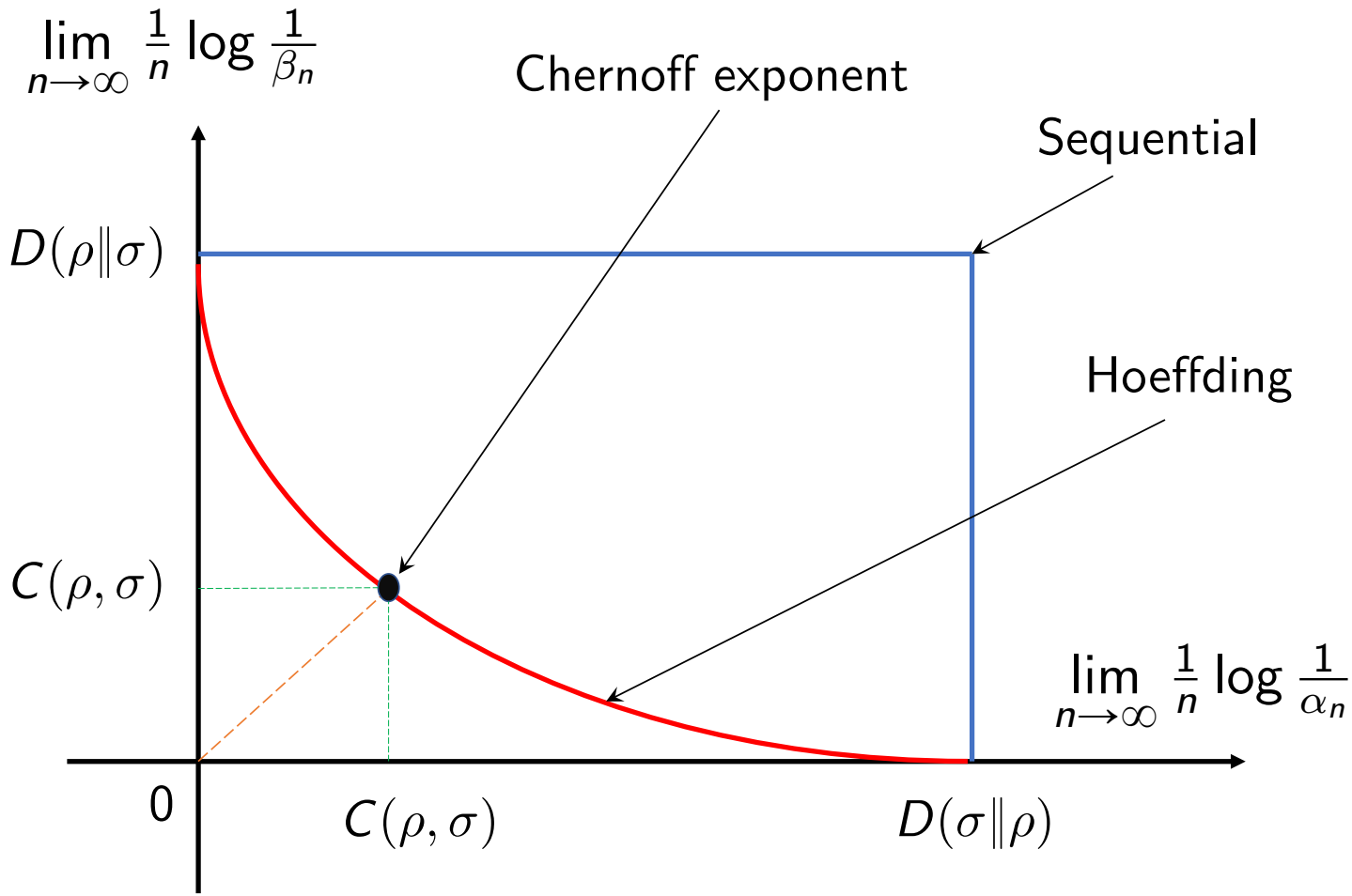
Sequential quantum hypothesis testing

- In our model a decision has to be made after receiving exactly n copies of the unknown state.
- We might also consider sequential strategies that use no more than n copies in expectation, but where the number is not a priori fixed.[29]
- Surprisingly tradeoffs change significantly with this slight change! In fact, there are no more tradeoffs and we can ensure that both

$$-\frac{1}{n}\log\alpha_n \to D(\sigma\|\rho) \quad \text{and} \quad -\frac{1}{n}\log\beta_n \to D(\rho\|\sigma)\,.$$

[29] Martínez-Vargas et al: PRL 126, 2021; Li, T, Tan: CMP (accepted), 2022

Achievable error exponents using sequential strategies



Part III

Applications to channel coding and randomness extraction

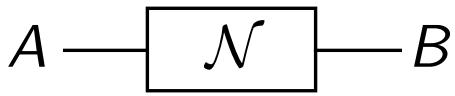
(just two examples out of many)

From one-shot to asymptotic information theory

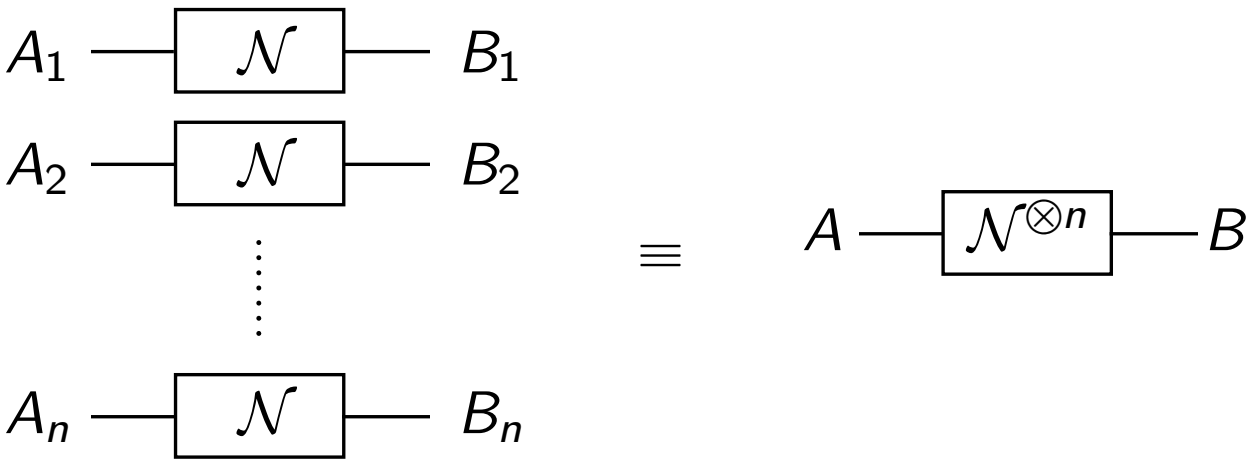
- In information theory we ideally want to express operational quantities (e.g. the maximal rate at which we can transmit information over a channel) in terms of relatively simple information quantities (e.g. the capacity formula).
- A modern appraoch splits this into two parts:
  - The information-theoretic part gives upper and lower bounds on the one-shot operational quantity. This does not use the i.i.d. or memoryless structure. Often these bounds can be expressed in terms of a hypothesis testing problem.
  - The statistical part then evaluates these one-shot bounds asymptotically, in different regimes.
- We explore this now with the example of entanglement-assisted communication over quantum channels.

III.A: Quantum channel coding

- Quantum channel: completely positive trace-preserving map $\mathcal{N} \equiv \mathcal{N}_{A \to B}$ from (linear operators on) A to B.



- The channel is memoryless:



- Channel coding simplifies considerably if we allow the sender and receiver to share entanglement.

Entanglement-assisted codes

- Entanglement-assisted code: quadruple

$$\mathcal{C}_n = \left\{ \mathcal{M},\ |\varphi\rangle_{A'B'},\ \{\mathcal{E}^m_{A'\to A}\}_{m\in\mathcal{M}},\ \{\Lambda^m_{BB'}\}_{m\in\mathcal{M}} \right\}.$$

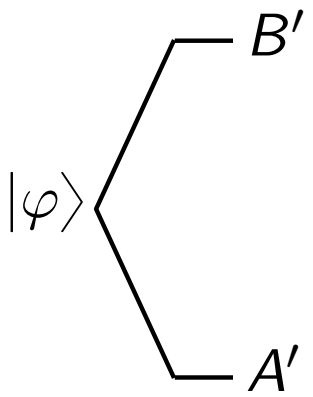
1. Set of messages: $\mathcal{M}$.
2. Resource state: $|\varphi\rangle_{A'B'}$.
3. encoder $\mathcal{E}$: a quantum channel $\mathcal{E}^m_{A'\to A}$ for each message m.
4. decoder $\mathcal{D}$: a positive operator valued measure where $\Lambda^m_{BB'}$ indicates that we decode to m.

Entanglement-assisted codes

- Entanglement-assisted code: quadruple

$$\mathcal{C}_n = \left\{ \mathcal{M},\ |\varphi\rangle_{A'B'},\ \{\mathcal{E}^m_{A'\to A}\}_{m\in\mathcal{M}},\ \{\Lambda^m_{BB'}\}_{m\in\mathcal{M}} \right\}.$$

1. Set of messages: $\mathcal{M}$.
2. Resource state: $|\varphi\rangle_{A'B'}$.
3. encoder $\mathcal{E}$: a quantum channel $\mathcal{E}^m_{A'\to A}$ for each message m.
4. decoder $\mathcal{D}$: a positive operator valued measure where $\Lambda^m_{BB'}$ indicates that we decode to m.

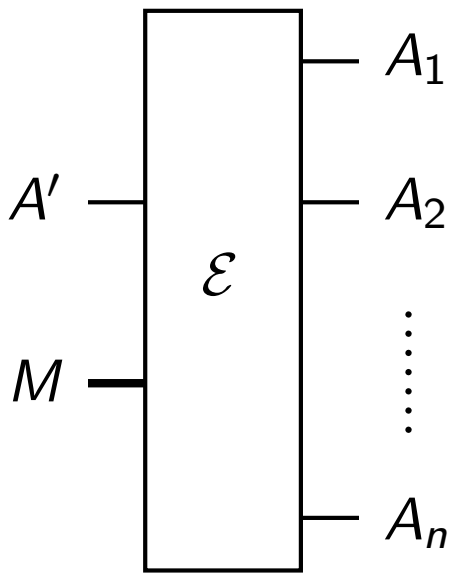


Entanglement-assisted codes

- Entanglement-assisted code: quadruple

$$\mathcal{C}_n = \left\{ \mathcal{M},\ |\varphi\rangle_{A'B'},\ \{\mathcal{E}^m_{A'\to A}\}_{m\in\mathcal{M}},\ \{\Lambda^m_{BB'}\}_{m\in\mathcal{M}} \right\}.$$

1. Set of messages: $\mathcal{M}$.
2. Resource state: $|\varphi\rangle_{A'B'}$.
3. encoder $\mathcal{E}$: a quantum channel $\mathcal{E}^m_{A'\to A}$ for each message m.
4. decoder $\mathcal{D}$: a positive operator valued measure where $\Lambda^m_{BB'}$ indicates that we decode to m.

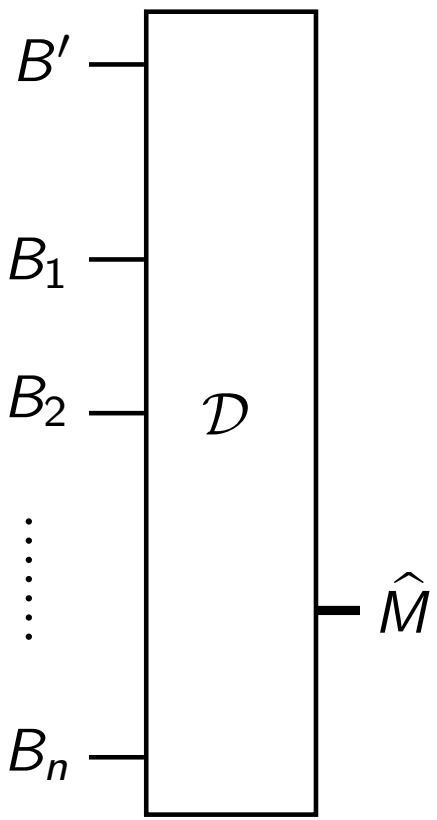


Entanglement-assisted codes

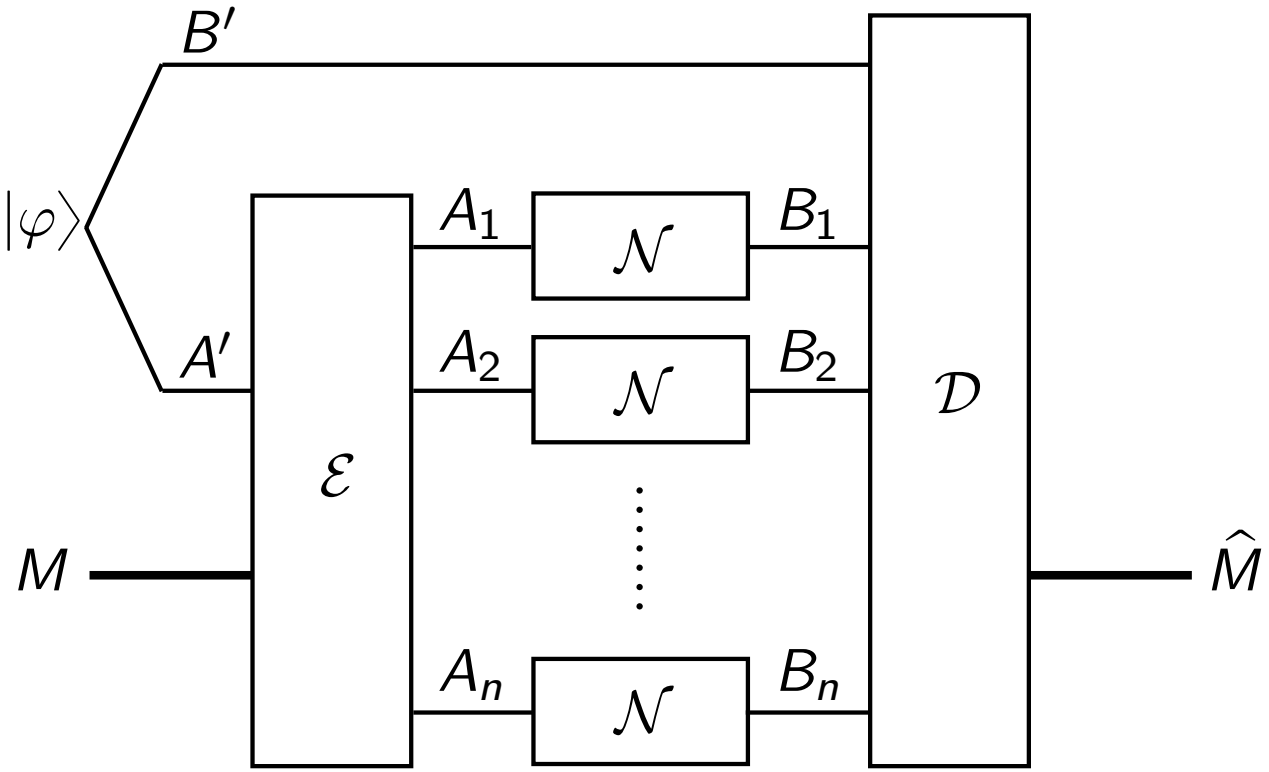
- Entanglement-assisted code: quadruple

$$\mathcal{C}_n = \left\{ \mathcal{M},\ |\varphi\rangle_{A'B'},\ \{\mathcal{E}^m_{A'\to A}\}_{m\in\mathcal{M}},\ \{\Lambda^m_{BB'}\}_{m\in\mathcal{M}} \right\}.$$

1. Set of messages: $\mathcal{M}$.
2. Resource state: $|\varphi\rangle_{A'B'}$.
3. encoder $\mathcal{E}$: a quantum channel $\mathcal{E}^m_{A'\to A}$ for each message m.
4. decoder $\mathcal{D}$: a positive operator valued measure where $\Lambda^m_{BB'}$ indicates that we decode to m.



Average error probability



- Average probability of error: M uniformly random in $\mathcal{M}$,

$$p_{\mathrm{err}}(\mathcal{C}_n, \mathcal{N}^{\otimes n}) := \Pr\left[M \neq \hat{M}\right].$$

Achievable region for finite block-length n

- Non-Asymptotic achievable region: A triple $\{R, n, \varepsilon\}$ is achievable if there exists a code $\mathcal{C}_n$ for $\mathcal{N}^{\otimes n}$ with

$$\frac{1}{n} \log |\mathcal{M}| \geqslant R, \quad \text{and} \quad p_{\text{err}}(\mathcal{C}_n, \mathcal{N}^{\otimes n}) \leqslant \varepsilon$$

- The tolerated error is fixed to $\varepsilon \in (0, 1)$.
- The boundary of the achievable region, $R^*_{\mathcal{N}}(n; \varepsilon)$, is the maximum rate R such that $\{R, n, \varepsilon\}$ is achievable.

- We want to first understand the one-shot quantity $R^*_{\mathcal{N}}(1; \varepsilon)$ and then the function $n \mapsto R^*_{\mathcal{N}}(n; \varepsilon)$ for large n and fixed ε.

One-shot bounds: converse

- One can tightly relate the channel coding problem with hypothesis testing:[30]

$$R^*_{\mathcal{N}}(1;\varepsilon) \leqslant \max_{\rho_A} \min_{\sigma_B} D^{\varepsilon}_h\big(\mathcal{N}_{A\to B}(\rho_{AA'})\big\|\rho_{A'}\otimes\sigma_B\big),$$

  where $\rho_{AA'}$ is a purification of ρ_A.
- This can be understood as a hypothesis test between the actual channel output, $\mathcal{N}_{A\to B}(\rho_{AA'})$, and the output of a replacer channel that always outputs σ_B.
- This is called the meta-converse for entanglement-assisted quantum channel coding, since various other converse bounds can be derived from it.

[30] Matthews, Wehner: IEEE T-IT 60, 2014

One-shot bounds: achievability

- We can achieve this bound using a technique called position-based coding:[31]

$$R^*_{\mathcal{N}}(1;\varepsilon) \geqslant \max_{\rho_A} D_h^{\varepsilon-\delta}\left(\mathcal{N}_{A\to B}(\rho_{AA'})\middle\|\rho_{A'}\otimes\mathcal{N}_{A\to B}(\rho_A)\right) - \log\frac{4\varepsilon}{\delta^2}$$

- In this sense the bound is tight.
- And so hopefully now we can just apply what we know about hypothesis testing to get the asymptotics...

[31] Qi, Wang, Wilde: J. Phys. A 51(44), 2018

First-order asymptotics

- It turns out that this is not so easy because of the optimisation over the input states!
- The initial result considered vanishing error:[32]

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} R^*_{\mathcal{N}}(n;\varepsilon) = C_{\mathrm{ea}}(\mathcal{N})\,.$$

  where $C_{\mathrm{ea}}(\mathcal{N}) = \max_{\rho_A} I(A:B)_\tau$ is the entanglement-assisted capacity and $\tau_{AB} = \mathcal{N}_{A'\to B}\left(|\psi^\rho\rangle\langle\psi^\rho|_{AA'}\right)$ the output state.
- The strong converse follows from the quantum reverse Shannon theorem.[33]

$$R^*_{\mathcal{N}}(n;\varepsilon) = C_{\mathrm{ea}}(\mathcal{N}) + o(1)$$

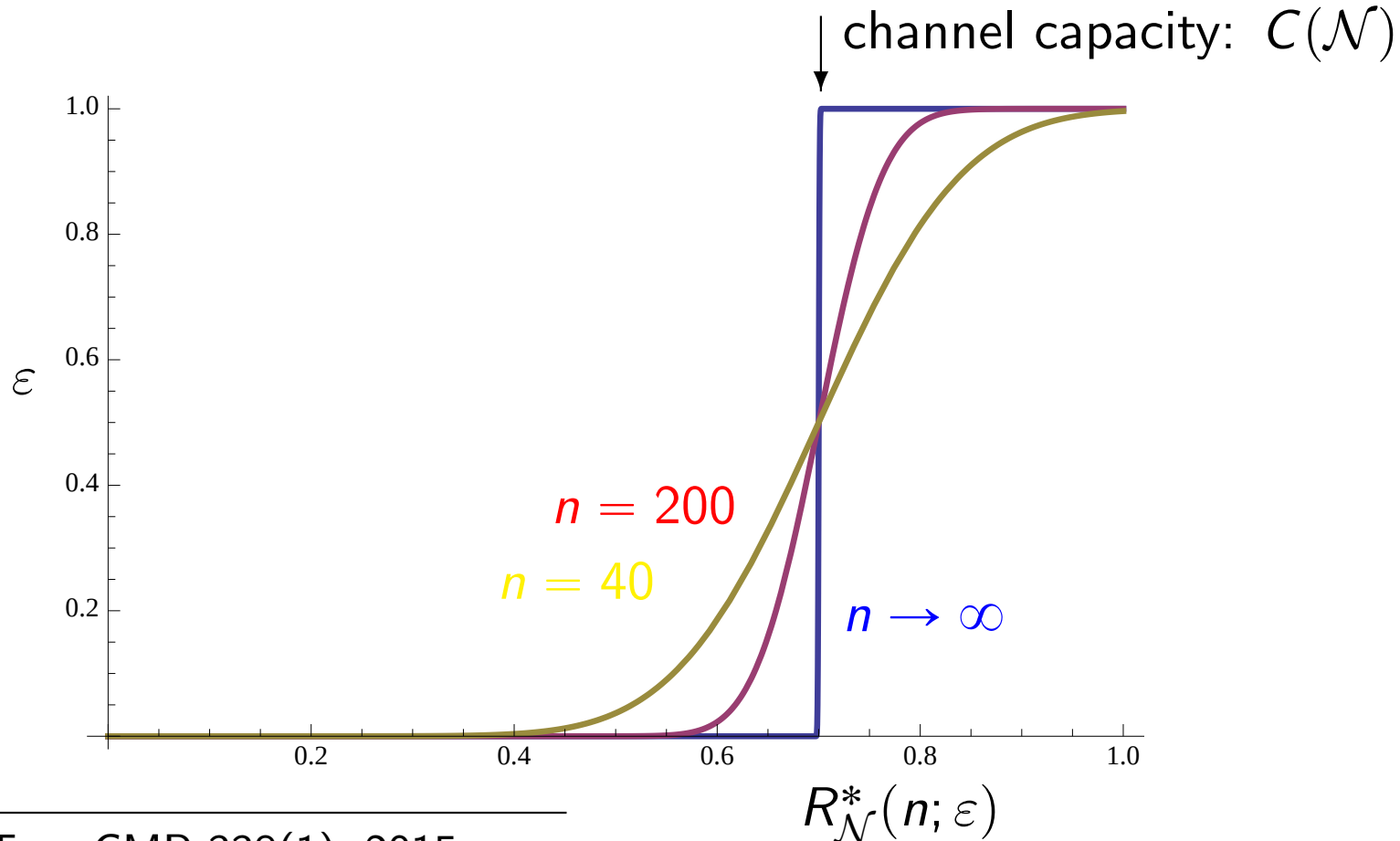
- Many more refined results for entanglement-assisted coding are known, but there are also important open problems (we will get back to that at the end).

[32] Bennett, Shor, Smolin Thapliyal: PRL 83,1999

[33] Bennett et al.: IEEE T-IT 60(5), 2014

Higher-order asymptotics for classical-quantum channels

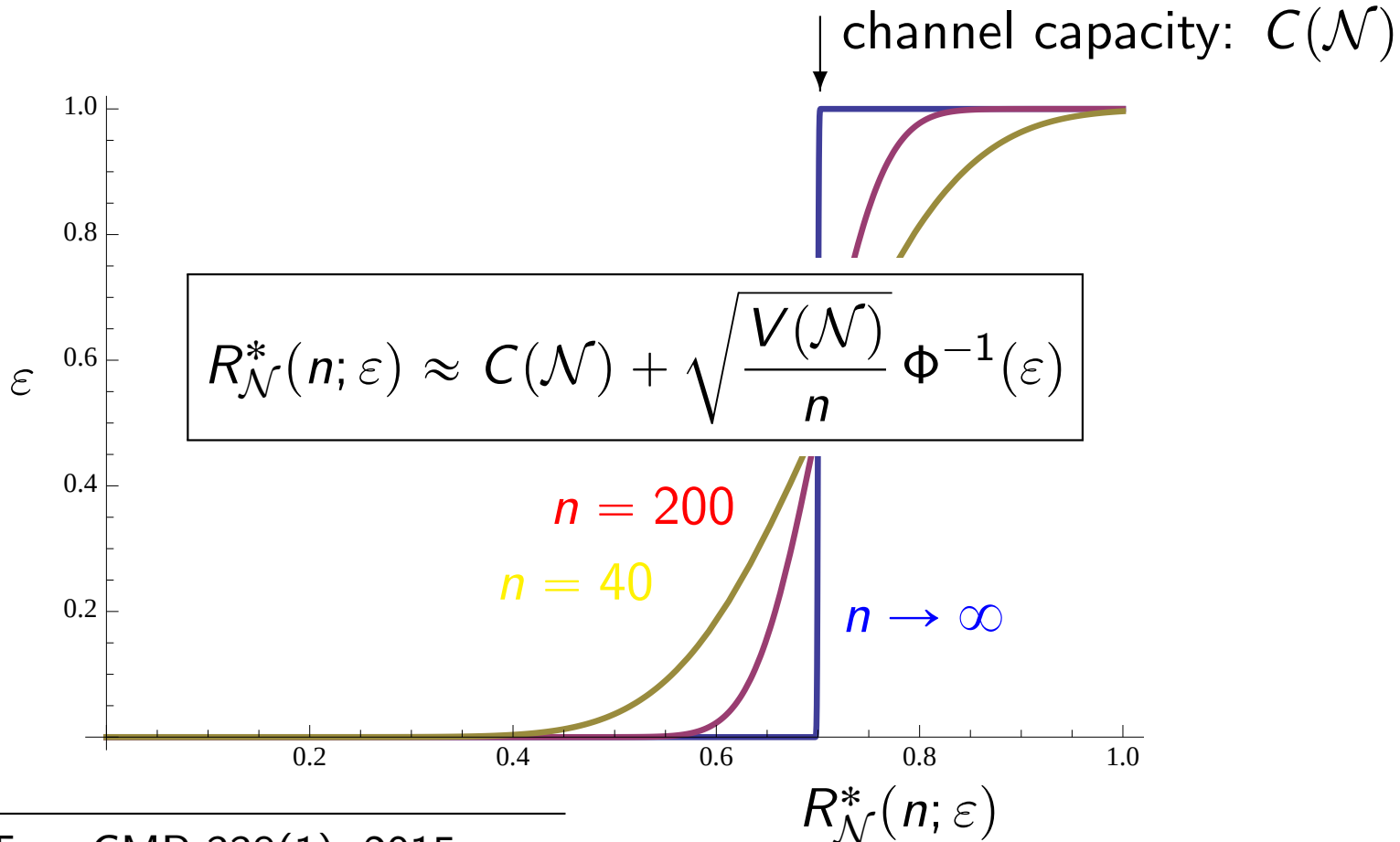
- For channels with only classical inputs the analysis is simpler.
- We have for example a second-order expansion:[34]



[34] T, Tan: CMP 338(1), 2015

Higher-order asymptotics for classical-quantum channels

- For channels with only classical inputs the analysis is simpler.
- We have for example a second-order expansion:[34]



[34] T, Tan: CMP 338(1), 2015

Multi-terminal problems

- There are a lot of open problems still when there is more than one sender and/or receiver.
- For multi-terminal problems we barely understand the asymptotic capacities and very little is known about tight one-shot bounds that would allow us to get corrections terms for them.
- The multiple access channel (MAC) is arguably the simplest multiterminal communication channel where there are several independent senders but only one genuine receiver.[35]

[35] Chakraborty, Sen, Nema: One-shot inner bounds for sending private classical information over a quantum MAC

III.B: Randomness extraction

- Randomness extraction against side information is the art of distilling uniform and independent randomness from a correlated source. It is a central task in cryptography.
- We are given a classical-quantum state

$$\rho_{XB} = \sum_x P(x)|x\rangle\langle x| \otimes \rho_{B,x}$$

- In the one-shot setting the goal is to use a seeded function $f : X \to \{0,1\}^\ell$ such that the resulting state satisfies

$$\mathbb{E} \left\| \rho_{ZB} - \pi_Z \otimes \rho_B \right\|_{\mathsf{tr}} \leqslant \varepsilon$$

for some error ε, where π_Z is the fully mixed state and the expectation is over the seed.

Characterization of the one-shot task

- The goal is to find the optimal tradeoffs between ℓ and ε. (In practice we also want to optimise over the seed length, but we ignore this here.)
- We can define the boundary of the allowed tradeoffs:

$$\ell^*(\varepsilon) := \max\left\{\ell \in \mathbb{N} : \exists f \text{ s.t. } \mathbb{E}\left\|\rho_{ZB} - \pi_Z \otimes \rho_B\right\|_{\mathsf{tr}} \leqslant \varepsilon\right\}$$

- This quantity is usually characterised using the smooth min-entropy:[36]

$$H_{\min}^{\varepsilon-\delta}(X|B) - \log\frac{1}{\delta^4} \leqslant \ell^*(\varepsilon) \leqslant H_{\min}^{2\sqrt{\varepsilon}}(X|B)$$

- Many variations are possible in these one-shot bounds. Hard to define what bounds to consider tight.

[36] Renner: Ph.D. thesis, ETH Zurich, 2005; T, Schaffner, Smith, Renner: IEEE T-IT 57(8), 2011

Memoryless randomness sources

- In cryptography we usually do not assume memoryless sources. Nonetheless, we can study the rate at which randomness can be extracted from such sources.
- Let ℓ_n^ε be as before, but now with n i.i.d. copies of the source.
- The one-shot bounds from the last slide yield

$$\ell_n^*(\varepsilon) = nH(X|B) + O(\sqrt{n})$$

  via the asymptotic equipartition property of the smooth min-entropy. But there is a gap in the second-order term!
- Fresh from the press: For memoryless sources:[37]

$$\ell_n^*(\varepsilon) = nH(X|B) + \sqrt{nV(X|B)}\Phi^{-1}(\varepsilon) + O(\log n)$$

  This result is achieved by avoiding smooth min-entropy and going directly to a hypothesis testing characterisation.

[37] Shen, Gao, Cheng: arXiv:2202.11590, 2022

Error exponents for randomness extraction

- A complementary question fixes first a rate of randomness extraction: $\ell = nR$ and then asks how quickly the error

$$\varepsilon_n = \mathbb{E} \left\| \rho_{ZB^n} - \pi_Z \otimes \rho_B^{\otimes n} \right\|_{\mathrm{tr}}$$

  vanishes when we use an optimal extractor.

- This question has been (almost) resolved for this QIP:[38] [39]

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\varepsilon_n} \geqslant \sup_{\alpha>1} \frac{\alpha-1}{\alpha} \left(\widetilde{H}_\alpha^{\uparrow}(X|B) - R \right)$$

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\varepsilon_n} \leqslant \sup_{\alpha>1} (\alpha-1) \left(\widetilde{H}_\alpha^{\downarrow}(X|B) - R \right)$$

[38] Dupuis: Privacy amplification and decoupling without smoothing

[39] Li, Yao: Reliability Function of Quantum Information Decoupling and Privacy Amplification Via the Sandwiched Renyi Divergence

Some more open questions and conclusion

Some open questions in finite resource QIT

1. One-shot/Rényi conditional mutual information
   - The quantum conditional mutual information is defined as

$$\begin{aligned} I(A:B|C) &= H(A|C) - H(A|BC) \\ &= H(AC) + H(BC) - H(ABC) - H(C) \end{aligned}$$

   - It depends on ρ_{ABC} and the marginals ρ_C, ρ_{AC} and ρ_{BC}.
   - Countless ways of ordering operators for our definitions!
   - Various candidates for Rényi conditional mutual information have been proposed, but some desirable properties could not be shown.[40]
   - Without appropriate definitions for one-shot and Rényi conditional mutual information error exponents and correction terms for more complex quantum information processing tasks like state redistribution appear out of reach.

[40] Berta, Seshadreesan, Wilde: JMP 56(2), 2015.

Some open questions in finite resource QIT

2. Error exponent (sphere packing bound) for classical-quantum channels.
   - For classical channels when the rate $R < C(W)$ is close to capacity, the error exponent is given as

$$\sup_{0<\alpha\leqslant 1} \frac{1-\alpha}{\alpha}\left(\sup_{P_X} I_\alpha^{\downarrow}(X:Y) - R\right)$$

   - Almost all known results translate from classical channels to classical-quantum (cq) channels — except for the error exponent of cq channels, which is generally unknown!
   - For partial results and more related open questions see Hao-Chung's thesis.[41]

[41] Hao-Chung Cheng: Ph.D. thesis, University of Technology Sydney, 2018

Some open questions in finite resource QIT

3. Second-order and moderate deviation analysis for entanglement-assisted communication.
   - The capacity is given by an analogue of Shannon's formula:

$$C_{\mathrm{ea}} = \max_{\psi_{AA'}} I(A:B)$$

   - Strong converse holds and converse for the strong converse exponent is known.[42] We also have a moderate deviation expansion for error probability approaching unity.[43]
   - For small deviations and moderate deviations with vanishing error we only have achievability results.[44] Matching converse bounds are missing!
   - Usual ideas to analyse the converse do not give sufficiently tight bounds, so this seems to require new techniques!

[42] Gupta, Wilde: CMP 334(2), 2013
[43] Ramakrishnan, T, Berta: arXiv:2112.07167, 2021
[44] Datta, T, Wilde: Quant. Inf. Proc. 15, 2016

Conclusions

- We have a pretty good understanding of the structure of quantum information, with some interesting questions still open when it comes to conditional mutual information.
- Even if the open problems in quantum Shannon theory are increasingly difficult and specialised, the techniques required to solve them will often have applications elsewhere too.

- So if you are mathematically inclined, it is still an area where you can make an impact; the tools you derive might be used to solve problems that you never even thought about.