# Entropies

Milán Mosonyi

Mathematical Institute, Budapest University of Technology and Economics

MTA-BME Lendület Quantum Information Theory Research Group

QIP 2019, Boulder, Colorado

## User's guide

- The whole of (quantum) Shannon theory could be presented under this title

# User's guide

- The whole of (quantum) Shannon theory could be presented under this title, but we won't do that.

# User's guide

- The whole of (quantum) Shannon theory could be presented under this title, but we won't do that.

  See previous QIP tutorials by Aram Harrow (2016) and Patrick Hayden (2011), and the textbooks on the next slide.

# User's guide

- The whole of (quantum) Shannon theory could be presented under this title, but we won't do that.

  See previous QIP tutorials by Aram Harrow (2016) and Patrick Hayden (2011), and the textbooks on the next slide.

- Tutorial aimed at students who have been less exposed to entropies (so far).

# User's guide

- The whole of (quantum) Shannon theory could be presented under this title, but we won't do that.

  See previous QIP tutorials by Aram Harrow (2016) and Patrick Hayden (2011), and the textbooks on the next slide.

- Tutorial aimed at students who have been less exposed to entropies (so far).

- Big amount of material - plenty of exercises to help digest it.

# User's guide

- The whole of (quantum) Shannon theory could be presented under this title, but we won't do that.

  See previous QIP tutorials by Aram Harrow (2016) and Patrick Hayden (2011), and the textbooks on the next slide.

- Tutorial aimed at students who have been less exposed to entropies (so far).

- Big amount of material - plenty of exercises to help digest it.

- Completely incomplete list of references, not necessarily to the original works - see the textbooks on the next slide for more detailed references.

# Recommended reading

Masahito Hayashi: *Quantum Information Theory: Mathematical Foundation*, 2nd ed., Springer, 2017.

Dénes Petz: *Quantum Information Theory and Quantum Statistics*, Springer, 2008.

Marco Tomamichel: *Quantum Information Processing with Finite Resources*, SpringerBriefs in Mathematical Physics, 2016.

Mark M. Wilde: *Quantum Information Theory*, 2nd ed., Cambridge University Press, 2017.

# I. Prelude: Shannon entropy

# Shannon entropy

- Problem: How to quantify the information gained by learning the outcome of a random event?

$$x \in \mathcal{X} \text{ happens with probability } P(x)$$

# Shannon entropy

- Problem: How to quantify the information gained by learning the outcome of a random event?

$$x \in \mathcal{X} \ \text{ happens with probability } \ P(x)$$

- Define the surprisal of the outcome $x$ as $\ f(P(x))$, where $f : [0,1] \to \mathbb{R}_+$ is monotone decreasing, $f(1) = 0$.

# Shannon entropy

- Problem: How to quantify the information gained by learning the outcome of a random event?

$$x \in \mathcal{X} \ \text{ happens with probability } \ P(x)$$

- Define the surprisal of the outcome $x$ as $f(P(x))$, where $f : [0,1] \to \mathbb{R}_+$ is monotone decreasing, $f(1) = 0$.

  Average surprisal: $\sum_{x \in \mathcal{X}} P(x) f(P(x))$.

  Dual interpretation: uncertainty about the outcome before it happens.

# Shannon entropy

- Problem: How to quantify the information gained by learning the outcome of a random event?

$$x \in \mathcal{X} \ \text{ happens with probability } \ P(x)$$

- Define the surprisal of the outcome $x$ as $f(P(x))$, where $f : [0, 1] \to \mathbb{R}_+$ is monotone decreasing, $f(1) = 0$.

  Average surprisal: $\sum_{x\in\mathcal{X}} P(x) f(P(x))$.

  Dual interpretation: uncertainty about the outcome before it happens.

- Shannon: $f := -\log$

$$H(P) := -\sum_{x\in\mathcal{X}} P(x) \log P(x) \qquad \text{Shannon entropy of } P.$$

- Why should we prefer this choice over others?

# Classical source coding

- Assume we want to efficiently store the outcomes of many independent events with the same probability distribution $P$.

$$f_n : \mathcal{X}^n \to \{0,1\}^{k_n} \quad \text{encoding} \qquad g_n : \{0,1\}^{k_n} \to \mathcal{X}^n \quad \text{decoding}$$

- block coding, we allow a small error probability $\varepsilon$ in the recovery

# Classical source coding

- Assume we want to efficiently store the outcomes of many independent events with the same probability distribution $P$.

$$f_n : \mathcal{X}^n \to \{0,1\}^{k_n} \text{ encoding} \qquad g_n : \{0,1\}^{k_n} \to \mathcal{X}^n \text{ decoding}$$

- block coding, we allow a small error probability $\varepsilon$ in the recovery
- Theorem 1. ([CsK, Theorem 1.1]) $\forall \varepsilon \in (0,1)$

$$\lim_{n\to+\infty} \frac{1}{n} \min \left\{ k_n : P^{\otimes n}\left(\{\underline{x} \in \mathcal{X}^n : g_n(f_n(\underline{x})) \neq \underline{x}\}\right) \leq \varepsilon \right\} = H(P).$$

Shannon's fixed-length source coding theorem

# Classical source coding

- Assume we want to efficiently store the outcomes of many independent events with the same probability distribution $P$.

$$f_n : \mathcal{X}^n \to \{0,1\}^{k_n} \quad \text{encoding} \qquad g_n : \{0,1\}^{k_n} \to \mathcal{X}^n \quad \text{decoding}$$

- block coding, we allow a small error probability $\varepsilon$ in the recovery
- Theorem 1. ([CsK, Theorem 1.1]) $\forall \varepsilon \in (0,1)$

$$\lim_{n\to+\infty} \frac{1}{n} \min \left\{ k_n : P^{\otimes n} \left( \{ \underline{x} \in \mathcal{X}^n : g_n(f_n(\underline{x})) \neq \underline{x} \} \right) \leq \varepsilon \right\} = H(P).$$

Shannon's fixed-length source coding theorem

- Operational interpretation of the Shannon entropy

# Classical source coding

- Assume we want to efficiently store the outcomes of many independent events with the same probability distribution $P$.

$$f_n : \mathcal{X}^n \to \{0,1\}^{k_n} \quad \text{encoding} \qquad g_n : \{0,1\}^{k_n} \to \mathcal{X}^n \quad \text{decoding}$$

- block coding, we allow a small error probability $\varepsilon$ in the recovery
- Theorem 1. ([CsK, Theorem 1.1]) $\forall \varepsilon \in (0,1)$

$$\lim_{n\to+\infty} \frac{1}{n} \min\left\{k_n : P^{\otimes n}\left(\{\underline{x} \in \mathcal{X}^n : g_n(f_n(\underline{x})) \neq \underline{x}\}\right) \leq \varepsilon\right\} = H(P).$$

  Shannon's fixed-length source coding theorem

- Operational interpretation of the Shannon entropy

  as the minimum number of bits/outcome needed to reliably store the outcomes in the asymptotics.

# Classical source coding

- Assume we want to efficiently store the outcomes of many independent events with the same probability distribution $P$.

$$f_n : \mathcal{X}^n \to \{0,1\}^{k_n} \quad \text{encoding} \qquad g_n : \{0,1\}^{k_n} \to \mathcal{X}^n \quad \text{decoding}$$

- block coding, we allow a small error probability $\varepsilon$ in the recovery

- Theorem 1. ([CsK, Theorem 1.1]) $\forall \varepsilon \in (0,1)$

$$\lim_{n\to+\infty} \frac{1}{n} \min\left\{k_n :\ P^{\otimes n}\left(\{\underline{x} \in \mathcal{X}^n :\ g_n(f_n(\underline{x})) \neq \underline{x}\}\right) \leq \varepsilon\right\} = H(P).$$

Shannon's fixed-length source coding theorem

- Operational interpretation of the Shannon entropy

# Classical source coding

- Assume we want to efficiently store the outcomes of many independent events with the same probability distribution $P$.

$$f_n : \mathcal{X}^n \to \{0,1\}^{k_n} \quad \text{encoding} \qquad g_n : \{0,1\}^{k_n} \to \mathcal{X}^n \quad \text{decoding}$$

- block coding, we allow a small error probability $\varepsilon$ in the recovery
- Theorem 1. ([CsK, Theorem 1.1]) $\forall \varepsilon \in (0,1)$

$$\lim_{n\to+\infty} \frac{1}{n} \min\left\{k_n : P^{\otimes n}\left(\{\underline{x} \in \mathcal{X}^n : g_n(f_n(\underline{x})) \neq \underline{x}\}\right) \leq \varepsilon\right\} = H(P).$$

  Shannon's fixed-length source coding theorem

- Operational interpretation of the Shannon entropy

  minimum number of bits needed to achieve $\varepsilon$ error: operational quantity

# Classical source coding

- Assume we want to efficiently store the outcomes of many independent events with the same probability distribution $P$.

$f_n : \mathcal{X}^n \to \{0,1\}^{k_n}$ encoding $\qquad g_n : \{0,1\}^{k_n} \to \mathcal{X}^n$ decoding

- block coding, we allow a small error probability $\varepsilon$ in the recovery

- Theorem 1. ([CsK, Theorem 1.1]) $\forall \varepsilon \in (0,1)$

$$\lim_{n\to+\infty} \frac{1}{n} \min\left\{k_n : P^{\otimes n}\left(\{\underline{x} \in \mathcal{X}^n : g_n(f_n(\underline{x})) \neq \underline{x}\}\right) \leq \varepsilon\right\} = H(P).$$

  Shannon's fixed-length source coding theorem

- Operational interpretation of the Shannon entropy

  minimum number of bits needed to achieve $\varepsilon$ error: operational quantity

  $H(P)$: entropic quantity, or information measure

# General picture

- Information theoretic problems have an operational side and an entropic side, and coding theorems connect the two

# General picture

- Information theoretic problems have an operational side and an entropic side, and coding theorems connect the two, giving an operational justification to use the entropic quantities as information measures.

# General picture

- Information theoretic problems have an operational side and an entropic side, and coding theorems connect the two, giving an operational justification to use the entropic quantities as information measures.

- Are there operationally relevant measures of information, other than the Shannon entropy?

# General picture

- Information theoretic problems have an operational side and an entropic side, and coding theorems connect the two, giving an operational justification to use the entropic quantities as information measures.

- Are there operationally relevant measures of information, other than the Shannon entropy? Yes.

# General picture

- Information theoretic problems have an operational side and an entropic side, and coding theorems connect the two, giving an operational justification to use the entropic quantities as information measures.

- Are there operationally relevant measures of information, other than the Shannon entropy?   Yes.

- Source coding with side information: $H(A|B)_P = H(P_{AB}) - H(P_B)$ conditional entropy.

# General picture

- Information theoretic problems have an operational side and an entropic side, and coding theorems connect the two, giving an operational justification to use the entropic quantities as information measures.

- Are there operationally relevant measures of information, other than the Shannon entropy? Yes.

- Source coding with side information: $H(A|B)_P = H(P_{AB}) - H(P_B)$ conditional entropy.

  Channel coding: $\sup_P\{H(P_A) + H(P_B) - H(P_{AB})\}$ mutual information.

# General picture

- Information theoretic problems have an operational side and an entropic side, and coding theorems connect the two, giving an operational justification to use the entropic quantities as information measures.

- Are there operationally relevant measures of information, other than the Shannon entropy? Yes.

- Source coding with side information: $H(A|B)_P = H(P_{AB}) - H(P_B)$ conditional entropy.

  Channel coding: $\sup_P \{H(P_A) + H(P_B) - H(P_{AB})\}$ mutual information.

- Error exponents: Rényi entropies.

# General picture

- Information theoretic problems have an operational side and an entropic side, and coding theorems connect the two, giving an operational justification to use the entropic quantities as information measures.

- Are there operationally relevant measures of information, other than the Shannon entropy? Yes.

- Source coding with side information: $H(A|B)_P = H(P_{AB}) - H(P_B)$ conditional entropy.

  Channel coding: $\sup_P\{H(P_A) + H(P_B) - H(P_{AB})\}$ mutual information.

- Error exponents: Rényi entropies.

- Aim of this talk: give an introduction to the most relevant ones

# General picture

- Information theoretic problems have an operational side and an entropic side, and coding theorems connect the two, giving an operational justification to use the entropic quantities as information measures.

- Are there operationally relevant measures of information, other than the Shannon entropy? Yes.

- Source coding with side information: $H(A|B)_P = H(P_{AB}) - H(P_B)$ conditional entropy.

  Channel coding: $\sup_P\{H(P_A) + H(P_B) - H(P_{AB})\}$ mutual information.

- Error exponents: Rényi entropies.

- Aim of this talk: give an introduction to the most relevant ones in the quantum case.

# Plan

- First part: Derive information measures from distance-like quantities on quantum states (divergences).

  Explore general properties and a few simple examples.

# Plan

- First part: Derive information measures from distance-like quantities on quantum states (divergences).

  Explore general properties and a few simple examples.

- Second part: Focus on Rényi information measures and applications.

# II. General divergences and simple examples

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

- Notations:

$\mathcal{B}(\mathcal{H})$: linear operators on $\mathcal{H}$

$\mathcal{B}(\mathcal{H})_{\mathrm{sa}}$: self-adjoint linear operators on $\mathcal{H}$

$\mathcal{B}(\mathcal{H})_+$: non-zero positive semidefinite (PSD) linear operators on $\mathcal{H}$

$\mathcal{S}(\mathcal{H})$: density operators (=states) on $\mathcal{H}$

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

- Examples:

$$\Delta(\varrho\|\sigma) := \frac{1}{2}\|\varrho - \sigma\|_1 \quad \text{trace norm distance}$$

$$\Delta(\varrho\|\sigma) := F(\varrho\|\sigma) := \left\|\sqrt{\varrho}\sqrt{\sigma}\right\|_1 = \operatorname{Tr}\sqrt{\sqrt{\varrho}\sigma\sqrt{\varrho}} \quad \text{fidelity}$$

$$\Delta(\varrho\|\sigma) := D^*_{1/2}(\varrho\|\sigma) := -2\log F(\varrho\|\sigma) + 2\log \operatorname{Tr}\varrho$$

sandwiched $1/2$-Rényi divergence

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

- Definition 2. ([Renner05, Datta09])

Max-relative entropy of $\varrho$ w.r.t. $\sigma$,
or sandwiched Rényi divergence with parameter $\infty$:

$$D_\infty^*(\varrho\|\sigma) := \log\inf\{\lambda > 0 : \varrho \le \lambda\sigma\} = \begin{cases} \log\left\|\sigma^{-1/2}\varrho\sigma^{-1/2}\right\|, & \varrho^0 \le \sigma^0 \\ +\infty, & \text{o.w.} \end{cases}$$

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

- Definition 2. ([Renner05, Datta09])

Max-relative entropy of $\varrho$ w.r.t. $\sigma$,
or sandwiched Rényi divergence with parameter $\infty$:

$$D_\infty^*(\varrho\|\sigma) := \log\inf\{\lambda > 0 : \varrho \le \lambda\sigma\} = \begin{cases} \log\left\|\sigma^{-1/2}\varrho\sigma^{-1/2}\right\|, & \varrho^0 \le \sigma^0 \\ +\infty, & \text{o.w.} \end{cases}$$

- Convention:

$$0 \le \varrho = \sum_{i=1}^d r_i |e_i\rangle\langle e_i| \implies \varrho^t := \sum_{i:\, r_i>0} r_i^t |e_i\rangle\langle e_i|, \;\; t \in \mathbb{R}$$

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

- Definition 2. ([Renner05, Datta09])

Max-relative entropy of $\varrho$ w.r.t. $\sigma$,
or sandwiched Rényi divergence with parameter $\infty$:

$$D_\infty^*(\varrho\|\sigma) := \log\inf\{\lambda > 0 : \varrho \le \lambda\sigma\} = \begin{cases} \log\left\|\sigma^{-1/2}\varrho\sigma^{-1/2}\right\|, & \varrho^0 \le \sigma^0 \\ +\infty, & \text{o.w.} \end{cases}$$

- Convention:

$$0 \le \varrho = \sum_{i=1}^d r_i|e_i\rangle\langle e_i| \quad \Longrightarrow \quad \varrho^t := \sum_{i:\,r_i>0} r_i^t|e_i\rangle\langle e_i|, \;\; t \in \mathbb{R}$$

$\varrho^{-1} = \sum_{i:\,r_i>0}(1/r_i)|e_i\rangle\langle e_i|$ generalized inverse

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

- Definition 2. ([Renner05, Datta09])

Max-relative entropy of $\varrho$ w.r.t. $\sigma$,
or sandwiched Rényi divergence with parameter $\infty$:

$$D_\infty^*(\varrho\|\sigma) := \log\inf\{\lambda > 0 : \varrho \le \lambda\sigma\} = \begin{cases} \log\left\|\sigma^{-1/2}\varrho\sigma^{-1/2}\right\|, & \varrho^0 \le \sigma^0 \\ +\infty, & \text{o.w.} \end{cases}$$

- Convention:

$$0 \le \varrho = \textstyle\sum_{i=1}^d r_i |e_i\rangle\langle e_i| \quad \Longrightarrow \quad \varrho^t := \textstyle\sum_{i:\, r_i>0} r_i^t |e_i\rangle\langle e_i|, \quad t \in \mathbb{R}$$

$\varrho^{-1} = \sum_{i:\, r_i>0}(1/r_i)|e_i\rangle\langle e_i|$ generalized inverse

$\varrho^0 = \lim_{\alpha\searrow 0}\varrho^\alpha$ projection onto $\operatorname{supp}\varrho := \operatorname{span}\{e_i : r_i > 0\}$.

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

# General quantum divergences

- Most general definition of a quantum divergence:

$$\Delta : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$$

defined on every finite-dimensional Hilbert space $\mathcal{H}$.

- Definition 3. ([Umegaki62])

(Umegaki's) relative entropy of $\varrho$ w.r.t. $\sigma$:

$$D_1(\varrho\|\sigma) := \begin{cases} \frac{1}{\operatorname{Tr}\varrho} \operatorname{Tr}\varrho(\log\varrho - \log\sigma), & \varrho^0 \le \sigma^0, \\ +\infty, & \text{o.w.} \end{cases}$$

($\log$ defined only on the support)

# General quantum divergences

Definition 4. A quantum divergence $\Delta$ is

- positive, if $\Delta(\varrho\|\sigma) \geq 0 \quad \forall \varrho, \sigma \in \mathcal{S}(\mathcal{H})$, and $\Delta(\varrho\|\varrho) = 0$,

  and strictly positive, if, moreover, $\Delta(\varrho\|\sigma) = 0 \Longrightarrow \varrho = \sigma$;

# General quantum divergences

Definition 4. A quantum divergence $\Delta$ is

- positive, if $\Delta(\varrho\|\sigma) \geq 0 \quad \forall \varrho, \sigma \in \mathcal{S}(\mathcal{H})$, and $\Delta(\varrho\|\varrho) = 0$,

  and strictly positive, if, moreover, $\Delta(\varrho\|\sigma) = 0 \Longrightarrow \varrho = \sigma$;
- additive, if $\Delta(\varrho^{\otimes n}\|\sigma^{\otimes n}) = n\,\Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\, n \in \mathbb{N}$;

# General quantum divergences

Definition 4. A quantum divergence $\Delta$ is

- positive, if $\Delta(\varrho\|\sigma) \geq 0 \quad \forall \varrho, \sigma \in \mathcal{S}(\mathcal{H})$, and $\Delta(\varrho\|\varrho) = 0$, and strictly positive, if, moreover, $\Delta(\varrho\|\sigma) = 0 \Longrightarrow \varrho = \sigma$;
- additive, if $\Delta(\varrho^{\otimes n}\|\sigma^{\otimes n}) = n\,\Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\ n \in \mathbb{N}$;
- stable, if $\Delta(\varrho \otimes \omega\|\sigma \otimes \omega) = \Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})+,\ \omega \in \mathcal{B}(\mathcal{K})_+$;

# General quantum divergences

Definition 4. A quantum divergence $\Delta$ is

- positive, if $\Delta(\varrho\|\sigma) \geq 0 \quad \forall \varrho, \sigma \in \mathcal{S}(\mathcal{H})$, and $\Delta(\varrho\|\varrho) = 0$,

  and strictly positive, if, moreover, $\Delta(\varrho\|\sigma) = 0 \Longrightarrow \varrho = \sigma$;
- additive, if $\Delta(\varrho^{\otimes n}\|\sigma^{\otimes n}) = n\,\Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\ n \in \mathbb{N}$;
- stable, if $\Delta(\varrho \otimes \omega\|\sigma \otimes \omega) = \Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})+,\ \omega \in \mathcal{B}(\mathcal{K})_+$;
- monotone, if

$$\Delta(\mathcal{E}(\varrho)\|\,\mathcal{E}(\sigma)) \leq \Delta(\varrho\|\sigma), \qquad \forall \ \ \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+, \quad \mathcal{E} \ \text{ CPTP},$$

**Definition 4.** A quantum divergence $\Delta$ is

- positive, if $\Delta(\varrho\|\sigma) \geq 0 \quad \forall \varrho, \sigma \in \mathcal{S}(\mathcal{H})$, and $\Delta(\varrho\|\varrho) = 0$,

  and strictly positive, if, moreover, $\Delta(\varrho\|\sigma) = 0 \Longrightarrow \varrho = \sigma$;
- additive, if $\Delta(\varrho^{\otimes n}\|\sigma^{\otimes n}) = n\,\Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\ n \in \mathbb{N}$;
- stable, if $\Delta(\varrho \otimes \omega\|\sigma \otimes \omega) = \Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})+,\ \omega \in \mathcal{B}(\mathcal{K})_+$;
- monotone, if

$$\Delta(\mathcal{E}(\varrho)\|\,\mathcal{E}(\sigma)) \leq \Delta(\varrho\|\sigma), \qquad \forall \ \ \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+, \quad \mathcal{E} \ \text{CPTP},$$

- and $\Delta$ satisfies the logarithmic scaling property if

$$\Delta(\lambda\varrho\|\eta\sigma) = \Delta(\varrho\|\sigma) + \log\lambda - \log\eta, \qquad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\ \lambda, \eta > 0.$$

# General quantum divergences

**Definition 4.** A quantum divergence $\Delta$ is

- positive, if $\Delta(\varrho\|\sigma) \geq 0 \quad \forall \varrho, \sigma \in \mathcal{S}(\mathcal{H})$, and $\Delta(\varrho\|\varrho) = 0$,

  and strictly positive, if, moreover, $\Delta(\varrho\|\sigma) = 0 \Longrightarrow \varrho = \sigma$;
- additive, if $\Delta(\varrho^{\otimes n}\|\sigma^{\otimes n}) = n\,\Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\ n \in \mathbb{N}$;
- stable, if $\Delta(\varrho \otimes \omega\|\sigma \otimes \omega) = \Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})+,\ \omega \in \mathcal{B}(\mathcal{K})_+$;
- monotone, if

$$\Delta(\mathcal{E}(\varrho)\|\,\mathcal{E}(\sigma)) \leq \Delta(\varrho\|\sigma), \qquad \forall \quad \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+, \quad \mathcal{E} \ \text{CPTP},$$

- and $\Delta$ satisfies the logarithmic scaling property if

$$\Delta(\lambda\varrho\|\eta\sigma) = \Delta(\varrho\|\sigma) + \log\lambda - \log\eta, \qquad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\ \lambda, \eta > 0.$$

**Exercise 5.** $D^*_{1/2}$, $D_1$, and $D^*_\infty$ satisfy all the above, and they are monotone even under PTP maps.

*Solution:* Easy, except for the monotonicity of $D^*_{1/2}$ ([NC, Section 9]), and positivity and monotonicity of $D_1$ (later).

# General quantum divergences

**Definition 4.** A quantum divergence $\Delta$ is

- positive, if $\Delta(\varrho\|\sigma) \geq 0 \quad \forall \varrho, \sigma \in \mathcal{S}(\mathcal{H})$, and $\Delta(\varrho\|\varrho) = 0$,

  and strictly positive, if, moreover, $\Delta(\varrho\|\sigma) = 0 \Longrightarrow \varrho = \sigma$;
- additive, if $\Delta(\varrho^{\otimes n}\|\sigma^{\otimes n}) = n\,\Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\ n \in \mathbb{N}$;
- stable, if $\Delta(\varrho \otimes \omega\|\sigma \otimes \omega) = \Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})+,\ \omega \in \mathcal{B}(\mathcal{K})_+$;
- monotone, if

  $$\Delta(\mathcal{E}(\varrho)\|\,\mathcal{E}(\sigma)) \leq \Delta(\varrho\|\sigma), \qquad \forall \ \ \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+, \quad \mathcal{E} \ \text{CPTP},$$
- and $\Delta$ satisfies the logarithmic scaling property if

  $$\Delta(\lambda\varrho\|\eta\sigma) = \Delta(\varrho\|\sigma) + \log\lambda - \log\eta, \qquad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\ \lambda, \eta > 0.$$

**Exercise 6.** Monotonicity implies stability and isometric invariance: $\Delta(V\varrho V^*\|V\sigma V^*) = \Delta(\varrho\|\sigma), \quad \forall \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+,\ V : \mathcal{H} \to \mathcal{K}$ isometry. *Solution:* Easy.

# Monotonicity and convexity

Definition 7. A quantum divergence $\Delta$ is jointly convex if

$$\Delta\left(\sum_i p_i\varrho_i \| \sum_i p_i\sigma_i\right) \leq \sum_i p_i\,\Delta(\varrho_i\|\sigma_i), \qquad \sum_i p_i = 1.$$

# Monotonicity and convexity

**Definition 7.** A quantum divergence $\Delta$ is jointly convex if

$$\Delta\left(\sum_i p_i\varrho_i \| \sum_i p_i\sigma_i\right) \leq \sum_i p_i \,\Delta(\varrho_i\|\sigma_i), \qquad \sum_i p_i = 1.$$

**Theorem 8.** (Uhlmann) If $\Delta$ is jointly convex and invariant under isometries then it is monotone.

*Proof:* $\mathcal{E}_{A\to B}(.) = \mathrm{Tr}_E\, V(.)V^*$ Stinespring dilation,

$$\mathrm{Tr}_E(.) = \frac{1}{d_B^2}\sum_{k,l=0}^{d_B-1}(I_B\otimes W_{k,l})(.)(I_B\otimes W_{k,l})^*$$

$$W_{k,l} := X^k Z^l, \quad X = \sum_{a=0}^{d_B-1}|a+1\rangle\langle a|, \quad Z = \sum_{a=0}^{d_B-1} e^{i\frac{2\pi}{d_B}a}|a\rangle\langle a|$$

discrete Weyl unitaries.

**Definition 9.** A quantum divergence $\Delta$

- has the direct sum property if

$$\Delta\left(\sum_i \varrho_i \| \sum_i \sigma_i\right) = \sum_i \Delta\left(\varrho_i \| \sigma_i\right), \qquad \text{when} \quad \varrho_i^0 \vee \sigma_i^0 \perp_{i \neq j} \varrho_j^0 \vee \sigma_j^0.$$

- is homogeneous if $\lambda > 0$.

$$\Delta(\lambda\varrho\|\lambda\sigma) = \lambda\,\Delta(\varrho\|\sigma), \qquad \lambda > 0.$$

# Monotonicity and convexity

Definition 9. A quantum divergence $\Delta$

- has the direct sum property if

$$\Delta\left(\sum_i \varrho_i \| \sum_i \sigma_i\right) = \sum_i \Delta\left(\varrho_i \| \sigma_i\right), \qquad \text{when} \;\; \varrho_i^0 \vee \sigma_i^0 \perp_{i \neq j} \varrho_j^0 \vee \sigma_j^0.$$

- is homogeneous if $\lambda > 0$.

$$\Delta(\lambda\varrho \| \lambda\sigma) = \lambda\, \Delta(\varrho \| \sigma), \qquad \lambda > 0.$$

Theorem 10. (Petz) If $\Delta$ is monotone and homogeneous, and has the direct sum property, then it is jointly convex.

*Proof:* Apply monotonicity to

$$\varrho = \sum_i p_i |i\rangle\langle i| \otimes \varrho_i, \qquad \sigma = \sum_i p_i |i\rangle\langle i| \otimes \sigma_i.$$

# III. Information measures from divergences

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

Definition 11. $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log |A| - \Delta\left(\varrho \middle\| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

Definition 11. $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log|A| - \Delta\left(\varrho \| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".
- Interpretation: uncertainty about system $A$ in state $\varrho$.

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

Definition 11. $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log |A| - \Delta\left(\varrho \middle\| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".
- Interpretation: uncertainty about system $A$ in state $\varrho$.
- The first formula works also in infinite dimension.

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

Definition 11. $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log |A| - \Delta\left(\varrho \middle\| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".
- Interpretation: uncertainty about system $A$ in state $\varrho$.
- The first formula works also in infinite dimension.

Exercise 12. For the relative entropy $\Delta = D_1$ we get

$$H_1(A)_\varrho := H_{D_1}(A)_\varrho = -\operatorname{Tr} \varrho \log \varrho.$$

von Neumann entropy Also denoted by $S(\varrho) = S(A)_\varrho$.

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

**Definition 11.** $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log |A| - \Delta\left(\varrho \middle\| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".
- Interpretation: uncertainty about system $A$ in state $\varrho$.
- The first formula works also in infinite dimension.

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

Definition 11. $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log |A| - \Delta\left(\varrho \middle\| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".
- Interpretation: uncertainty about system $A$ in state $\varrho$.
- The first formula works also in infinite dimension.

Exercise 13.

$$H^*_{1/2}(\varrho) := H^*_{1/2}(A)_\varrho := H_{D^*_{1/2}}(A)_\varrho = 2 \log \operatorname{Tr} \sqrt{\varrho} \qquad \text{max-entropy}$$

$$H^*_\infty(\varrho) := H^*_\infty(A)_\varrho := H_{D^*_\infty}(A)_\varrho = -\log \|\varrho\|_\infty \qquad \text{min-entropy}$$

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

Definition 11. $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log |A| - \Delta\left(\varrho \middle\| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".
- The first formula works also in infinite dimension.

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

Definition 11. $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log |A| - \Delta\left(\varrho \middle\| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".
- The first formula works also in infinite dimension.
- If $\Delta$ is positive then $H_\Delta(A)_\varrho \le \log |A|$, equality at $\varrho = \pi_A$.

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

Definition 11. $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log |A| - \Delta\left(\varrho \middle\| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".
- The first formula works also in infinite dimension.
- If $\Delta$ is positive then $H_\Delta(A)_\varrho \leq \log |A|$, equality at $\varrho = \pi_A$.
- If $\Delta$ is additive then so is $H_\Delta$:

$$H_\Delta(AB)_{\varrho_A \otimes \varrho_B} = H_\Delta(A)_{\varrho_A} + H_\Delta(B)_{\varrho_B}.$$

# Entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Notation: $|A| := \dim \mathcal{H}_A$, $\pi_A := \frac{1}{|A|} I_A$

Definition 11. $\Delta$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\Delta(\varrho) := H_\Delta(A)_\varrho := -\Delta(\varrho \| I_A) = \log|A| - \Delta\left(\varrho \middle\| \pi_A\right).$$

- Measures how far the state is from the maximally mixed state in $\Delta$ "distance".
- The first formula works also in infinite dimension.
- If $\Delta$ is positive then $H_\Delta(A)_\varrho \le \log|A|$, equality at $\varrho = \pi_A$.
- If $\Delta$ is additive then so is $H_\Delta$:

$$H_\Delta(AB)_{\varrho_A \otimes \varrho_B} = H_\Delta(A)_{\varrho_A} + H_\Delta(B)_{\varrho_B}.$$

- If $\Delta$ is monotone then $H_\Delta$ is monotone non-decreasing under unital CPTP maps (e.g., rank $1$ projective measurements).

# Monotonicity of entropy and majorization

Definition 12. A state $\varrho$ majorizes another state $\sigma$ ($\varrho \succeq \sigma$), or $\sigma$ is more mixed than $\varrho$, if

$$\sum_{k=1}^{m} \lambda_k^{\downarrow}(\sigma) \leq \sum_{k=1}^{m} \lambda_k^{\downarrow}(\varrho) \qquad \forall m,$$

where $\lambda_1^{\downarrow} \geq \lambda_2^{\downarrow} \geq \dots$ are the decreasingly ordered eigenvalues.

# Monotonicity of entropy and majorization

Definition 12. A state $\varrho$ majorizes another state $\sigma$ ($\varrho \succeq \sigma$), or $\sigma$ is more mixed than $\varrho$, if

$$\sum_{k=1}^{m} \lambda_k^{\downarrow}(\sigma) \leq \sum_{k=1}^{m} \lambda_k^{\downarrow}(\varrho) \qquad \forall m,$$

where $\lambda_1^{\downarrow} \geq \lambda_2^{\downarrow} \geq \ldots$ are the decreasingly ordered eigenvalues.

Exercise 13. $|\psi\rangle\langle\psi| \succeq \varrho \succeq \pi$ for every $\varrho$, i.e., pure states are the least mixed states, and the maximally mixed state is the most.

# Monotonicity of entropy and majorization

Definition 12. A state $\varrho$ majorizes another state $\sigma$ ($\varrho \succeq \sigma$), or $\sigma$ is more mixed than $\varrho$, if

$$\sum_{k=1}^{m} \lambda_k^{\downarrow}(\sigma) \leq \sum_{k=1}^{m} \lambda_k^{\downarrow}(\varrho) \qquad \forall m,$$

where $\lambda_1^{\downarrow} \geq \lambda_2^{\downarrow} \geq \ldots$ are the decreasingly ordered eigenvalues.

Exercise 13. $|\psi\rangle\langle\psi| \succeq \varrho \succeq \pi$ for every $\varrho$, i.e., pure states are the least mixed states, and the maximally mixed state is the most.

Exercise 14. For two states $\varrho, \sigma$, the following are equivalent:

(i). $\varrho \succeq \sigma$

(ii). $\sum_i p_i U_i \varrho U_i^* = \sigma$ for some unitaries $U_i$ and prob. distr. $(p_i)_i$.

(iii). $\mathcal{E}(\varrho) = \sigma$ for some unital CPTP map $\mathcal{E}$

*Solution:* Use the classical characterization of majorization for (i)$\Longrightarrow$(ii) [Hiai10, Proposition 4.1.1].

# Monotonicity of entropy and majorization

Corollary 15. A function $H$ on quantum states is monotone non-decreasing under unital CPTP maps if and only if it is monotone non-increasing w.r.t. majorization (Schur concave).

In particular, it takes its smallest value on pure states and its largest value on the maximally mixed state.

# Monotonicity of entropy and majorization

Corollary 15. A function $H$ on quantum states is monotone non-decreasing under unital CPTP maps if and only if it is monotone non-increasing w.r.t. majorization (Schur concave).

In particular, it takes its smallest value on pure states and its largest value on the maximally mixed state.

Exercise 16. Such a function takes the same value on every pure state on the same Hilbert space.

# Monotonicity of entropy and majorization

Corollary 15. A function $H$ on quantum states is monotone non-decreasing under unital CPTP maps if and only if it is monotone non-increasing w.r.t. majorization (Schur concave).

In particular, it takes its smallest value on pure states and its largest value on the maximally mixed state.

Exercise 16. Such a function takes the same value on every pure state on the same Hilbert space.

- To call $H$ an entropy function, we require Schur concavity with the normalization $H(\text{pure state}) = 0$,
  $H(\text{maximally mixed state}) = \log \dim(\text{Hilbert space})$,

# Monotonicity of entropy and majorization

Corollary 15. A function $H$ on quantum states is monotone non-decreasing under unital CPTP maps if and only if it is monotone non-increasing w.r.t. majorization (Schur concave).

In particular, it takes its smallest value on pure states and its largest value on the maximally mixed state.

Exercise 16. Such a function takes the same value on every pure state on the same Hilbert space.

- To call $H$ an entropy function, we require Schur concavity with the normalization $H(\text{pure state}) = 0$,
  $H(\text{maximally mixed state}) = \log \dim(\text{Hilbert space})$,

  plus additivity.

# Monotonicity of entropy and majorization

Corollary 15. A function $H$ on quantum states is monotone non-decreasing under unital CPTP maps if and only if it is monotone non-increasing w.r.t. majorization (Schur concave).

In particular, it takes its smallest value on pure states and its largest value on the maximally mixed state.

Exercise 16. Such a function takes the same value on every pure state on the same Hilbert space.

- To call $H$ an entropy function, we require Schur concavity with the normalization $H(\text{pure state}) = 0$,
  $H(\text{maximally mixed state}) = \log\dim(\text{Hilbert space})$,

  plus additivity.

Exercise 17. If $H$ is concave and unitarily invariant then it is Schur concave.
Holds for $H_\Delta$ if $\Delta$ is unitarily invariant and convex in its first variable.

# Conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

# Conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

# Conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

- Measures how far the state is from a state where $A$ and $B$ are independent, and $A$ is in the maximally mixed state.

# Conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

- Measures how far the state is from a state where $A$ and $B$ are independent, and $A$ is in the maximally mixed state.
- Interpretation: uncertainty about system $A$ given access to system $B$.

# Conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$
\begin{aligned}
H_\Delta(A|B)_\varrho &:= -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B) \\
&= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).
\end{aligned}
$$

- Measures how far the state is from a state where $A$ and $B$ are independent, and $A$ is in the maximally mixed state.
- Interpretation: uncertainty about system $A$ given access to system $B$.
- Variant:

$$
H_\Delta^\downarrow(A|B)_\varrho := -\Delta(\varrho_{AB} \| I_A \otimes \varrho_B) = \log|A| - \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \varrho_B\right).
$$

# Conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

**Definition 18.** **Conditional $\Delta$-entropy** of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log |A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

# Conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

Exercise 19.

a). If $|B| = 1$ then $H_\Delta(A|B)_\varrho = H_\Delta(A)_\varrho$.

b). If $\Delta$ is additive and positive then $H_\Delta(A|B)_{\varrho_A \otimes \varrho_B} = H_\Delta(A)_{\varrho_A}$.

# Conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

**Definition 18.** **Conditional $\Delta$-entropy** of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

# Conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

Terminology: $H^*_{1/2}(A|B)_\varrho := H_{D^*_{1/2}}(A|B)_\varrho$ (conditional) max-entropy

$H^*_\infty(A|B)_\varrho := H_{D^*_\infty}(A|B)_\varrho$ (conditional) min-entropy

No explicit formulas in general.

# Conditional entropy can be negative

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log |A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

# Conditional entropy can be negative

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

Exercise 20. For the relative entropy $\Delta = D_1$ we get

$$H_1(A|B)_\varrho := H_{D_1}(A|B)_\varrho = H_{D_1}^{\downarrow}(A|B)_\varrho = H_1(AB)_\varrho - H_1(B)_\varrho$$

conditional von Neumann entropy

# Conditional entropy can be negative

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

Exercise 20. For the relative entropy $\Delta = D_1$ we get

$$H_1(A|B)_\varrho := H_{D_1}(A|B)_\varrho = H^{\downarrow}_{D_1}(A|B)_\varrho = H_1(AB)_\varrho - H_1(B)_\varrho$$

conditional von Neumann entropy

It is negative on every pure entangled state.

Operational interpretation: Entanglement cost in state merging [HOW05, HOW07].

# Concavity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

**Definition 18.** Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

# Concavity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

**Definition 18.** **Conditional $\Delta$-entropy** of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

**Exercise 20.** If $f : C_1 \times C_2 \to \overline{\mathbb{R}}$ is jointly convex then $y \mapsto \inf_x f(x,y)$ is convex.

# Concavity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

Exercise 20. If $f : C_1 \times C_2 \to \overline{\mathbb{R}}$ is jointly convex then $y \mapsto \inf_x f(x,y)$ is convex.

Exercise 21. If $\Delta$ is jointly convex then $H_\Delta(.|.)_\varrho$ is concave in $\varrho$, i.e.,

$$H_\Delta(A|B)_{\sum_i p_i \varrho_i} \geq \sum_i p_i H_\Delta(A|B)_{\varrho_i}.$$

# Concavity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log |A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

Exercise 20. If $f : C_1 \times C_2 \to \overline{\mathbb{R}}$ is jointly convex then $y \mapsto \inf_x f(x,y)$ is convex.

Exercise 21. If $\Delta$ is jointly convex then $H_\Delta(.|.)_\varrho$ is concave in $\varrho$, i.e.,

$$H_\Delta(A|B)_{\sum_i p_i \varrho_i} \geq \sum_i p_i H_\Delta(A|B)_{\varrho_i}.$$

Exercise 22. If $\Delta$ is jointly convex, additive and positive then the conditional $H_\Delta$ is non-negative on separable states.

# Concavity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log |A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

Exercise 20. If $f : C_1 \times C_2 \to \overline{\mathbb{R}}$ is jointly convex then $y \mapsto \inf_x f(x, y)$ is convex.

Exercise 21. If $\Delta$ is jointly convex then $H_\Delta(.|.)_\varrho$ is concave in $\varrho$, i.e.,

$$H_\Delta(A|B)_{\sum_i p_i \varrho_i} \geq \sum_i p_i H_\Delta(A|B)_{\varrho_i}.$$

Exercise 22. If $\Delta$ is jointly convex, additive and positive then the conditional $H_\Delta$ is non-negative on separable states.

Negative conditional entropy is really a quantum phenomenon.

# Monotonicity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log |A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta \left( \varrho_{AB} \middle\| \pi_A \otimes \sigma_B \right).$$

# Monotonicity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

**Definition 18.** Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B) \\ = \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

- If $\Delta$ is monotone then so is the conditional $\Delta$-entropy:

  $\mathcal{E} : A \to A'$ unital CPTP, $\quad \mathcal{F} : B \to B'$ CPTP

$$H_\Delta(A'|B')_{\varrho'} \geq H_\Delta(A|B)_\varrho, \qquad \varrho' = (\mathcal{E} \otimes \mathcal{F})\varrho.$$

# Monotonicity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

**Definition 18.** Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

- If $\Delta$ is monotone then so is the conditional $\Delta$-entropy:

  $\mathcal{E} : A \to A'$ unital CPTP, $\quad \mathcal{F} : B \to B'$ CPTP

  $$H_\Delta(A'|B')_{\varrho'} \geq H_\Delta(A|B)_\varrho, \qquad \varrho' = (\mathcal{E} \otimes \mathcal{F})\varrho.$$

  In particular, conditioning reduces uncertainty:

  $$H_\Delta(A)_\varrho \geq H_\Delta(A|B)_\varrho \geq H_\Delta(A|BB')_\varrho$$

# Superadditivity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

# Superadditivity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

**Definition 18.** Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

- If $\Delta$ is additive then the conditional $\Delta$-entropy is superadditive:

$$H_\Delta(A_1A_2|B_1B_2)_{\varrho_{A_1B_1} \otimes \varrho_{A_2B_2}} \geq H_\Delta(A_1|B_1)_{\varrho_{A_1B_1}} + H_\Delta(A_2|B_2)_{\varrho_{A_2B_2}}$$

# Superadditivity of conditional entropy

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 18. Conditional $\Delta$-entropy of a bi-partite state $\varrho = \varrho_{AB}$:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$
$$= \log|A| - \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta\left(\varrho_{AB} \middle\| \pi_A \otimes \sigma_B\right).$$

- If $\Delta$ is additive then the conditional $\Delta$-entropy is superadditive:

$$H_\Delta(A_1 A_2|B_1 B_2)_{\varrho_{A_1 B_1} \otimes \varrho_{A_2 B_2}} \geq H_\Delta(A_1|B_1)_{\varrho_{A_1 B_1}} + H_\Delta(A_2|B_2)_{\varrho_{A_2 B_2}}$$

- Additivity?

# Duality and additivity

Exercise 23. For any tripartite pure state $\varrho_{ABC}$,

$$H_1(A|B)_\varrho = -H_1(A|C)_\varrho.$$

# Duality and additivity

Exercise 23. For any tripartite pure state $\varrho_{ABC}$,

$$H_1(A|B)_\varrho = -H_1(A|C)_\varrho.$$

Definition 24. Two functions $H$ and $\tilde{H}$ on bi-partite states are dual to each other, if $H(\varrho_{AB}) + \tilde{H}(\varrho_{AC}) = 0$ for any tri-partite pure state $\varrho_{ABC}$.

# Duality and additivity

Exercise 23. For any tripartite pure state $\varrho_{ABC}$,

$$H_1(A|B)_\varrho = -H_1(A|C)_\varrho.$$

Definition 24. Two functions $H$ and $\tilde{H}$ on bi-partite states are dual to each other, if $H(\varrho_{AB}) + \tilde{H}(\varrho_{AC}) = 0$ for any tri-partite pure state $\varrho_{ABC}$.

Example 25. The conditional von Neumann entropy is self-dual.

# Duality and additivity

Exercise 23. For any tripartite pure state $\varrho_{ABC}$,

$$H_1(A|B)_\varrho = -H_1(A|C)_\varrho.$$

Definition 24. Two functions $H$ and $\tilde{H}$ on bi-partite states are dual to each other, if $H(\varrho_{AB}) + \tilde{H}(\varrho_{AC}) = 0$ for any tri-partite pure state $\varrho_{ABC}$.

Example 25. The conditional von Neumann entropy is self-dual.

Example 26. The conditional min- and max-entropies are dual to each other. ([KRS09]). Special case of duality for conditional Rényi entropies ([Tomamichel]).

# Duality and additivity

Exercise 23. For any tripartite pure state $\varrho_{ABC}$,

$$H_1(A|B)_\varrho = -H_1(A|C)_\varrho.$$

Definition 24. Two functions $H$ and $\tilde{H}$ on bi-partite states are dual to each other, if $H(\varrho_{AB}) + \tilde{H}(\varrho_{AC}) = 0$ for any tri-partite pure state $\varrho_{ABC}$.

Example 25. The conditional von Neumann entropy is self-dual.

Example 26. The conditional min- and max-entropies are dual to each other. ([KRS09]). Special case of duality for conditional Rényi entropies ([Tomamichel]).

Exercise 27. If $H$ and $\tilde{H}$ are dual to each other, and both are superadditive then both are additive.

# Duality and additivity

Exercise 23. For any tripartite pure state $\varrho_{ABC}$,

$$H_1(A|B)_\varrho = -H_1(A|C)_\varrho.$$

Definition 24. Two functions $H$ and $\tilde{H}$ on bi-partite states are dual to each other, if $H(\varrho_{AB}) + \tilde{H}(\varrho_{AC}) = 0$ for any tri-partite pure state $\varrho_{ABC}$.

Example 25. The conditional von Neumann entropy is self-dual.

Example 26. The conditional min- and max-entropies are dual to each other. ([KRS09]). Special case of duality for conditional Rényi entropies ([Tomamichel]).

Exercise 27. If $H$ and $\tilde{H}$ are dual to each other, and both are superadditive then both are additive.

Example 28. The conditional min- and max-entropies are additive.

# Duality and additivity

Exercise 23. For any tripartite pure state $\varrho_{ABC}$,

$$H_1(A|B)_\varrho = -H_1(A|C)_\varrho.$$

Definition 24. Two functions $H$ and $\tilde{H}$ on bi-partite states are dual to each other, if $H(\varrho_{AB}) + \tilde{H}(\varrho_{AC}) = 0$ for any tri-partite pure state $\varrho_{ABC}$.

Example 25. The conditional von Neumann entropy is self-dual.

Example 26. The conditional min- and max-entropies are dual to each other. ([KRS09]). Special case of duality for conditional Rényi entropies ([Tomamichel]).

Exercise 27. If $H$ and $\tilde{H}$ are dual to each other, and both are superadditive then both are additive.

Example 28. The conditional min- and max-entropies are additive.

Remark: Duality of conditional $H_\Delta$ and $H_{\tilde{\Delta}}$ cannot be formulated on the level of the divergences.

Exercise 29. If $H_\Delta$ and $H_{\tilde{\Delta}}$ are dual, and both $\Delta$ and $\tilde{\Delta}$ monotone then

$$H_\Delta(A)_\varrho \geq H_\Delta(A|B)_\varrho \geq -H_{\tilde{\Delta}}(A)_\varrho.$$

Only depends on the first subsystem.

# Duality and entropy bounds

Exercise 29. If $H_\Delta$ and $H_{\tilde{\Delta}}$ are dual, and both $\Delta$ and $\tilde{\Delta}$ monotone then

$$H_\Delta(A)_\varrho \geq H_\Delta(A|B)_\varrho \geq -H_{\tilde{\Delta}}(A)_\varrho.$$

Only depends on the first subsystem.

Corollary 30. If $\Delta$, $\tilde{\Delta}$ are positive then $H_\Delta(A)_\varrho, H_{\tilde{\Delta}}(A)_\varrho \leq \log|A|$, and

$$\log|A| \geq H_\Delta(A|B)_\varrho \geq -\log|A|$$

independently of the state.

# Duality and entropy bounds

Exercise 29. If $H_\Delta$ and $H_{\tilde{\Delta}}$ are dual, and both $\Delta$ and $\tilde{\Delta}$ monotone then

$$H_\Delta(A)_\varrho \geq H_\Delta(A|B)_\varrho \geq -H_{\tilde{\Delta}}(A)_\varrho.$$

Only depends on the first subsystem.

Corollary 30. If $\Delta$, $\tilde{\Delta}$ are positive then $H_\Delta(A)_\varrho$, $H_{\tilde{\Delta}}(A)_\varrho \leq \log|A|$, and

$$\log|A| \geq H_\Delta(A|B)_\varrho \geq -\log|A|$$

independently of the state.

Application: Alicki-Fannes continuity bound on the conditional von Neumann entropy, continuity of channel capacities. [LS09]

Exercise 31. If $H$ and $\tilde{H}$ are dual and $H$ or $\tilde{H}$ is monotone under partial trace over the second system then

$$H(\varrho_{AB}) + \tilde{H}(\varrho_{AC}) \geq 0$$

for any tri-partite state $\varrho_{ABC}$.

Exercise 31. If $H$ and $\tilde{H}$ are dual and $H$ or $\tilde{H}$ is monotone under partial trace over the second system then

$$H(\varrho_{AB}) + \tilde{H}(\varrho_{AC}) \geq 0$$

for any tri-partite state $\varrho_{ABC}$.

Interpretation: Monogamy of correlations.

E.g. $A$ and $B$ are highly entangled

$\Longrightarrow -H_1(A|B)_\varrho$ large $\Longrightarrow H_1(A|C)_\varrho$ large,

i.e., $A$ is close to decoupled from $C$.

# Duality, monogamy, and uncertainty

Exercise 31. If $H$ and $\tilde{H}$ are dual and $H$ or $\tilde{H}$ is monotone under partial trace over the second system then

$$H(\varrho_{AB}) + \tilde{H}(\varrho_{AC}) \geq 0$$

for any tri-partite state $\varrho_{ABC}$.

Interpretation: Monogamy of correlations.

E.g. $A$ and $B$ are highly entangled

$\Longrightarrow -H_1(A|B)_\varrho$ large $\Longrightarrow H_1(A|C)_\varrho$ large,

i.e., $A$ is close to decoupled from $C$.

Application: General tri-partite uncertainty relations for dual conditional entropies $H_\Delta$ and $H_{\tilde{\Delta}}$ [CCYZ12, CBTW17]:

$$H_\Delta(X|B) + H_{\tilde{\Delta}}(Z|C) \geq \mathrm{const}(\mathbb{X}, \mathbb{Z}),$$

where $\mathbb{X}$, $\mathbb{Z}$ are two measurements on $A$ with output spaces $X$, $Z$.

# Mutual information

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 32. $\Delta$-mutual information in a bi-partite state $\varrho_{AB}$:

$$I_{\Delta}^{\uparrow\downarrow}(A:B)_{\varrho} := \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \varrho_A \otimes \sigma_B)$$

# Mutual information

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 32. $\Delta$-mutual information in a bi-partite state $\varrho_{AB}$:

$$I_{\Delta}^{\uparrow\downarrow}(A:B)_{\varrho} := \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \varrho_A \otimes \sigma_B)$$

- Variants:

$$I_{\Delta}^{\downarrow\uparrow}(A:B)_{\varrho} := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A)} \Delta(\varrho_{AB} \| \sigma_A \otimes \varrho_B)$$

$$I_{\Delta}^{\downarrow\downarrow}(A:B)_{\varrho} := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A), \sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \sigma_A \otimes \sigma_B)$$

$$I_{\Delta}^{\uparrow\uparrow}(A:B)_{\varrho} := \Delta(\varrho_{AB} \| \varrho_A \otimes \varrho_B)$$

## Mutual information

Let $\Delta$ be a positive divergence with logarithmic scaling.

**Definition 32.** $\Delta$-mutual information in a bi-partite state $\varrho_{AB}$:

$$I_\Delta^{\uparrow\downarrow}(A:B)_\varrho := \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \varrho_A \otimes \sigma_B)$$

- Variants: $$I_\Delta^{\downarrow\uparrow}(A:B)_\varrho := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A)} \Delta(\varrho_{AB} \| \sigma_A \otimes \varrho_B)$$
$$I_\Delta^{\downarrow\downarrow}(A:B)_\varrho := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A), \sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \sigma_A \otimes \sigma_B)$$
$$I_\Delta^{\uparrow\uparrow}(A:B)_\varrho := \Delta(\varrho_{AB} \| \varrho_A \otimes \varrho_B)$$

- Measures how far the state is from the set of uncorrelated states.

# Mutual information

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 32. $\Delta$-mutual information in a bi-partite state $\varrho_{AB}$:

$$I_\Delta^{\uparrow\downarrow}(A:B)_\varrho := \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \varrho_A \otimes \sigma_B)$$

- Variants:

$$I_\Delta^{\downarrow\uparrow}(A:B)_\varrho := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A)} \Delta(\varrho_{AB} \| \sigma_A \otimes \varrho_B)$$

$$I_\Delta^{\downarrow\downarrow}(A:B)_\varrho := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A), \sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \sigma_A \otimes \sigma_B)$$

$$I_\Delta^{\uparrow\uparrow}(A:B)_\varrho := \Delta(\varrho_{AB} \| \varrho_A \otimes \varrho_B)$$

- Measures how far the state is from the set of uncorrelated states.

- Exercise 33. For Umegaki's relative entropy these all coincide, and are equal to

$$I_1(A:B)_\varrho := I_{D_1}(A:B)_\varrho = H_1(A)_\varrho + H_1(B)_\varrho - H_1(AB)_\varrho.$$

# Mutual information

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 32. $\Delta$-mutual information in a bi-partite state $\varrho_{AB}$:

$$I_{\Delta}^{\uparrow\downarrow}(A:B)_{\varrho} := \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \varrho_A \otimes \sigma_B)$$

- Variants:

$$I_{\Delta}^{\downarrow\uparrow}(A:B)_{\varrho} := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A)} \Delta(\varrho_{AB} \| \sigma_A \otimes \varrho_B)$$

$$I_{\Delta}^{\downarrow\downarrow}(A:B)_{\varrho} := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A), \sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \sigma_A \otimes \sigma_B)$$

$$I_{\Delta}^{\uparrow\uparrow}(A:B)_{\varrho} := \Delta(\varrho_{AB} \| \varrho_A \otimes \varrho_B)$$

- Measures how far the state is from the set of uncorrelated states.
- Exercise 33. For Umegaki's relative entropy these all coincide, and are equal to

$$I_1(A:B)_{\varrho} := I_{D_1}(A:B)_{\varrho} = H_1(A)_{\varrho} + H_1(B)_{\varrho} - H_1(AB)_{\varrho}.$$

- In general they are different.

# Mutual information

Let $\Delta$ be a positive divergence with logarithmic scaling.

**Definition 32.** **$\Delta$-mutual information** in a bi-partite state $\varrho_{AB}$:

$$I_{\Delta}^{\uparrow\downarrow}(A:B)_{\varrho} := \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \varrho_A \otimes \sigma_B)$$

- Variants:

$$I_{\Delta}^{\downarrow\uparrow}(A:B)_{\varrho} := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A)} \Delta(\varrho_{AB} \| \sigma_A \otimes \varrho_B)$$

$$I_{\Delta}^{\downarrow\downarrow}(A:B)_{\varrho} := \inf_{\sigma_A \in \mathcal{S}(\mathcal{H}_A), \sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| \sigma_A \otimes \sigma_B)$$

$$I_{\Delta}^{\uparrow\uparrow}(A:B)_{\varrho} := \Delta(\varrho_{AB} \| \varrho_A \otimes \varrho_B)$$

- Measures how far the state is from the set of uncorrelated states.

# Mutual information

Let $\Delta$ be a positive divergence with logarithmic scaling.

Definition 32. $\Delta$-mutual information in a bi-partite state $\varrho_{AB}$:

$$I_\Delta^{\uparrow\downarrow}(A:B)_\varrho := \inf_{\sigma_B\in\mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB}\|\varrho_A\otimes\sigma_B)$$

- Variants:
$$I_\Delta^{\downarrow\uparrow}(A:B)_\varrho := \inf_{\sigma_A\in\mathcal{S}(\mathcal{H}_A)} \Delta(\varrho_{AB}\|\sigma_A\otimes\varrho_B)$$
$$I_\Delta^{\downarrow\downarrow}(A:B)_\varrho := \inf_{\sigma_A\in\mathcal{S}(\mathcal{H}_A),\sigma_B\in\mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB}\|\sigma_A\otimes\sigma_B)$$
$$I_\Delta^{\uparrow\uparrow}(A:B)_\varrho := \Delta(\varrho_{AB}\|\varrho_A\otimes\varrho_B)$$

- Measures how far the state is from the set of uncorrelated states.

- It is not completely clear which one is the "right" definition.

# Conditional mutual information

- Von Neumann conditional mutual information:

$$I(A:B|C)_\varrho \quad := \quad H_1(A|C)_\varrho + H_1(B|C)_\varrho - H_1(AB|C)_\varrho$$

# Conditional mutual information

- Von Neumann conditional mutual information:

$$
\begin{aligned}
I(A:B|C)_\varrho &:= H_1(A|C)_\varrho + H_1(B|C)_\varrho - H_1(AB|C)_\varrho \\
&= H_1(A|C)_\varrho - H_1(A|BC)_\varrho
\end{aligned}
$$

# Conditional mutual information

- Von Neumann conditional mutual information:

$$\begin{aligned} I(A:B|C)_\varrho &:= H_1(A|C)_\varrho + H_1(B|C)_\varrho - H_1(AB|C)_\varrho \\ &= H_1(A|C)_\varrho - H_1(A|BC)_\varrho \geq 0 \end{aligned}$$

# Conditional mutual information

- Von Neumann conditional mutual information:

$$\begin{aligned} I(A:B|C)_\varrho &:= H_1(A|C)_\varrho + H_1(B|C)_\varrho - H_1(AB|C)_\varrho \\ &= H_1(A|C)_\varrho - H_1(A|BC)_\varrho \geq 0 \end{aligned}$$

- Non-negativity $\Longleftrightarrow$ strong subadditivity of entropy ([LR73])

$$H_1(ABC)_\varrho + H_1(C)_\varrho \leq H_1(AC)_\varrho + H_1(BC)_\varrho$$

# Conditional mutual information

- Von Neumann conditional mutual information:

$$\begin{aligned} I(A:B|C)_\varrho &:= H_1(A|C)_\varrho + H_1(B|C)_\varrho - H_1(AB|C)_\varrho \\ &= H_1(A|C)_\varrho - H_1(A|BC)_\varrho \geq 0 \end{aligned}$$

- Non-negativity $\iff$ strong subadditivity of entropy ([LR73])

$$H_1(ABC)_\varrho + H_1(C)_\varrho \leq H_1(AC)_\varrho + H_1(BC)_\varrho$$

Equality $\iff$ $\varrho_{ABC}$ is a short quantum Markov chain. [HJPW04]

# Conditional mutual information

- Von Neumann conditional mutual information:

$$\begin{aligned} I(A:B|C)_\varrho &:= H_1(A|C)_\varrho + H_1(B|C)_\varrho - H_1(AB|C)_\varrho \\ &= H_1(A|C)_\varrho - H_1(A|BC)_\varrho \geq 0 \end{aligned}$$

- Non-negativity $\Longleftrightarrow$ strong subadditivity of entropy ([LR73])

$$H_1(ABC)_\varrho + H_1(C)_\varrho \leq H_1(AC)_\varrho + H_1(BC)_\varrho$$

Equality $\Longleftrightarrow$ $\varrho_{ABC}$ is a short quantum Markov chain. [HJPW04]

Approximate equality $\Longleftrightarrow$ $\varrho_{ABC}$ approximate Markov chain. [FR15]

# Conditional mutual information

- Von Neumann conditional mutual information:

$$\begin{aligned} I(A:B|C)_\varrho &:= H_1(A|C)_\varrho + H_1(B|C)_\varrho - H_1(AB|C)_\varrho \\ &= H_1(A|C)_\varrho - H_1(A|BC)_\varrho \geq 0 \end{aligned}$$

- Non-negativity $\iff$ strong subadditivity of entropy ([LR73])

$$H_1(ABC)_\varrho + H_1(C)_\varrho \leq H_1(AC)_\varrho + H_1(BC)_\varrho$$

Equality $\iff$ $\varrho_{ABC}$ is a short quantum Markov chain. [HJPW04]

Approximate equality $\iff$ $\varrho_{ABC}$ approximate Markov chain. [FR15]

- Operational interpretation in quantum state redistribution. [YD09]

# Conditional mutual information

- Von Neumann conditional mutual information:

$$
\begin{aligned}
I(A:B|C)_\varrho &:= H_1(A|C)_\varrho + H_1(B|C)_\varrho - H_1(AB|C)_\varrho \\
&= H_1(A|C)_\varrho - H_1(A|BC)_\varrho \geq 0
\end{aligned}
$$

- Non-negativity $\Longleftrightarrow$ strong subadditivity of entropy ([LR73])

$$H_1(ABC)_\varrho + H_1(C)_\varrho \leq H_1(AC)_\varrho + H_1(BC)_\varrho$$

  Equality $\Longleftrightarrow$ $\varrho_{ABC}$ is a short quantum Markov chain. [HJPW04]

  Approximate equality $\Longleftrightarrow$ $\varrho_{ABC}$ approximate Markov chain. [FR15]

- Operational interpretation in quantum state redistribution. [YD09]

- Not clear how to generalize to divergences other than Umegaki's relative entropy. [BSW15]

# Divergence radius and center

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$.

Definition 33. $\Delta$-radius of $\Sigma$:

$$R_\Delta(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\varrho \in \Sigma} \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $\Delta$-center for $\Sigma$.

# Divergence radius and center

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$.

Definition 33. $\Delta$-radius of $\Sigma$:

$$R_\Delta(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\varrho \in \Sigma} \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $\Delta$-center for $\Sigma$.

- No explicit formula in general.

# Divergence radius and center

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$.

**Definition 33.** $\Delta$-radius of $\Sigma$:

$$R_\Delta(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\varrho \in \Sigma} \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $\Delta$-center for $\Sigma$.

- No explicit formula in general.
- Monotonicity if $\Delta$ is monotone: $R_\Delta(\mathcal{E}(\Sigma)) \leq R_\Delta(\Sigma)$.

# Divergence radius and center

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$.

Definition 33. $\Delta$-radius of $\Sigma$:

$$R_\Delta(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\varrho \in \Sigma} \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $\Delta$-center for $\Sigma$.

- No explicit formula in general.
- Monotonicity if $\Delta$ is monotone: $R_\Delta(\mathcal{E}(\Sigma)) \leq R_\Delta(\Sigma)$.
- Additivity?

# Divergence radius and center

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$.

Definition 33. $\Delta$-radius of $\Sigma$:

$$R_\Delta(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\varrho \in \Sigma} \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $\Delta$-center for $\Sigma$.

- No explicit formula in general.
- Monotonicity if $\Delta$ is monotone: $R_\Delta(\mathcal{E}(\Sigma)) \leq R_\Delta(\Sigma)$.
- Additivity? No general method to establish.

# Divergence radius and center

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$.

Definition 33. $\Delta$-radius of $\Sigma$:

$$R_\Delta(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\varrho \in \Sigma} \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $\Delta$-center for $\Sigma$.

- No explicit formula in general.
- Monotonicity if $\Delta$ is monotone: $R_\Delta(\mathcal{E}(\Sigma)) \leq R_\Delta(\Sigma)$.
- Additivity? No general method to establish.

Exercise 34. If $\Sigma \subseteq \mathcal{S}(\mathcal{H})$ and $\Delta$ is a metric on $\mathcal{S}(\mathcal{H})$ then $R_\Delta(\Sigma)$ is the radius of the smallest ball that can be circumscribed around $\Sigma$, and its center is a $\Delta$-center.

# Weighted divergence radius

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$,

$P \in \mathcal{P}_f(\Sigma)$ be a finitely supported probability disctribution on $\Sigma$.

Definition 35. $P$-weighted $\Delta$-radius of $\Sigma$:

$$R_{\Delta,P}(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sum_{\varrho \in \Sigma} P(\varrho)\, \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $P$-weighted $\Delta$-center for $\Sigma$.

# Weighted divergence radius

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$,

$P \in \mathcal{P}_f(\Sigma)$ be a finitely supported probability disctribution on $\Sigma$.

Definition 35. $P$-weighted $\Delta$-radius of $\Sigma$:

$$R_{\Delta,P}(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sum_{\varrho \in \Sigma} P(\varrho)\, \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $P$-weighted $\Delta$-center for $\Sigma$.

- Monotone if $\Delta$ monotone, no explicit formula, additivity not obvious.

# Weighted divergence radius

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$,

$P \in \mathcal{P}_f(\Sigma)$ be a finitely supported probability disctribution on $\Sigma$.

Definition 35. $P$-weighted $\Delta$-radius of $\Sigma$:

$$R_{\Delta,P}(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sum_{\varrho \in \Sigma} P(\varrho)\, \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $P$-weighted $\Delta$-center for $\Sigma$.

- Monotone if $\Delta$ monotone, no explicit formula, additivity not obvious.

Exercise 36. If $\Delta = D_1$ and $\Sigma \subseteq \mathcal{S}(\mathcal{H})$ finite then $\sum_{\varrho \in \Sigma} P(\varrho)\varrho$ is the unique $D_1$-center, and

$$R_{1,P}(\Sigma) := R_{D_1,P}(\Sigma) \quad = \quad \underbrace{H_1\left(\sum_{\varrho \in \Sigma} P(\varrho)\varrho\right) - \sum_{\varrho \in \Sigma} P(\varrho) H_1(\varrho)}_{\text{Holevo quantity}}$$

Holevo quantity

# Weighted divergence radius

Let $\Delta$ be a positive divergence, $\Sigma \subseteq \mathcal{B}(\mathcal{H})_+$,

$P \in \mathcal{P}_f(\Sigma)$ be a finitely supported probability disctribution on $\Sigma$.

Definition 35. $P$-weighted $\Delta$-radius of $\Sigma$:

$$R_{\Delta,P}(\Sigma) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sum_{\varrho \in \Sigma} P(\varrho)\, \Delta(\varrho \| \sigma)$$

If $\sigma$ attains the infimum above, it is called a $P$-weighted $\Delta$-center for $\Sigma$.

- Monotone if $\Delta$ monotone, no explicit formula, additivity not obvious.

Exercise 36. If $\Delta = D_1$ and $\Sigma \subseteq \mathcal{S}(\mathcal{H})$ finite then $\sum_{\varrho \in \Sigma} P(\varrho)\varrho$ is the unique $D_1$-center, and

$$R_{1,P}(\Sigma) := R_{D_1,P}(\Sigma) = H_1\left(\sum_{\varrho \in \Sigma} P(\varrho)\varrho\right) - \sum_{\varrho \in \Sigma} P(\varrho) H_1(\varrho)$$

Holevo quantity

$$= I_1(A:B)_{\varrho}, \qquad \varrho = \sum_{\varrho \in \operatorname{supp} P} P(\varrho) |\varrho\rangle\langle\varrho| \otimes \varrho.$$

# Divergence radii

Proposition 37. $R_{\Delta,P}(\Sigma) \leq R_{\Delta}(\Sigma)$ for every $P$,

# Divergence radii

Proposition 37. $R_{\Delta,P}(\Sigma) \leq R_{\Delta}(\Sigma)$ for every $P$, and if $\Delta$ is convex and lower semi-continuous in its second argument then

$$\sup_{P \in \mathcal{P}_f(\Sigma)} R_{\Delta,P}(S) = R_{\Delta}(\Sigma).$$

Proposition 37. $R_{\Delta,P}(\Sigma) \le R_\Delta(\Sigma)$ for every $P$, and if $\Delta$ is convex and lower semi-continuous in its second argument then

$$\sup_{P \in \mathcal{P}_f(\Sigma)} R_{\Delta,P}(S) = R_\Delta(\Sigma).$$

*Proof:*

$$R_\Delta(S) = \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\varrho \in \Sigma} \Delta(\varrho \| \sigma) \qquad = \qquad \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{P \in \mathcal{P}_f(\Sigma)} \sum_{\varrho \in \Sigma} P(\varrho)\, \Delta(\varrho \| \sigma)$$

# Divergence radii

Proposition 37. $R_{\Delta,P}(\Sigma) \le R_{\Delta}(\Sigma)$ for every $P$, and if $\Delta$ is convex and lower semi-continuous in its second argument then

$$\sup_{P \in \mathcal{P}_f(\Sigma)} R_{\Delta,P}(S) = R_{\Delta}(\Sigma).$$

*Proof:*

$$R_{\Delta}(S) = \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\varrho \in \Sigma} \Delta(\varrho \| \sigma) \quad = \quad \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{P \in \mathcal{P}_f(\Sigma)} \sum_{\varrho \in \Sigma} P(\varrho)\, \Delta(\varrho \| \sigma)$$

$$\overset{\text{minimax}}{=} \quad \sup_{P \in \mathcal{P}_f(\Sigma)} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sum_{\varrho \in \Sigma} P(\varrho)\, \Delta(\varrho \| \sigma)$$

# Divergence radii

Proposition 37. $R_{\Delta,P}(\Sigma) \le R_\Delta(\Sigma)$ for every $P$, and if $\Delta$ is convex and lower semi-continuous in its second argument then

$$\sup_{P\in\mathcal{P}_f(\Sigma)} R_{\Delta,P}(S) = R_\Delta(\Sigma).$$

*Proof:*

$$R_\Delta(S) = \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_{\varrho\in\Sigma} \Delta(\varrho\|\sigma) = \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_{P\in\mathcal{P}_f(\Sigma)} \sum_{\varrho\in\Sigma} P(\varrho)\,\Delta(\varrho\|\sigma)$$

$$\overset{\text{minimax}}{=} \sup_{P\in\mathcal{P}_f(\Sigma)} \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sum_{\varrho\in\Sigma} P(\varrho)\,\Delta(\varrho\|\sigma)$$

Application: Strong converse properties of the classical capacity of various channel models ([KW09, WWY13, MO17]).

# IV. Single-shot state discrimination and min-entropy

# State discrimination: Multiple, single-shot, Bayesian

- **Problem:** Alice wants to send Bob one of $M$ possible messages using a quantum system.

# State discrimination: Multiple, single-shot, Bayesian

- Problem: Alice wants to send Bob one of $M$ possible messages using a quantum system.

  She encodes message $i \in [M] := \{1, \ldots, M\}$ into some state of the quantum system, and sends it to Bob over a quantum channel; Bob receives the system in state $\varrho_i$.

  Bob performs a POVM $\{M_i\}_{i \in [M]}$ to decide which message was sent.

# State discrimination: Multiple, single-shot, Bayesian

- Problem: Alice wants to send Bob one of $M$ possible messages using a quantum system.

  She encodes message $i \in [M] := \{1, \ldots, M\}$ into some state of the quantum system, and sends it to Bob over a quantum channel; Bob receives the system in state $\varrho_i$.

  Bob performs a POVM $\{M_i\}_{i\in[M]}$ to decide which message was sent.

  If Alice sends the $i$-th message with probability $p_i$ then the optimal Bayesian success probability (guessing probability) is

$$P_s^*(\{\varrho_i, p_i\}) := \sup\left\{\sum_{i=1}^{M} p_i \operatorname{Tr} \varrho_i M_i : \{M_i\}_{i\in[M]} \text{ POVM}\right\}.$$

- Single-shot state discrimination with multiple hypothesis and Bayesian error criterion.

# Classical case

$$P_s^*(\{\varrho_i, p_i\}) = \sup \left\{ \sum_{i=1}^{M} p_i \operatorname{Tr} \varrho_i M_i \, : \, \{M_i\}_{i\in[M]} \text{ POVM} \right\}.$$

# Classical case

$$P_s^*(\{\varrho_i, p_i\}) = \sup\left\{\sum_{i=1}^{M} p_i \operatorname{Tr} \varrho_i M_i \,:\, \{M_i\}_{i\in[M]} \text{ POVM}\right\}.$$

Exercise 38.

Assume that all $\varrho_i$ commute: $\varrho_i = \sum_{x\in\mathcal{X}} \varrho_i(x)|x\rangle\langle x|$. Show that

a). It is enough to consider measurement operators diagonal in the same basis.

b). A measurement is optimal iff it is a maximum likelihood measurement:

$$M_i(x) = 0 \text{ when } p_i\varrho_i(x) < m(x) := \max_{i\in[M]} p_i\varrho_i(x).$$

c). $P_s^*(\{\varrho_i, p_i\}) = \sum_{x\in\mathcal{X}} m(x) = \operatorname{Tr}\max_{i\in[M]}\{p_i\varrho_i\}$,

where the maximum is in the (diagonal) entry-wise sense.

*Solution:* Easy; see, e.g., [AM14, Appendix B].

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \ \ \forall \psi \in \mathcal{H} \qquad \text{Notation: } A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \ \ \forall \psi \in \mathcal{H} \qquad\qquad \text{Notation: } \ A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases.

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \ \ \forall \psi \in \mathcal{H} \qquad \text{Notation: } \ A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases. Even if the operators commute!

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \ \ \forall \psi \in \mathcal{H} \qquad \text{Notation: } A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases. Even if the operators commute!

Exercise 39. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. Show that

$$\{X \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}} : X \leq A, \, X \leq B\} \qquad \text{and}$$
$$\{Y \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}} : A \leq Y, \, B \leq Y\}$$

do not admit a maximal (resp. minimal) element.

*Solution:* Easy; see, e.g., [AM14, Example A.1].

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \quad \forall \psi \in \mathcal{H} \qquad\qquad \text{Notation: } A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases. Even if the operators commute!

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \quad \forall \psi \in \mathcal{H} \qquad \text{Notation: } A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases. Even if the operators commute!
- Theorem 40. ([AM14, Theorem A.3]) Let $A_1, \ldots, A_M \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$. Then there is a unique element with minimal trace in $\{Y \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}} : A_1, \ldots, A_M \leq Y\}$. Definition: $\max_{\mathrm{Tr}}\{A_1, \ldots, A_M\}$

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \ \ \forall \psi \in \mathcal{H} \qquad \text{Notation: } \ A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases. Even if the operators commute!
- Theorem 40. ([AM14, Theorem A.3]) Let $A_1, \ldots, A_M \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$. Then there is a unique element with minimal trace in $\{Y \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}} : A_1, \ldots, A_M \leq Y\}$. Definition: $\max_{\mathrm{Tr}}\{A_1, \ldots, A_M\}$

  Similarly,

$$\begin{aligned} \min_{\mathrm{Tr}}\{A_1, \ldots, A_M\} &:= \operatorname{argmax}\{\operatorname{Tr} X : X \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}, X \leq A_i \ \ \forall i\} \\ &= -\max_{\mathrm{Tr}}\{-A_1, \ldots, -A_M\}. \end{aligned}$$

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \ \ \forall \psi \in \mathcal{H} \qquad \text{Notation: } \ A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases. Even if the operators commute!
- Theorem 40. ([AM14, Theorem A.3]) Let $A_1, \ldots, A_M \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$. Then there is a unique element with minimal trace in $\{Y \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}} : A_1, \ldots, A_M \leq Y\}$. Definition: $\max_{\mathrm{Tr}}\{A_1, \ldots, A_M\}$

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \;\; \forall \psi \in \mathcal{H} \qquad \text{Notation: } A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases. Even if the operators commute!
- Theorem 40. ([AM14, Theorem A.3]) Let $A_1, \ldots, A_M \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$. Then there is a unique element with minimal trace in $\{Y \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}} : A_1, \ldots, A_M \leq Y\}$. Definition: $\max_{\mathrm{Tr}}\{A_1, \ldots, A_M\}$

  Exercise 41. If all $A_i$ commute then $\max_{\mathrm{Tr}}\{A_1, \ldots, A_M\}$ is the usual entry-wise maximum.

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \quad \forall \psi \in \mathcal{H} \qquad \text{Notation: } A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases. Even if the operators commute!
- Theorem 40. ([AM14, Theorem A.3]) Let $A_1, \ldots, A_M \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$. Then there is a unique element with minimal trace in $\{Y \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}} : A_1, \ldots, A_M \leq Y\}$. Definition: $\max_{\mathrm{Tr}}\{A_1, \ldots, A_M\}$

# Maximum and minimum of operators

- Reminder: $A \in \mathcal{B}(\mathcal{H})$ is positive semi-definite (PSD) if

$$\langle \psi, A\psi \rangle \geq 0 \quad \forall \psi \in \mathcal{H} \qquad \text{Notation: } A \geq 0$$

- PSD order: $A \geq B$ if $A - B \geq 0$.
- Fun fact: A finite set of self-adjoint operators does not have a maximal or minimal element in the PSD order, except for very special cases. Even if the operators commute!
- Theorem 40. ([AM14, Theorem A.3]) Let $A_1, \ldots, A_M \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$. Then there is a unique element with minimal trace in $\{Y \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}} : A_1, \ldots, A_M \leq Y\}$. Definition: $\max_{\mathrm{Tr}}\{A_1, \ldots, A_M\}$

Exercise 42. For $\Sigma = \{A_1, \ldots, A_r\} \subseteq \mathcal{B}(\mathcal{H})_+$,

$$R_\infty^*(\{A_1, \ldots, A_r\}) = \log \mathrm{Tr} \max_{\mathrm{Tr}}\{A_1, \ldots, A_r\}$$

$$\sigma^* := \frac{\max_{\mathrm{Tr}}\{A_1, \ldots, A_r\}}{\mathrm{Tr}(\ldots)} \quad \text{unique } D_\infty^*\text{-center.}$$

# Guessing probability and the $D_\infty^*$-radius

Theorem 43. ([YKL75])

$$P_s^*(\{\varrho_i, p_i\}) = \operatorname{Tr}\max_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\}$$

*Proof:* : The first equality follows from the duality of semi-definite programming, the rest are trivial.

# Guessing probability and the $D_\infty^*$-radius

Theorem 43. ([YKL75])

$$\begin{aligned} P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr}\max\nolimits_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\ &= \min\{\operatorname{Tr} Y : Y = Y^*,\, p_i\varrho_i \le Y \quad \forall i\} \end{aligned}$$

*Proof:* : The first equality follows from the duality of semi-definite programming, the rest are trivial.

**Theorem 43.** ([YKL75])

$$\begin{aligned} P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr}\max\nolimits_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\ &= \min\{\operatorname{Tr} Y : Y = Y^*,\, p_i\varrho_i \le Y \quad \forall i\} \\ &= \min_{\sigma\in\mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \le \lambda\sigma \quad \forall i\} \end{aligned}$$

*Proof:* : The first equality follows from the duality of semi-definite programming, the rest are trivial.

**Theorem 43.** ([YKL75])

$$\begin{aligned} P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr}\max_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\ &= \min\{\operatorname{Tr} Y : Y = Y^*,\ p_i\varrho_i \le Y \quad \forall i\} \\ &= \min_{\sigma\in\mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \le \lambda\sigma \quad \forall i\} \\ &= \exp\left(\min_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_i D_\infty^*(p_i\varrho_i\|\sigma\right) \end{aligned}$$

*Proof:* : The first equality follows from the duality of semi-definite programming, the rest are trivial.

Theorem 43. ([YKL75])

$$\begin{aligned}
P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr}\max_{\operatorname{Tr}}\{p_1\varrho_1, \dots, p_M\varrho_M\} \\
&= \min\{\operatorname{Tr} Y : Y = Y^*, \, p_i\varrho_i \le Y \quad \forall i\} \\
&= \min_{\sigma\in\mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \le \lambda\sigma \quad \forall i\} \\
&= \exp\left(\min_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_i D_\infty^*(p_i\varrho_i\|\sigma\right) \\
&= \exp(R_\infty^*(\{p_i\varrho_i\})\,.
\end{aligned}$$

*Proof:* : The first equality follows from the duality of semi-definite programming, the rest are trivial.

# Guessing probability and the $D_\infty^*$-radius

**Theorem 43.** ([YKL75])

$$\begin{aligned} P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr}\max_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\ &= \min\{\operatorname{Tr} Y : Y = Y^*,\ p_i\varrho_i \le Y \quad \forall i\} \\ &= \min_{\sigma\in\mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \le \lambda\sigma \quad \forall i\} \\ &= \exp\left(\min_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_i D_\infty^*(p_i\varrho_i\|\sigma\right) \\ &= \exp(R_\infty^*(\{p_i\varrho_i\}) . \end{aligned}$$

*Proof:* : The first equality follows from the duality of semi-definite programming, the rest are trivial.

**Corollary 44.** When $p_1 = \ldots = p_M = \frac{1}{M}$, we have

$$P_s^*(\{\varrho_i, p_i\}) = \frac{1}{M}\exp\left(R_\infty^*(\{\varrho_i\})\right)$$

Operational interpretation of the $D_\infty^*$-radius.

# Guessing probability and conditional min-entropy

Theorem 43. ([YKL75])

$$
\begin{aligned}
P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr} \max_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\
&= \min\{\operatorname{Tr} Y : Y = Y^*,\ p_i\varrho_i \le Y \quad \forall i\} \\
&= \min_{\sigma \in \mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \le \lambda\sigma \quad \forall i\}
\end{aligned}
$$

# Guessing probability and conditional min-entropy

**Theorem 43.** ([YKL75])

$$\begin{aligned} P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr}\max\nolimits_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\ &= \min\{\operatorname{Tr} Y : Y = Y^*, \, p_i\varrho_i \le Y \ \ \forall i\} \\ &= \min_{\sigma\in\mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \le \lambda\sigma \ \ \forall i\} \\ &= \exp\left(-H_\infty^*(A|B)_\varrho\right), \end{aligned}$$

where $\varrho_{AB} = \sum_{i=1}^{M} p_i |i\rangle\langle i| \otimes \varrho_i$ classical-quantum state.

- Operational interpretation of the conditional min-entropy [KRS09].

# Guessing probability and conditional min-entropy

**Theorem 43.** ([YKL75])

$$
\begin{aligned}
P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr} \max\nolimits_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\
&= \min\{\operatorname{Tr} Y : Y = Y^*,\ p_i\varrho_i \le Y \quad \forall i\} \\
&= \min_{\sigma \in \mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \le \lambda\sigma \quad \forall i\} \\
&= \exp\left(-H_\infty^*(A|B)_\varrho\right),
\end{aligned}
$$

where $\varrho_{AB} = \sum_{i=1}^{M} p_i |i\rangle\langle i| \otimes \varrho_i$ classical-quantum state.

- Operational interpretation of the conditional min-entropy [KRS09].
- $P_s^*(\{\varrho_i, p_i\})$ is the optimal probability of correctly guessing the classical value of $i$ in system $A$ based on the quantum side-information in system $B$.

# Guessing probability and conditional min-entropy

Theorem 43. ([YKL75])

$$\begin{aligned} P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr} \max_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\ &= \min\{\operatorname{Tr} Y : Y = Y^*, \, p_i\varrho_i \leq Y \quad \forall i\} \\ &= \min_{\sigma \in \mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \leq \lambda\sigma \quad \forall i\} \\ &= \exp\left(-H_\infty^*(A|B)_\varrho\right), \end{aligned}$$

where $\varrho_{AB} = \sum_{i=1}^{M} p_i |i\rangle\langle i| \otimes \varrho_i$ classical-quantum state.

- Operational interpretation of the conditional min-entropy [KRS09].
- $P_s^*(\{\varrho_i, p_i\})$ is the optimal probability of correctly guessing the classical value of $i$ in system $A$ based on the quantum side-information in system $B$.
- No side-information: $P_s^*(\{p_i\}) = \max_i p_i = \exp\left(-H_\infty^*(A)_\varrho\right)$.

# Guessing probability and conditional min-entropy

**Theorem 43.** ([YKL75])

$$\begin{aligned} P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr}\max_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\ &= \min\{\operatorname{Tr} Y : Y = Y^*,\ p_i\varrho_i \le Y \quad \forall i\} \\ &= \min_{\sigma \in \mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \le \lambda\sigma \quad \forall i\} \\ &= \exp\left(-H_\infty^*(A|B)_\varrho\right), \end{aligned}$$

where $\varrho_{AB} = \sum_{i=1}^{M} p_i |i\rangle\langle i| \otimes \varrho_i$ classical-quantum state.

- Operational interpretation of the conditional min-entropy [KRS09].

# Guessing probability and conditional min-entropy

**Theorem 43.** ([YKL75])

$$\begin{aligned} P_s^*(\{\varrho_i, p_i\}) &= \operatorname{Tr} \max_{\operatorname{Tr}}\{p_1\varrho_1, \ldots, p_M\varrho_M\} \\ &= \min\{\operatorname{Tr} Y : Y = Y^*,\, p_i\varrho_i \leq Y \quad \forall i\} \\ &= \min_{\sigma\in\mathcal{S}(\mathcal{H})} \inf\{\lambda > 0 : p_i\varrho_i \leq \lambda\sigma \quad \forall i\} \\ &= \exp\left(-H_\infty^*(A|B)_\varrho\right), \end{aligned}$$

where $\quad \varrho_{AB} = \sum_{i=1}^{M} p_i |i\rangle\langle i| \otimes \varrho_i \quad$ classical-quantum state.

- Operational interpretation of the conditional min-entropy [KRS09].

  Also for a quantum-quantum state as the maximal singlet fraction.

# A geometric problem

Exercise 44. Let $\varrho_1, \dots, \varrho_M \in \mathcal{S}(\mathcal{H})$, and let $q_{\max}$ be the largest $q \in (0, 1]$ such that there exist states $\varrho_1', \dots, \varrho_r' \in \mathcal{S}(\mathcal{H})$ with

$$q\varrho_i + (1-q)\varrho_i' = q\varrho_j + (1-q)\varrho_j' \quad \forall i, j.$$

Show that $-\log q_{\max} = R_\infty^*(\{\varrho_i\})$, and $q\varrho_i + (1-q)\varrho_i'$ is the $D_\infty^*$ divergence center.

Exercise 45. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\max\{\operatorname{Tr} AT + \operatorname{Tr} B(I - T) : 0 \le T \le I\} = \operatorname{Tr} \max_{\operatorname{Tr}}\{A, B\}, \qquad \text{and}$$

$$\max_{\operatorname{Tr}}\{A, B\} = A + (A - B)_- = B + (A - B)_+ = \frac{1}{2}\left(A + B + |A - B|\right).$$

Exercise 45. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\max\{\operatorname{Tr} AT + \operatorname{Tr} B(I-T) : 0 \le T \le I\} = \operatorname{Tr} \max_{\operatorname{Tr}}\{A, B\}, \qquad \text{and}$$

$$\max_{\operatorname{Tr}}\{A, B\} = A + (A-B)_- = B + (A-B)_+ = \frac{1}{2}\left(A + B + |A-B|\right).$$

Corollary 46. Optimal guessing probability for $\varrho_1$ vs. $\varrho_2$ with equal priors is

$$P_s^* = \frac{1}{2}\left(1 + \frac{1}{2}\left\|\varrho_1 - \varrho_2\right\|_1\right).$$

Operational interpretation of the trace norm.

Exercise 45. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\max\{\operatorname{Tr} AT + \operatorname{Tr} B(I-T) : 0 \leq T \leq I\} = \operatorname{Tr} \max_{\operatorname{Tr}}\{A, B\}, \qquad \text{and}$$

$$\max_{\operatorname{Tr}}\{A, B\} = A + (A-B)_- = B + (A-B)_+ = \frac{1}{2}\left(A + B + |A-B|\right).$$

Corollary 46. Optimal guessing probability for $\varrho_1$ vs. $\varrho_2$ with equal priors is

$$P_s^* = \frac{1}{2}\left(1 + \frac{1}{2}\left\|\varrho_1 - \varrho_2\right\|_1\right).$$

Operational interpretation of the trace norm.

Corollary 47. $-\log q_{\max} = R_\infty^*(\{\varrho_1, \varrho_2\}) = \log\left(1 + \frac{1}{2}\left\|\varrho_1 - \varrho_2\right\|_1\right)$,

and $\frac{\varrho_1+\varrho_2+|\varrho_1-\varrho_2|}{2+\|\varrho_1-\varrho_2\|_1}$ is the $D_\infty^*$ center.

Exercise 45. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\max\{\operatorname{Tr} AT + \operatorname{Tr} B(I-T) : 0 \le T \le I\} = \operatorname{Tr} \max_{\operatorname{Tr}}\{A, B\}, \quad \text{and}$$

$$\max_{\operatorname{Tr}}\{A, B\} = A + (A-B)_- = B + (A-B)_+ = \frac{1}{2}\left(A + B + |A-B|\right).$$

Corollary 46. Optimal guessing probability for $\varrho_1$ vs. $\varrho_2$ with equal priors is

$$P_s^* = \frac{1}{2}\left(1 + \frac{1}{2}\|\varrho_1 - \varrho_2\|_1\right).$$

Operational interpretation of the trace norm.

Corollary 47. $-\log q_{\max} = R_\infty^*(\{\varrho_1, \varrho_2\}) = \log\left(1 + \frac{1}{2}\|\varrho_1 - \varrho_2\|_1\right)$,

and $\frac{\varrho_1 + \varrho_2 + |\varrho_1 - \varrho_2|}{2 + \|\varrho_1 - \varrho_2\|_1}$ is the $D_\infty^*$ center.

Application: Improved Alicki-Fannes inequalities for the continuity of the conditional von Neumann entropy. [MH11, Winter16]

# Binary state discrimination

Exercise 45. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\max\{\operatorname{Tr} AT + \operatorname{Tr} B(I-T) : 0 \le T \le I\} = \operatorname{Tr} \max_{\operatorname{Tr}}\{A, B\}, \qquad \text{and}$$

$$\max_{\operatorname{Tr}}\{A, B\} = A + (A-B)_- = B + (A-B)_+ = \frac{1}{2}\left(A + B + |A-B|\right).$$

Corollary 46. Optimal guessing probability for $\varrho_1$ vs. $\varrho_2$ with equal priors is

$$P_s^* = \frac{1}{2}\left(1 + \frac{1}{2}\left\|\varrho_1 - \varrho_2\right\|_1\right).$$

Operational interpretation of the trace norm.

# Binary state discrimination

Exercise 45. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\max\{\operatorname{Tr} AT + \operatorname{Tr} B(I-T) : 0 \le T \le I\} = \operatorname{Tr} \max_{\operatorname{Tr}}\{A, B\}, \quad \text{and}$$

$$\max_{\operatorname{Tr}}\{A, B\} = A + (A-B)_- = B + (A-B)_+ = \frac{1}{2}\left(A + B + |A-B|\right).$$

Corollary 46. Optimal guessing probability for $\varrho_1$ vs. $\varrho_2$ with equal priors is

$$P_s^* = \frac{1}{2}\left(1 + \frac{1}{2}\left\|\varrho_1 - \varrho_2\right\|_1\right).$$

Operational interpretation of the trace norm.

Corollary 48. Optimal error probability:

$$P_e^* = 1 - P_s^* = \frac{1}{2}\left(1 - \frac{1}{2}\left\|\varrho_1 - \varrho_2\right\|_1\right).$$

# Binary state discrimination

Exercise 45. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\max\{\operatorname{Tr} AT + \operatorname{Tr} B(I-T) : 0 \le T \le I\} = \operatorname{Tr} \max_{\operatorname{Tr}}\{A, B\}, \quad \text{and}$$

$$\max_{\operatorname{Tr}}\{A, B\} = A + (A-B)_- = B + (A-B)_+ = \frac{1}{2}(A + B + |A-B|).$$

Corollary 46. Optimal guessing probability for $\varrho_1$ vs. $\varrho_2$ with equal priors is

$$P_s^* = \frac{1}{2}\left(1 + \frac{1}{2}\|\varrho_1 - \varrho_2\|_1\right).$$

Operational interpretation of the trace norm.

Corollary 48. Optimal error probability:

$$P_e^* = 1 - P_s^* = \frac{1}{2}\left(1 - \frac{1}{2}\|\varrho_1 - \varrho_2\|_1\right).$$

What is the asymptotics of this quantity for many copies of $\varrho_1, \varrho_2$?

$$\frac{1}{2}\left(1 - \frac{1}{2}\left\|\varrho_1^{\otimes n} - \varrho_2^{\otimes n}\right\|_1\right) \sim_{n\to+\infty} \quad ?$$

# V. Asymptotic binary state discrimination and Rényi divergences

# Binary state discrimination

- Problem: Alice wants to send Bob one of $2$ possible messages using a quantum system.

# Binary state discrimination

- Problem: Alice wants to send Bob one of $2$ possible messages using a quantum system.

  She encodes message $i \in \{0,1\}$ into some state of the quantum system, and sends it to Bob over a quantum channel; Bob receives the system in state $\varrho$ if the meassage was $0$ and in state $\sigma$ if the message was $1$.

# Binary state discrimination

- Problem: Alice wants to send Bob one of $2$ possible messages using a quantum system.

  She encodes message $i \in \{0,1\}$ into some state of the quantum system, and sends it to Bob over a quantum channel; Bob receives the system in state $\varrho$ if the meassage was $0$ and in state $\sigma$ if the message was $1$.

  Bob performs a POVM $(T(0) = T,\, T(1) = I - T)$ to decide which message was sent.

# Binary state discrimination

- Problem: Alice wants to send Bob one of $2$ possible messages using a quantum system.

  She encodes message $i \in \{0, 1\}$ into some state of the quantum system, and sends it to Bob over a quantum channel; Bob receives the system in state $\varrho$ if the meassage was $0$ and in state $\sigma$ if the message was $1$.

  Bob performs a POVM $(T(0) = T, T(1) = I - T)$ to decide which message was sent.

- type I error: A: 0 but B: 1 $\quad \alpha(T) := \mathrm{Tr}\,\varrho(I - T)$
  type II error: A: 1 but B: 0 $\quad \beta(T) := \mathrm{Tr}\,\sigma T$

# Binary state discrimination

- Problem: Alice wants to send Bob one of $2$ possible messages using a quantum system.

  She encodes message $i \in \{0,1\}$ into some state of the quantum system, and sends it to Bob over a quantum channel; Bob receives the system in state $\varrho$ if the meassage was $0$ and in state $\sigma$ if the message was $1$.

  Bob performs a POVM $(T(0) = T, T(1) = I - T)$ to decide which message was sent.

- type I error: A: 0 but B: 1 $\quad \alpha(T) := \operatorname{Tr} \varrho(I - T)$
  type II error: A: 1 but B: 0 $\quad \beta(T) := \operatorname{Tr} \sigma T$

- Optimal symmetric error probability:

$$P_e^*(\varrho, \sigma) := \min \left\{ \frac{1}{2}\alpha(T) + \frac{1}{2}\beta(T) \right\} = \frac{1}{2}\left(1 - \frac{1}{2}\|\varrho - \sigma\|_1\right)$$

  Strictly positive unless $\varrho \perp \sigma$.

# Binary state discrimination

- To reduce the error, Alice can send the same message $n > 1$ times.

# Binary state discrimination

- To reduce the error, Alice can send the same message $n > 1$ times.
- type I error: A: 0 but B: 1 $\qquad \alpha_n(T) := \operatorname{Tr} \varrho^{\otimes n}(I - T)$
  type II error: A: 1 but B: 0 $\qquad \beta_n(T) := \operatorname{Tr} \sigma^{\otimes n} T$

  Optimal symmetric error probability:

$$P^*_{e,n}(\varrho, \sigma) := \min \left\{ \frac{1}{2}\alpha_n(T) + \frac{1}{2}\beta_n(T) \right\} = \frac{1}{2}\left(1 - \frac{1}{2}\left\|\varrho^{\otimes n} - \sigma^{\otimes n}\right\|_1\right)$$

# Binary state discrimination

- To reduce the error, Alice can send the same message $n > 1$ times.
- type I error: A: 0 but B: 1 $\qquad \alpha_n(T) := \operatorname{Tr} \varrho^{\otimes n}(I - T)$
  type II error: A: 1 but B: 0 $\qquad \beta_n(T) := \operatorname{Tr} \sigma^{\otimes n} T$

  Optimal symmetric error probability:

$$P_{e,n}^*(\varrho, \sigma) := \min \left\{ \frac{1}{2}\alpha_n(T) + \frac{1}{2}\beta_n(T) \right\} = \frac{1}{2}\left(1 - \frac{1}{2}\left\|\varrho^{\otimes n} - \sigma^{\otimes n}\right\|_1\right)$$

- Exercise 49. Use the Fuchs - van de Graaf inequalities ([NC, Chapter 9]) to show that $P_{e,n}^*$ decays exponentially fast in $n$.

# Binary state discrimination

- To reduce the error, Alice can send the same message $n > 1$ times.
- type I error: A: 0 but B: 1 $\alpha_n(T) := \operatorname{Tr} \varrho^{\otimes n}(I - T)$
  type II error: A: 1 but B: 0 $\beta_n(T) := \operatorname{Tr} \sigma^{\otimes n} T$

  Optimal symmetric error probability:

  $$P_{e,n}^*(\varrho, \sigma) := \min\left\{\frac{1}{2}\alpha_n(T) + \frac{1}{2}\beta_n(T)\right\} = \frac{1}{2}\left(1 - \frac{1}{2}\left\|\varrho^{\otimes n} - \sigma^{\otimes n}\right\|_1\right)$$

- Exercise 49. Use the Fuchs - van de Graaf inequalities ([NC, Chapter 9]) to show that $P_{e,n}^*$ decays exponentially fast in $n$.
- What is the optimal exponent?

# Binary state discrimination

- To reduce the error, Alice can send the same message $n > 1$ times.
- type I error: A: 0 but B: 1 $\alpha_n(T) := \operatorname{Tr} \varrho^{\otimes n}(I - T)$
  type II error: A: 1 but B: 0 $\beta_n(T) := \operatorname{Tr} \sigma^{\otimes n} T$

  Optimal symmetric error probability:

$$P^*_{e,n}(\varrho, \sigma) := \min \left\{ \frac{1}{2}\alpha_n(T) + \frac{1}{2}\beta_n(T) \right\} = \frac{1}{2}\left(1 - \frac{1}{2}\left\|\varrho^{\otimes n} - \sigma^{\otimes n}\right\|_1\right)$$

- Exercise 49. Use the Fuchs - van de Graaf inequalities ([NC, Chapter 9]) to show that $P^*_{e,n}$ decays exponentially fast in $n$.
- What is the optimal exponent?
- More generally, what are the achievable rate pairs ?

$$\left(\lim_n -\frac{1}{n}\log \alpha_n(T_n),\ -\frac{1}{n}\log \beta_n(T_n)\right)$$

# Binary state discrimination

- Idea: Adjust the balance between the rates by a parameter $c \in \mathbb{R}$:

$$\begin{aligned} e_n(c) &:= \min_{0 \le T \le I} \{\alpha_n(T_n) + e^{nc}\beta_n(T_n)\} \\ &= \min\{\operatorname{Tr} \varrho^{\otimes n}(I - T) + e^{nc} \operatorname{Tr} \sigma^{\otimes n} T\} \end{aligned} \tag{1}$$

# Binary state discrimination

- Idea: Adjust the balance between the rates by a parameter $c \in \mathbb{R}$:

$$\begin{aligned} e_n(c) &:= \min_{0\le T\le I}\{\alpha_n(T_n)+e^{nc}\beta_n(T_n)\} \\ &= \min\{\operatorname{Tr}\varrho^{\otimes n}(I-T)+e^{nc}\operatorname{Tr}\sigma^{\otimes n}T\} \end{aligned} \tag{1}$$

- Exercise 50. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\min\{\operatorname{Tr} A(I-T)+\operatorname{Tr} BT:\ 0\le T\le I\} = \operatorname{Tr}\min_{\mathrm{Tr}}\{A,B\}, \qquad \text{and}$$

$$\min_{\mathrm{Tr}}\{A,B\} = A-(A-B)_+ = B-(A-B)_- = \frac{1}{2}\left(A+B-|A-B|\right).$$

Not necessarily PSD!

# Binary state discrimination

- Idea: Adjust the balance between the rates by a parameter $c \in \mathbb{R}$:

$$\begin{aligned} e_n(c) &:= \min_{0\le T\le I} \{\alpha_n(T_n) + e^{nc}\beta_n(T_n)\} \\ &= \min\{\operatorname{Tr}\varrho^{\otimes n}(I-T) + e^{nc}\operatorname{Tr}\sigma^{\otimes n}T\} \end{aligned} \tag{1}$$

- Exercise 50. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\min\{\operatorname{Tr} A(I-T) + \operatorname{Tr} BT : \, 0 \le T \le I\} = \operatorname{Tr}\min_{\operatorname{Tr}}\{A, B\}, \quad \text{and}$$

$$\min_{\operatorname{Tr}}\{A, B\} = A - (A-B)_+ = B - (A-B)_- = \frac{1}{2}\left(A + B - |A-B|\right).$$

Not necessarily PSD!

Moreover, $T$ is optimal iff $\{A - B > 0\} \le T \le \{A - B \ge 0\}$.

(For $X^* = X = \sum_i x_i |e_i\rangle\langle e_i|$: $\{X \ge c\} := \sum_{i:\, x_i \ge c} |e_i\rangle\langle e_i|$.)

# Binary state discrimination

- Idea: Adjust the balance between the rates by a parameter $c \in \mathbb{R}$:

$$e_n(c) := \min_{0 \leq T \leq I} \{\alpha_n(T_n) + e^{nc}\beta_n(T_n)\} = \min\{\operatorname{Tr} \varrho^{\otimes n}(I - T) + e^{nc} \operatorname{Tr} \sigma^{\otimes n} T\} \qquad (1)$$

- Exercise 50. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\min\{\operatorname{Tr} A(I - T) + \operatorname{Tr} BT : 0 \leq T \leq I\} = \operatorname{Tr} \min_{\operatorname{Tr}}\{A, B\}, \quad \text{and}$$

$$\min_{\operatorname{Tr}}\{A, B\} = A - (A - B)_+ = B - (A - B)_- = \frac{1}{2}(A + B - |A - B|).$$

Not necessarily PSD!

Moreover, $T$ is optimal iff $\{A - B > 0\} \leq T \leq \{A - B \geq 0\}$.

(For $X^* = X = \sum_i x_i |e_i\rangle\langle e_i|$: $\{X \geq c\} := \sum_{i:\, x_i \geq c} |e_i\rangle\langle e_i|$.)

- Min. in (1): Neyman-Pearson test $T_{n,c} := \{\varrho^{\otimes n} - e^{nc}\sigma^{\otimes n} \geq 0\}$

# Binary state discrimination

- Idea: Adjust the balance between the rates by a parameter $c \in \mathbb{R}$:

$$\begin{aligned} e_n(c) &:= \min_{0 \le T \le I} \{\alpha_n(T_n) + e^{nc}\beta_n(T_n)\} \\ &= \alpha_n(T_{n,c}) + e^{nc}\beta_n(T_{n,c}) \end{aligned}$$

- Exercise 50. Let $A, B \in \mathcal{B}(\mathcal{H})_+$. Give a direct proof that

$$\min\{\operatorname{Tr} A(I-T) + \operatorname{Tr} BT : 0 \le T \le I\} = \operatorname{Tr} \min_{\operatorname{Tr}}\{A, B\}, \quad \text{and}$$

$$\min_{\operatorname{Tr}}\{A, B\} = A - (A-B)_+ = B - (A-B)_- = \frac{1}{2}(A + B - |A - B|).$$

Not necessarily PSD!

Moreover, $T$ is optimal iff $\{A - B > 0\} \le T \le \{A - B \ge 0\}$.

(For $X^* = X = \sum_i x_i |e_i\rangle\langle e_i|$: $\{X \ge c\} := \sum_{i:\, x_i \ge c} |e_i\rangle\langle e_i|$.)

- Min. in (1): Neyman-Pearson test $T_{n,c} := \{\varrho^{\otimes n} - e^{nc}\sigma^{\otimes n} \ge 0\}$

# Binary state discrimination

- Idea: Adjust the balance between the rates by a parameter $c \in \mathbb{R}$:

$$
\begin{aligned}
e_n(c) \quad &:= \quad \min_{0 \leq T \leq I} \{\alpha_n(T_n) + e^{nc}\beta_n(T_n)\} \\
&= \quad \alpha_n(T_{n,c}) + e^{nc}\beta_n(T_{n,c})
\end{aligned}
$$

# Binary state discrimination

- Idea: Adjust the balance between the rates by a parameter $c \in \mathbb{R}$:

$$\begin{aligned} e_n(c) &:= \min_{0 \leq T \leq I} \{\alpha_n(T_n) + e^{nc}\beta_n(T_n)\} \\ &= \alpha_n(T_{n,c}) + e^{nc}\beta_n(T_{n,c}) \end{aligned}$$

- Neyman-Pearson lemma: Among test sequences with a given type II error rate, the Neyman-Pearson tests have the best type I error rate.

# Binary state discrimination

- Idea: Adjust the balance between the rates by a parameter $c \in \mathbb{R}$:

$$\begin{aligned} e_n(c) &:= \min_{0 \leq T \leq I} \{\alpha_n(T_n) + e^{nc}\beta_n(T_n)\} \\ &= \alpha_n(T_{n,c}) + e^{nc}\beta_n(T_{n,c}) \end{aligned}$$

- Neyman-Pearson lemma: Among test sequences with a given type II error rate, the Neyman-Pearson tests have the best type I error rate.

  Basis of the information spectrum method.

# Binary state discrimination

- Classical case: $\varrho, \sigma$ functions on $\mathcal{X}$

$$T_{n,c} \quad \leftrightarrow \quad \left\{ \underline{x} \in \mathcal{X}^n : \frac{1}{n} \log \frac{\varrho^{\otimes n}(\underline{x})}{\sigma^{\otimes n}(\underline{x})} \geq c \right\}$$

modified maximum likelihood test, log-likelihood ratio test.

# Binary state discrimination

- Classical case: $\varrho, \sigma$ functions on $\mathcal{X}$

$$T_{n,c} \quad \leftrightarrow \quad \left\{ \underline{x} \in \mathcal{X}^n : \frac{1}{n} \log \frac{\varrho^{\otimes n}(\underline{x})}{\sigma^{\otimes n}(\underline{x})} \geq c \right\}$$

modified maximum likelihood test, log-likelihood ratio test.

- Markov inequality: $\alpha_n(T_{n,c}) \leq e^{-n \sup_{\alpha>0}\{c(\alpha-1)-(\alpha-1)D_\alpha(\varrho\|\sigma)\}}$

$$\beta_n(T_{n,c}) \leq e^{-n \sup_{\alpha>0}\{c\alpha-(\alpha-1)D_\alpha(\varrho\|\sigma)\}}$$

$$D_\alpha(\varrho\|\sigma) := \frac{1}{\alpha-1} \log \sum\nolimits_x \varrho(x)^\alpha \sigma(x)^{1-\alpha}$$ classical Rényi divergences

# Binary state discrimination

- Classical case: $\varrho, \sigma$ functions on $\mathcal{X}$

$$T_{n,c} \quad \leftrightarrow \quad \left\{ \underline{x} \in \mathcal{X}^n : \frac{1}{n} \log \frac{\varrho^{\otimes n}(\underline{x})}{\sigma^{\otimes n}(\underline{x})} \geq c \right\}$$

modified maximum likelihood test, log-likelihood ratio test.

- Markov inequality: $\alpha_n(T_{n,c}) \leq e^{-n \sup_{\alpha>0}\{c(\alpha-1)-(\alpha-1)D_\alpha(\varrho\|\sigma)\}}$

$\beta_n(T_{n,c}) \leq e^{-n \sup_{\alpha>0}\{c\alpha-(\alpha-1)D_\alpha(\varrho\|\sigma)\}}$

$$D_\alpha(\varrho\|\sigma) := \frac{1}{\alpha-1} \log \sum\nolimits_x \varrho(x)^\alpha \sigma(x)^{1-\alpha}$$ classical Rényi divergences

Cramér's large deviation theorem: The bounds are sharp in the asymptotics.

# Binary state discrimination

- Classical case: $\varrho, \sigma$ functions on $\mathcal{X}$

$$T_{n,c} \quad \leftrightarrow \quad \left\{ \underline{x} \in \mathcal{X}^n : \, \frac{1}{n} \log \frac{\varrho^{\otimes n}(\underline{x})}{\sigma^{\otimes n}(\underline{x})} \geq c \right\}$$

modified maximum likelihood test, log-likelihood ratio test.

- Markov inequality: $\alpha_n(T_{n,c}) \leq e^{-n \sup_{\alpha>0}\{c(\alpha-1)-(\alpha-1)D_\alpha(\varrho\|\sigma)\}}$

$$\beta_n(T_{n,c}) \leq e^{-n \sup_{\alpha>0}\{c\alpha-(\alpha-1)D_\alpha(\varrho\|\sigma)\}}$$

$$D_\alpha(\varrho\|\sigma) := \frac{1}{\alpha-1} \log \sum\nolimits_x \varrho(x)^\alpha \sigma(x)^{1-\alpha}$$ classical Rényi divergences

Cramér's large deviation theorem: The bounds are sharp in the asymptotics.

- The classical Rényi divergences appear as the logarithmic moment generating function for the log-likelihood ratio test.

# Operator monotone functions

Reminder: $f : (0, +\infty) \to \mathbb{R}$ operator monotone if

$0 \leq A \leq B \in \mathcal{B}(\mathcal{H}) \implies f(A) \leq f(B)$.

$t \mapsto t^\alpha$ operator monotone $\iff \alpha \in [0, 1]$.

[Bhatia]

# Operator monotone functions

Reminder: $f : (0, +\infty) \to \mathbb{R}$ operator monotone if

$$0 \le A \le B \in \mathcal{B}(\mathcal{H}) \implies f(A) \le f(B).$$

$t \mapsto t^\alpha$ operator monotone $\iff \alpha \in [0, 1]$.

[Bhatia]

Exercise 51. Prove that $t \mapsto t^\alpha$ is not operator monotone for $\alpha > 1$.

Hint: Use $2 \times 2$ matrices.

# Quantum large deviation bound

**Theorem 52.** (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\operatorname{Tr} \frac{1}{2} \left( A + B - |A - B| \right) \leq \operatorname{Tr} A^\alpha B^{1-\alpha}, \qquad \alpha \in [0, 1].$$

# Quantum large deviation bound

Theorem 52. (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\mathrm{Tr}\,\frac{1}{2}\,(A + B - |A - B|) \le \mathrm{Tr}\,A^\alpha B^{1-\alpha}, \qquad \alpha \in [0, 1].$$

*Proof:* (N. Ozawa) $\quad A \le A + (A - B)_- = B + (A - B)_+ \ge B$

# Quantum large deviation bound

Theorem 52. (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\mathrm{Tr}\, \frac{1}{2}(A + B - |A - B|) \leq \mathrm{Tr}\, A^\alpha B^{1-\alpha}, \qquad \alpha \in [0,1].$$

*Proof:* (N. Ozawa) $A \leq A + (A - B)_- = B + (A - B)_+ \geq B$

$$\mathrm{Tr}\, A - \mathrm{Tr}\, A^\alpha B^{1-\alpha} \quad = \quad \mathrm{Tr}\, A^\alpha A^{1-\alpha} - \mathrm{Tr}\, A^\alpha B^{1-\alpha}$$

# Quantum large deviation bound

Theorem 52. (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\mathrm{Tr}\,\frac{1}{2}\,(A + B - |A - B|) \leq \mathrm{Tr}\,A^\alpha B^{1-\alpha}, \qquad \alpha \in [0, 1].$$

*Proof:* (N. Ozawa) $\quad A \leq A + (A-B)_- = B + (A-B)_+ \geq B$

$$\begin{aligned}\mathrm{Tr}\,A - \mathrm{Tr}\,A^\alpha B^{1-\alpha} &= \mathrm{Tr}\,A^\alpha A^{1-\alpha} - \mathrm{Tr}\,A^\alpha B^{1-\alpha} \\ &\leq \mathrm{Tr}\,A^\alpha (B + (A-B)_+)^{1-\alpha} - \mathrm{Tr}\,A^\alpha B^{1-\alpha}\end{aligned}$$

# Quantum large deviation bound

Theorem 52. (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\mathrm{Tr}\,\frac{1}{2}\left(A + B - |A - B|\right) \le \mathrm{Tr}\,A^\alpha B^{1-\alpha}, \qquad \alpha \in [0,1].$$

*Proof:* (N. Ozawa) $A \le A + (A-B)_- = B + (A-B)_+ \ge B$

$$\begin{aligned}
\mathrm{Tr}\,A - \mathrm{Tr}\,A^\alpha B^{1-\alpha} &= \mathrm{Tr}\,A^\alpha A^{1-\alpha} - \mathrm{Tr}\,A^\alpha B^{1-\alpha} \\
&\le \mathrm{Tr}\,A^\alpha (B + (A-B)_+)^{1-\alpha} - \mathrm{Tr}\,A^\alpha B^{1-\alpha} \\
&= \mathrm{Tr}\,A^\alpha \left((B + (A-B)_+)^{1-\alpha} - B^{1-\alpha}\right)
\end{aligned}$$

# Quantum large deviation bound

Theorem 52. (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\mathrm{Tr}\,\frac{1}{2}\,(A + B - |A - B|) \le \mathrm{Tr}\,A^{\alpha} B^{1-\alpha}, \qquad \alpha \in [0,1].$$

*Proof:* (N. Ozawa) $\quad A \le A + (A-B)_- = B + (A-B)_+ \ge B$

$$\begin{aligned}
\mathrm{Tr}\,A - \mathrm{Tr}\,A^{\alpha} B^{1-\alpha} &= \mathrm{Tr}\,A^{\alpha} A^{1-\alpha} - \mathrm{Tr}\,A^{\alpha} B^{1-\alpha} \\
&\le \mathrm{Tr}\,A^{\alpha} (B + (A-B)_+)^{1-\alpha} - \mathrm{Tr}\,A^{\alpha} B^{1-\alpha} \\
&= \mathrm{Tr}\,A^{\alpha} \left( (B + (A-B)_+)^{1-\alpha} - B^{1-\alpha} \right) \\
&\le \mathrm{Tr}(B + (A-B)_+)^{\alpha} \left( (B + (A-B)_+)^{1-\alpha} - B^{1-\alpha} \right)
\end{aligned}$$

# Quantum large deviation bound

**Theorem 52.** (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\mathrm{Tr}\,\frac{1}{2}\,(A+B-|A-B|) \le \mathrm{Tr}\,A^\alpha B^{1-\alpha}, \qquad \alpha \in [0,1].$$

*Proof:* (N. Ozawa) $A \le A + (A-B)_- = B + (A-B)_+ \ge B$

$$\begin{aligned}
\mathrm{Tr}\,A - \mathrm{Tr}\,A^\alpha B^{1-\alpha} &= \mathrm{Tr}\,A^\alpha A^{1-\alpha} - \mathrm{Tr}\,A^\alpha B^{1-\alpha} \\
&\le \mathrm{Tr}\,A^\alpha (B+(A-B)_+)^{1-\alpha} - \mathrm{Tr}\,A^\alpha B^{1-\alpha} \\
&= \mathrm{Tr}\,A^\alpha \left((B+(A-B)_+)^{1-\alpha} - B^{1-\alpha}\right) \\
&\le \mathrm{Tr}(B+(A-B)_+)^\alpha \left((B+(A-B)_+)^{1-\alpha} - B^{1-\alpha}\right) \\
&= \mathrm{Tr}\,B + \mathrm{Tr}(A-B)_+ - \mathrm{Tr}(B+(A-B)_+)^\alpha B^{1-\alpha}
\end{aligned}$$

# Quantum large deviation bound

**Theorem 52.** (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\operatorname{Tr} \frac{1}{2}(A + B - |A - B|) \leq \operatorname{Tr} A^\alpha B^{1-\alpha}, \qquad \alpha \in [0, 1].$$

*Proof:* (N. Ozawa) $A \leq A + (A - B)_- = B + (A - B)_+ \geq B$

$$\begin{aligned}
\operatorname{Tr} A - \operatorname{Tr} A^\alpha B^{1-\alpha} &= \operatorname{Tr} A^\alpha A^{1-\alpha} - \operatorname{Tr} A^\alpha B^{1-\alpha} \\
&\leq \operatorname{Tr} A^\alpha (B + (A - B)_+)^{1-\alpha} - \operatorname{Tr} A^\alpha B^{1-\alpha} \\
&= \operatorname{Tr} A^\alpha \left((B + (A - B)_+)^{1-\alpha} - B^{1-\alpha}\right) \\
&\leq \operatorname{Tr}(B + (A - B)_+)^\alpha \left((B + (A - B)_+)^{1-\alpha} - B^{1-\alpha}\right) \\
&= \operatorname{Tr} B + \operatorname{Tr}(A - B)_+ - \operatorname{Tr}(B + (A - B)_+)^\alpha B^{1-\alpha} \\
&= \operatorname{Tr}(A - B)_+ + \operatorname{Tr}\left[B^\alpha - (B + (A - B)_+)^\alpha\right] B^{1-\alpha}
\end{aligned}$$

# Quantum large deviation bound

Theorem 52. (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\operatorname{Tr} \frac{1}{2}(A + B - |A - B|) \leq \operatorname{Tr} A^\alpha B^{1-\alpha}, \qquad \alpha \in [0, 1].$$

*Proof:* (N. Ozawa) $A \leq A + (A - B)_- = B + (A - B)_+ \geq B$

$$\begin{aligned}
\operatorname{Tr} A - \operatorname{Tr} A^\alpha B^{1-\alpha} &= \operatorname{Tr} A^\alpha A^{1-\alpha} - \operatorname{Tr} A^\alpha B^{1-\alpha} \\
&\leq \operatorname{Tr} A^\alpha (B + (A - B)_+)^{1-\alpha} - \operatorname{Tr} A^\alpha B^{1-\alpha} \\
&= \operatorname{Tr} A^\alpha \left((B + (A - B)_+)^{1-\alpha} - B^{1-\alpha}\right) \\
&\leq \operatorname{Tr}(B + (A - B)_+)^\alpha \left((B + (A - B)_+)^{1-\alpha} - B^{1-\alpha}\right) \\
&= \operatorname{Tr} B + \operatorname{Tr}(A - B)_+ - \operatorname{Tr}(B + (A - B)_+)^\alpha B^{1-\alpha} \\
&= \operatorname{Tr}(A - B)_+ + \operatorname{Tr}\left[B^\alpha - (B + (A - B)_+)^\alpha\right] B^{1-\alpha} \\
&\leq \operatorname{Tr}(A - B)_+.
\end{aligned}$$

Rearranging, and noting that $A - (A - B)_+ = \frac{1}{2}(A + B - |A - B|)$, we get the desired inequality.

# Quantum large deviation bound

Theorem 52. (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\operatorname{Tr} \frac{1}{2}\left(A + B - |A - B|\right) \leq \operatorname{Tr} A^\alpha B^{1-\alpha}, \qquad \alpha \in [0, 1].$$

**Theorem 52.** (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\operatorname{Tr}\frac{1}{2}(A+B-|A-B|) \le \operatorname{Tr} A^\alpha B^{1-\alpha}, \qquad \alpha \in [0,1].$$

Apply the above to $A := \varrho^{\otimes n}$ and $B := e^{nc}\sigma^{\otimes n}$ to get

$$\begin{aligned}&\max\{\alpha_n(T_{n,c}),\, e^{nc}\beta_n(T_{n,c})\}\\ &\quad \le \alpha_n(T_{n,c}) + e^{nc}\beta_n(T_{n,c}) = \frac{1+e^{nc}}{2} - \frac{1}{2}\left\|\varrho^{\otimes n} - e^{nc}\sigma^{\otimes n}\right\|_1\end{aligned}$$

# Quantum large deviation bound

Theorem 52. (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\operatorname{Tr} \frac{1}{2}(A + B - |A - B|) \leq \operatorname{Tr} A^\alpha B^{1-\alpha}, \qquad \alpha \in [0,1].$$

Apply the above to $A := \varrho^{\otimes n}$ and $B := e^{nc}\sigma^{\otimes n}$ to get

$$\begin{aligned}
&\max\{\alpha_n(T_{n,c}),\, e^{nc}\beta_n(T_{n,c})\} \\
&\quad \leq \alpha_n(T_{n,c}) + e^{nc}\beta_n(T_{n,c}) = \frac{1 + e^{nc}}{2} - \frac{1}{2}\left\|\varrho^{\otimes n} - e^{nc}\sigma^{\otimes n}\right\|_1 \\
&\quad \leq \inf_{0<\alpha<1} \operatorname{Tr}\left(\varrho^{\otimes n}\right)^\alpha \left(e^{nc}\sigma^{\otimes n}\right)^{1-\alpha}
\end{aligned}$$

# Quantum large deviation bound

Theorem 52. (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\operatorname{Tr} \frac{1}{2}(A + B - |A - B|) \le \operatorname{Tr} A^\alpha B^{1-\alpha}, \qquad \alpha \in [0,1].$$

Apply the above to $A := \varrho^{\otimes n}$ and $B := e^{nc}\sigma^{\otimes n}$ to get

$$\begin{aligned} &\max\{\alpha_n(T_{n,c}),\, e^{nc}\beta_n(T_{n,c})\} \\ &\quad\le \alpha_n(T_{n,c}) + e^{nc}\beta_n(T_{n,c}) = \frac{1+e^{nc}}{2} - \frac{1}{2}\left\|\varrho^{\otimes n} - e^{nc}\sigma^{\otimes n}\right\|_1 \\ &\quad\le \inf_{0<\alpha<1} \operatorname{Tr}\left(\varrho^{\otimes n}\right)^\alpha \left(e^{nc}\sigma^{\otimes n}\right)^{1-\alpha} = e^{-n\sup_{1<\alpha<1}\{c(\alpha-1)-\psi(\alpha)\}} \end{aligned}$$

$\psi(\alpha) := \log \operatorname{Tr} \varrho^\alpha \sigma^{1-\alpha}, \qquad \varphi(c) := \sup_{0<\alpha<1} \{c\alpha - \psi(\alpha)\}$ Legendre transform

# Quantum large deviation bound

**Theorem 52.** (Audenaert's inequality)

For $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\mathrm{Tr}\,\frac{1}{2}\left(A + B - |A - B|\right) \le \mathrm{Tr}\,A^\alpha B^{1-\alpha}, \qquad \alpha \in [0,1].$$

Apply the above to $A := \varrho^{\otimes n}$ and $B := e^{nc}\sigma^{\otimes n}$ to get

$$\begin{aligned} &\max\{\alpha_n(T_{n,c}),\, e^{nc}\beta_n(T_{n,c})\} \\ &\quad \le \alpha_n(T_{n,c}) + e^{nc}\beta_n(T_{n,c}) = \frac{1+e^{nc}}{2} - \frac{1}{2}\left\|\varrho^{\otimes n} - e^{nc}\sigma^{\otimes n}\right\|_1 \\ &\quad \le \inf_{0<\alpha<1} \mathrm{Tr}\left(\varrho^{\otimes n}\right)^\alpha \left(e^{nc}\sigma^{\otimes n}\right)^{1-\alpha} = e^{-n\sup_{1<\alpha<1}\{c(\alpha-1)-\psi(\alpha)\}} \end{aligned}$$

**Corollary 53.**

$$\alpha_n(T_{n,c}) \le e^{-n(\varphi(c)-c)} \qquad \beta_n(T_{n,c}) \le e^{-n\varphi(c)}$$

$\psi(\alpha) := \log \mathrm{Tr}\,\varrho^\alpha\sigma^{1-\alpha}, \qquad \varphi(c) := \sup_{0<\alpha<1}\{c\alpha - \psi(\alpha)\}$ Legendre transform

# Achievability bound for the direct exponent

**Corollary 53.**

$$\alpha_n(T_{n,c}) \le e^{-n(\varphi(c)-c)} \qquad \beta_n(T_{n,c}) \le e^{-n\varphi(c)}$$

# Achievability bound for the direct exponent

Corollary 53.

$$\alpha_n(T_{n,c}) \leq e^{-n(\varphi(c)-c)} \qquad \beta_n(T_{n,c}) \leq e^{-n\varphi(c)}$$

Direct error exponent:

$$d_r(\varrho\|\sigma) := \sup_{(T_n)_{n\in\mathbb{N}}} \left\{ \liminf_{n\to+\infty} -\frac{1}{n}\log \operatorname{Tr} \varrho_n(I-T_n) :\; \beta_n(T_n) \leq e^{-nr} \right\}$$

# Achievability bound for the direct exponent

**Corollary 53.**

$$\alpha_n(T_{n,c}) \le e^{-n(\varphi(c)-c)} \qquad \beta_n(T_{n,c}) \le e^{-n\varphi(c)}$$

**Direct error exponent:**

$$\begin{aligned} d_r(\varrho\|\sigma) \quad &:= \quad \sup_{(T_n)_{n\in\mathbb{N}}} \left\{ \liminf_{n\to+\infty} -\frac{1}{n}\log \operatorname{Tr}\varrho_n(I-T_n) : \ \beta_n(T_n) \le e^{-nr} \right\} \\ &\ge \quad \sup_c \{\varphi(c) - c : \ \varphi(c) > r\} \end{aligned}$$

# Achievability bound for the direct exponent

Corollary 53.

$$\alpha_n(T_{n,c}) \le e^{-n(\varphi(c)-c)} \qquad \beta_n(T_{n,c}) \le e^{-n\varphi(c)}$$

Direct error exponent:

$$\begin{aligned} d_r(\varrho\|\sigma) &:= \sup_{(T_n)_{n\in\mathbb{N}}} \left\{ \liminf_{n\to+\infty} -\frac{1}{n}\log \operatorname{Tr} \varrho_n(I-T_n) :\ \beta_n(T_n) \le e^{-nr} \right\} \\ &\ge \sup_c \{\varphi(c) - c :\ \varphi(c) > r\} \\ &\overset{\text{exercise}}{=} \sup_{0<\alpha<1} \frac{\alpha-1}{\alpha}\left[r - D_\alpha(\varrho\|\sigma)\right] \end{aligned}$$

(Petz-type) Rényi divergences:

$$D_\alpha(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log \frac{1}{\operatorname{Tr}\varrho} \operatorname{Tr} \varrho^\alpha \sigma^{1-\alpha}, & \alpha \in [0,1] \text{ or } \varrho^0 \le \sigma^0, \\ +\infty, & \text{o.w.} \end{cases}$$

# Optimality bound for the direct exponent

- Nussbaum-Szkoła distributions: ([NSz09])

$$\left.\begin{array}{l}\varrho = \sum_i r_i |e_i\rangle\langle e_i| \\ \sigma = \sum_j s_j |f_j\rangle\langle f_j|\end{array}\right\} \quad \mapsto \quad \left\{\begin{array}{l} p(i,j) := r_i |\langle e_i, f_j\rangle|^2, \\ q(i,j) := s_j |\langle e_i, f_j\rangle|^2 \end{array}\right.$$

# Optimality bound for the direct exponent

- Nussbaum-Szkoła distributions: ([NSz09])

$$\left.\begin{array}{l}\varrho = \sum_i r_i |e_i\rangle\langle e_i| \\ \sigma = \sum_j s_j |f_j\rangle\langle f_j|\end{array}\right\} \quad \mapsto \quad \left\{\begin{array}{l} p(i,j) := r_i |\langle e_i, f_j\rangle|^2, \\ q(i,j) := s_j |\langle e_i, f_j\rangle|^2 \end{array}\right.$$

- Observation: $D_\alpha(\varrho\|\sigma) = D_\alpha(p\|q)$

# Optimality bound for the direct exponent

- Nussbaum-Szkoła distributions: ([NSz09])

$$\left.\begin{array}{l} \varrho = \sum_i r_i |e_i\rangle\langle e_i| \\ \sigma = \sum_j s_j |f_j\rangle\langle f_j| \end{array}\right\} \mapsto \left\{\begin{array}{l} p(i,j) := r_i |\langle e_i, f_j\rangle|^2, \\ q(i,j) := s_j |\langle e_i, f_j\rangle|^2 \end{array}\right.$$

- Observation: $D_\alpha(\varrho\|\sigma) = D_\alpha(p\|q)$

- Optimality bound for the direct exponent: ([NSz09, Nagaoka06])

$$d_r(\varrho\|\sigma) \leq d_r(p\|q)$$

# Optimality bound for the direct exponent

- Nussbaum-Szkoła distributions: ([NSz09])

$$\left.\begin{array}{l}\varrho=\sum_i r_i|e_i\rangle\langle e_i| \\ \sigma=\sum_j s_j|f_j\rangle\langle f_j|\end{array}\right\} \quad\mapsto\quad \left\{\begin{array}{l}p(i,j):=r_i|\langle e_i,f_j\rangle|^2, \\ q(i,j):=s_j|\langle e_i,f_j\rangle|^2\end{array}\right.$$

- Observation: $D_\alpha(\varrho\|\sigma)=D_\alpha(p\|q)$

- Optimality bound for the direct exponent: ([NSz09, Nagaoka06])

$$d_r(\varrho\|\sigma)\le d_r(p\|q) \quad = \quad \sup_{0<\alpha<1}\frac{\alpha-1}{\alpha}\left[r-D_\alpha(p\|q)\right]$$

# Optimality bound for the direct exponent

- Nussbaum-Szkoła distributions: ([NSz09])

$$\left.\begin{array}{l} \varrho = \sum_i r_i |e_i\rangle\langle e_i| \\ \sigma = \sum_j s_j |f_j\rangle\langle f_j| \end{array}\right\} \mapsto \left\{\begin{array}{l} p(i,j) := r_i |\langle e_i, f_j\rangle|^2, \\ q(i,j) := s_j |\langle e_i, f_j\rangle|^2 \end{array}\right.$$

- Observation: $D_\alpha(\varrho\|\sigma) = D_\alpha(p\|q)$

- Optimality bound for the direct exponent: ([NSz09, Nagaoka06])

$$\begin{aligned} d_r(\varrho\|\sigma) \le d_r(p\|q) &= \sup_{0<\alpha<1} \frac{\alpha-1}{\alpha}\left[r - D_\alpha(p\|q)\right] \\ &= \sup_{0<\alpha<1} \frac{\alpha-1}{\alpha}\left[r - D_\alpha(\varrho\|\sigma)\right]. \end{aligned}$$

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ r - D_\alpha(\varrho\|\sigma) \right].$$

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ r - D_\alpha(\varrho\|\sigma) \right].$$

- Operational interpretation of the Petz-type Rényi divergences for $\alpha \in (0, 1)$.

# Quantum Hoeffding bound theorem

**Theorem 54.** For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ r - D_\alpha(\varrho\|\sigma) \right].$$

- Operational interpretation of the Petz-type Rényi divergences for $\alpha \in (0, 1)$.
- More explicit operational interpretation for a fixed $\alpha$: generalized cutoff rates [Csiszár95, MH11].

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha-1}{\alpha}\left[r - D_\alpha(\varrho\|\sigma)\right].$$

- Operational interpretation of the Petz-type Rényi divergences for $\alpha \in (0,1)$.
- More explicit operational interpretation for a fixed $\alpha$: generalized cutoff rates [Csiszár95, MH11].

Exercise 55. Prove that $\alpha \mapsto (\alpha-1)D_\alpha(\varrho\|\sigma)$ is convex. Write it as a Legendre transform, and use the bipolar theorem to prove

$$D_\alpha(\varrho\|\sigma) = \sup_{r>0}\left\{r - \frac{\alpha-1}{\alpha}d_r(\varrho\|\sigma)\right\}, \qquad \alpha \in (0,1).$$

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ r - D_\alpha(\varrho\|\sigma) \right].$$

- Operational interpretation of the Petz-type Rényi divergences for $\alpha \in (0, 1)$.
- More explicit operational interpretation for a fixed $\alpha$: generalized cutoff rates [Csiszár95, MH11].

Exercise 55. Prove that $\alpha \mapsto (\alpha - 1)D_\alpha(\varrho\|\sigma)$ is convex. Write it as a Legendre transform, and use the bipolar theorem to prove

$$D_\alpha(\varrho\|\sigma) = \sup_{r>0} \left\{ r - \frac{\alpha - 1}{\alpha} d_r(\varrho\|\sigma) \right\}, \qquad \alpha \in (0, 1).$$

Use this to prove that $D_\alpha$ is monotone under CPTP maps for $\alpha \in (0, 1)$.

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ r - D_\alpha(\varrho\|\sigma) \right].$$

- Operational interpretation of the Petz-type Rényi divergences for $\alpha \in (0, 1)$.
- More explicit operational interpretation for a fixed $\alpha$: generalized cutoff rates [Csiszár95, MH11].

Exercise 55. Prove that $\alpha \mapsto (\alpha - 1) D_\alpha(\varrho\|\sigma)$ is convex. Write it as a Legendre transform, and use the bipolar theorem to prove

$$D_\alpha(\varrho\|\sigma) = \sup_{r>0} \left\{ r - \frac{\alpha - 1}{\alpha} d_r(\varrho\|\sigma) \right\}, \qquad \alpha \in (0, 1).$$

Use this to prove that $D_\alpha$ is monotone under CPTP maps for $\alpha \in (0, 1)$.

Fully operational proof of the monotonicity of $D_\alpha$.

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha}\left[r - D_\alpha(\varrho\|\sigma)\right].$$

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ r - D_\alpha(\varrho\|\sigma) \right].$$

Exercise 56. Use the convexity of $\alpha \mapsto (\alpha - 1)D_\alpha(\varrho\|\sigma)$ to conclude that $\alpha \mapsto D_\alpha(\varrho\|\sigma)$ is monotone increasing. Show that

$$\lim_{\alpha\to 1} D_\alpha(\varrho\|\sigma) = D_1(\varrho\|\sigma) \qquad (= \text{Umegaki relative entropy}).$$

Conclude that $D_1$ is monotone under CPTP maps.

# Quantum Hoeffding bound theorem

**Theorem 54.** For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[r - D_\alpha(\varrho\|\sigma)\right].$$

**Exercise 56.** Use the convexity of $\alpha \mapsto (\alpha - 1)D_\alpha(\varrho\|\sigma)$ to conclude that $\alpha \mapsto D_\alpha(\varrho\|\sigma)$ is monotone increasing. Show that

$$\lim_{\alpha\to 1} D_\alpha(\varrho\|\sigma) = D_1(\varrho\|\sigma) \qquad (= \text{Umegaki relative entropy}).$$

Conclude that $D_1$ is monotone under CPTP maps.

**Corollary 57.** $\quad d_r(\varrho\|\sigma) > 0 \iff r < D_1(\varrho\|\sigma).$

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[r - D_\alpha(\varrho\|\sigma)\right].$$

Exercise 56. Use the convexity of $\alpha \mapsto (\alpha - 1)D_\alpha(\varrho\|\sigma)$ to conclude that $\alpha \mapsto D_\alpha(\varrho\|\sigma)$ is monotone increasing. Show that

$$\lim_{\alpha\to 1} D_\alpha(\varrho\|\sigma) = D_1(\varrho\|\sigma) \qquad (= \text{Umegaki relative entropy}).$$

Conclude that $D_1$ is monotone under CPTP maps.

Corollary 57. $$d_r(\varrho\|\sigma) > 0 \iff r < D_1(\varrho\|\sigma).$$

Corollary 58. For every type II error rate $r < D_1(\varrho\|\sigma)$, the optimal type I error goes to $0$.

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ r - D_\alpha(\varrho\|\sigma) \right].$$

Exercise 56. Use the convexity of $\alpha \mapsto (\alpha - 1)D_\alpha(\varrho\|\sigma)$ to conclude that $\alpha \mapsto D_\alpha(\varrho\|\sigma)$ is monotone increasing. Show that

$$\lim_{\alpha\to 1} D_\alpha(\varrho\|\sigma) = D_1(\varrho\|\sigma) \qquad (= \text{Umegaki relative entropy}).$$

Conclude that $D_1$ is monotone under CPTP maps.

Corollary 57. $\quad d_r(\varrho\|\sigma) > 0 \iff r < D_1(\varrho\|\sigma)$.

Corollary 58. For every type II error rate $r < D_1(\varrho\|\sigma)$, the optimal type I error goes to $0$. Direct part of Stein's lemma.

# Quantum Hoeffding bound theorem

Theorem 54. For every $r > 0$,

$$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1} \frac{\alpha-1}{\alpha}\left[r - D_\alpha(\varrho\|\sigma)\right].$$

Exercise 56. Use the convexity of $\alpha \mapsto (\alpha-1)D_\alpha(\varrho\|\sigma)$ to conclude that $\alpha \mapsto D_\alpha(\varrho\|\sigma)$ is monotone increasing. Show that

$$\lim_{\alpha\to 1} D_\alpha(\varrho\|\sigma) = D_1(\varrho\|\sigma) \qquad (= \text{Umegaki relative entropy}).$$

Conclude that $D_1$ is monotone under CPTP maps.

Corollary 57. $\quad d_r(\varrho\|\sigma) > 0 \iff r < D_1(\varrho\|\sigma).$

Corollary 58. For every type II error rate $r < D_1(\varrho\|\sigma)$, the optimal type I error goes to $0$. Direct part of Stein's lemma.

What about $r > D_1(\varrho\|\sigma)$? Operational interpretation of $D_\alpha$ for $\alpha > 1$?

# Strong converse exponent

- For type II rates $r > D_1(\varrho\|\sigma)$, we expect the type I error to go to $1$ exponentially fast.

# Strong converse exponent

- For type II rates $r > D_1(\varrho\|\sigma)$, we expect the type I error to go to $1$ exponentially fast. Strong converse property.

# Strong converse exponent

- For type II rates $r > D_1(\varrho\|\sigma)$, we expect the type I error to go to $1$ exponentially fast. Strong converse property.
- Strong converse exponent:

$$sc_r(\varrho\|\sigma) := \inf_{(T_n)_n} \left\{ \limsup_n -\frac{1}{n} \log(1 - \alpha_n(T_n)) : \beta_n(T_n) \leq e^{-nr} \right\}$$

# Strong converse exponent

- For type II rates $r > D_1(\varrho\|\sigma)$, we expect the type I error to go to $1$ exponentially fast. Strong converse property.
- Strong converse exponent:

$$sc_r(\varrho\|\sigma) := \inf_{(T_n)_n} \left\{ \limsup_n -\frac{1}{n} \log(1 - \alpha_n(T_n)) : \beta_n(T_n) \le e^{-nr} \right\}$$

- Classically: $$sc_r(\varrho\|\sigma) = \sup_{1<\alpha} \frac{\alpha - 1}{\alpha} \left[ r - D_\alpha(\varrho\|\sigma) \right] \qquad (2)$$

# Strong converse exponent

- For type II rates $r > D_1(\varrho\|\sigma)$, we expect the type I error to go to $1$ exponentially fast. Strong converse property.
- Strong converse exponent:

$$sc_r(\varrho\|\sigma) := \inf_{(T_n)_n} \left\{ \limsup_n -\frac{1}{n} \log(1 - \alpha_n(T_n)) : \beta_n(T_n) \le e^{-nr} \right\}$$

- Classically: $$sc_r(\varrho\|\sigma) = \sup_{1<\alpha} \frac{\alpha - 1}{\alpha} [r - D_\alpha(\varrho\|\sigma)] \qquad (2)$$

- Exercise 59. Prove that if (2) held in the quantum case then $D_\alpha$ would be monotone for every $\alpha > 1$

# Strong converse exponent

- For type II rates $r > D_1(\varrho\|\sigma)$, we expect the type I error to go to $1$ exponentially fast. Strong converse property.
- Strong converse exponent:

$$sc_r(\varrho\|\sigma) := \inf_{(T_n)_n} \left\{ \limsup_n -\frac{1}{n} \log(1-\alpha_n(T_n)) : \beta_n(T_n) \leq e^{-nr} \right\}$$

- Classically: $$sc_r(\varrho\|\sigma) = \sup_{1<\alpha} \frac{\alpha-1}{\alpha} [r - D_\alpha(\varrho\|\sigma)] \qquad (2)$$

- Exercise 59. Prove that if (2) held in the quantum case then $D_\alpha$ would be monotone for every $\alpha > 1$

$\Longrightarrow Q_\alpha(\varrho\|\sigma) := \mathrm{Tr}\,\varrho^\alpha\sigma^{1-\alpha}$ would be jointly convex

# Strong converse exponent

- For type II rates $r > D_1(\varrho\|\sigma)$, we expect the type I error to go to $1$ exponentially fast. Strong converse property.
- Strong converse exponent:

$$sc_r(\varrho\|\sigma) := \inf_{(T_n)_n} \left\{ \limsup_n -\frac{1}{n} \log(1-\alpha_n(T_n)) : \beta_n(T_n) \leq e^{-nr} \right\}$$

- Classically: $$sc_r(\varrho\|\sigma) = \sup_{1<\alpha} \frac{\alpha-1}{\alpha} [r - D_\alpha(\varrho\|\sigma)] \qquad (2)$$

- Exercise 59. Prove that if (2) held in the quantum case then $D_\alpha$ would be monotone for every $\alpha > 1$

$\Longrightarrow Q_\alpha(\varrho\|\sigma) := \operatorname{Tr} \varrho^\alpha \sigma^{1-\alpha}$ would be jointly convex

$\Longrightarrow t \mapsto t^\alpha$ would be operator convex, for all $\alpha > 1$.

# Strong converse exponent

- For type II rates $r > D_1(\varrho\|\sigma)$, we expect the type I error to go to $1$ exponentially fast. Strong converse property.
- Strong converse exponent:

$$sc_r(\varrho\|\sigma) := \inf_{(T_n)_n}\left\{\limsup_n -\frac{1}{n}\log(1-\alpha_n(T_n)) : \beta_n(T_n) \le e^{-nr}\right\}$$

- Classically: $$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha(\varrho\|\sigma)\right] \qquad (2)$$

- Exercise 59. Prove that if (2) held in the quantum case then $D_\alpha$ would be monotone for every $\alpha > 1$

$\implies Q_\alpha(\varrho\|\sigma) := \mathrm{Tr}\,\varrho^\alpha\sigma^{1-\alpha}$ would be jointly convex

$\implies t \mapsto t^\alpha$ would be operator convex, for all $\alpha > 1$.

- However, $t \mapsto t^\alpha$ is not operator convex for $t > 2$ [Bhatia]

# Strong converse exponent

- For type II rates $r > D_1(\varrho\|\sigma)$, we expect the type I error to go to $1$ exponentially fast. Strong converse property.
- Strong converse exponent:

$$sc_r(\varrho\|\sigma) := \inf_{(T_n)_n} \left\{ \limsup_n -\frac{1}{n} \log(1-\alpha_n(T_n)) : \beta_n(T_n) \le e^{-nr} \right\}$$

- Classically: $$sc_r(\varrho\|\sigma) = \sup_{1<\alpha} \frac{\alpha-1}{\alpha} \left[ r - D_\alpha(\varrho\|\sigma) \right] \qquad (2)$$

- Exercise 59. Prove that if (2) held in the quantum case then $D_\alpha$ would be monotone for every $\alpha > 1$

$\Longrightarrow Q_\alpha(\varrho\|\sigma) := \operatorname{Tr} \varrho^\alpha \sigma^{1-\alpha}$ would be jointly convex

$\Longrightarrow t \mapsto t^\alpha$ would be operator convex, for all $\alpha > 1$.

- However, $t \mapsto t^\alpha$ is not operator convex for $t > 2$ [Bhatia]

$\Longrightarrow$ (2) cannot hold with the Petz-type Rényi divergences.

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0 \le \sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0 \le \sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0 \le \sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

Operational interpretation of $D_\alpha^*$ for $\alpha > 1$

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1),\\ +\infty, & \text{o.w.}\end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

Operational interpretation of $D_\alpha^*$ for $\alpha > 1$

- Reminder: $$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1}\frac{\alpha-1}{\alpha}\left[r - D_\alpha(\varrho\|\sigma)\right].$$

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1),\\ +\infty, & \text{o.w.}\end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

Operational interpretation of $D_\alpha^*$ for $\alpha>1$

- Reminder: $$d_r(\varrho\|\sigma) = \sup_{0<\alpha<1}\frac{\alpha-1}{\alpha}\left[r - D_\alpha(\varrho\|\sigma)\right].$$

- Operationally motivated definition of quantum Rényi divergence:

$$D_\alpha^{(q)} := \begin{cases} D_\alpha(\varrho\|\sigma), & \alpha\in(0,1),\\ D_\alpha^*(\varrho\|\sigma), & \alpha>1.\end{cases}$$

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0 \le \sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

Operational interpretation of $D_\alpha^*$ for $\alpha > 1$.

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

  Operational interpretation of $D_\alpha^*$ for $\alpha > 1$.

- Remark: Monotonicity is used in the proof (following Nagaoka's method), and had been proved by various matrix analytic methods. ([Beigi13, FL13])

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

Operational interpretation of $D_\alpha^*$ for $\alpha > 1$.

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

Operational interpretation of $D_\alpha^*$ for $\alpha > 1$.

- Proposition 61. ([MLDSzFT13])

$$\lim_{\alpha\to+\infty} D_\alpha^*(\varrho\|\sigma) = D_\infty^*(\varrho\|\sigma)$$

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0 \leq \sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

  Operational interpretation of $D_\alpha^*$ for $\alpha > 1$.

- Proposition 61. ([MLDSzFT13])

$$\lim_{\alpha\to+\infty} D_\alpha^*(\varrho\|\sigma) = D_\infty^*(\varrho\|\sigma)$$

- Remark: $D_{1/2}^*(\varrho\|\sigma) = -2\log F(\varrho\|\sigma) + 2\log\operatorname{Tr}\varrho$ as before.

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0 \le \sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

  Operational interpretation of $D_\alpha^*$ for $\alpha > 1$.

- Proposition 61. ([MLDSzFT13])

$$\lim_{\alpha\to+\infty} D_\alpha^*(\varrho\|\sigma) = D_\infty^*(\varrho\|\sigma)$$

- Remark: $D_{1/2}^*(\varrho\|\sigma) = -2\log F(\varrho\|\sigma) + 2\log\operatorname{Tr}\varrho$ as before.

  No operational interpretation.

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

Operational interpretation of $D_\alpha^*$ for $\alpha > 1$.

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

  Operational interpretation of $D_\alpha^*$ for $\alpha > 1$.

- Relation of the two Rényi divegences:

$$D_\alpha(\varrho\|\sigma) \quad \ge \quad D_\alpha^*(\varrho\|\sigma)$$

  Araki-Lieb-Thirring inequality

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

  Operational interpretation of $D_\alpha^*$ for $\alpha > 1$.

- Relation of the two Rényi divegences:

$$\begin{aligned} D_\alpha(\varrho\|\sigma) &\ge D_\alpha^*(\varrho\|\sigma) && \text{Araki-Lieb-Thirring inequality} \\ &\ge \alpha D_\alpha(\varrho\|\sigma), \quad \alpha\in[0,1] && \text{([IRS17])} \end{aligned}$$

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D^*_\alpha(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1),\\ +\infty, & \text{o.w.}\end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D^*_\alpha(\varrho\|\sigma)\right].$$

Operational interpretation of $D^*_\alpha$ for $\alpha > 1$.

- Relation of the two Rényi divegences:

$$\begin{aligned} D_\alpha(\varrho\|\sigma) &\ge D^*_\alpha(\varrho\|\sigma) && \text{Araki-Lieb-Thirring inequality}\\ &\ge \alpha D_\alpha(\varrho\|\sigma), \quad \alpha\in[0,1] && \text{([IRS17])}\end{aligned}$$

- Corollary: $\lim_{\alpha\to 1} D^*_\alpha(\varrho\|\sigma) = D_1(\varrho\|\sigma)$. (Also by direct calculation.)

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

Operational interpretation of $D_\alpha^*$ for $\alpha>1$

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0 \le \sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

  Operational interpretation of $D_\alpha^*$ for $\alpha > 1$

- Proposition 62. $\alpha \mapsto D_\alpha^*(\varrho\|\sigma)$ is also monotone.

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

Operational interpretation of $D_\alpha^*$ for $\alpha > 1$

- Proposition 62. $\alpha\mapsto D_\alpha^*(\varrho\|\sigma)$ is also monotone.

- Corollary 63. $sc_r(\varrho\|\sigma) > 0 \iff r > D_1(\varrho\|\sigma)$.

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0 \le \sigma^0 \text{ or } \alpha \in (0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

  Operational interpretation of $D_\alpha^*$ for $\alpha > 1$

- Proposition 62. $\alpha \mapsto D_\alpha^*(\varrho\|\sigma)$ is also monotone.

- Corollary 63. $sc_r(\varrho\|\sigma) > 0 \iff r > D_1(\varrho\|\sigma)$.

- Corollary 64. For every type II error rate $r > D_1(\varrho\|\sigma)$, the optimal type I error goes to $1$.

# Sandwiched Rényi divergences

- Sandwiched Rényi divergences: ([WWY13, MLDSzFT13])

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^{\alpha}, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

- Theorem 60. ([MO15])

$$sc_r(\varrho\|\sigma) = \sup_{1<\alpha}\frac{\alpha-1}{\alpha}\left[r - D_\alpha^*(\varrho\|\sigma)\right].$$

  Operational interpretation of $D_\alpha^*$ for $\alpha > 1$

- Proposition 62. $\alpha \mapsto D_\alpha^*(\varrho\|\sigma)$ is also monotone.

- Corollary 63. $sc_r(\varrho\|\sigma) > 0 \iff r > D_1(\varrho\|\sigma)$.

- Corollary 64. For every type II error rate $r > D_1(\varrho\|\sigma)$, the optimal type I error goes to $1$.

  Strong converse part of Stein's lemma.

# Relative entropy and Stein's lemma

**Theorem 65.** Quantum Stein's lemma ([HP91, ON00])

The optimal rate of the type II error under the assumption that the type I error goes to $0$ is the relative entropy $D_1(\varrho\|\sigma)$.

Operational interpretation of the relative entropy.

# VI. Rényi information measures

# Properties of the Rényi divergences

Definition: (Reminder)

(Petz-type) Rényi divergences:

$$D_\alpha(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\varrho^\alpha\sigma^{1-\alpha}, & \alpha\in[0,1] \text{ or } \varrho^0\le\sigma^0, \\ +\infty, & \text{o.w.} \end{cases}$$

Sandwiched Rényi divergences:

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

# Properties of the Rényi divergences

Definition: (Reminder)

(Petz-type) Rényi divergences:

$$D_\alpha(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\varrho^\alpha\sigma^{1-\alpha}, & \alpha\in[0,1] \text{ or } \varrho^0\le\sigma^0,\\ +\infty, & \text{o.w.}\end{cases}$$

Sandwiched Rényi divergences:

$$D_\alpha^*(\varrho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho}\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1),\\ +\infty, & \text{o.w.}\end{cases}$$

Exercise 66. They are strictly positive, additive, stable, and satisfy the logarithmic scaling property for all $\alpha>0$.

# Properties of the Rényi divergences

Definition:

(Petz-type) trace quantities:

$$Q_\alpha(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\varrho^\alpha\sigma^{1-\alpha}, & \alpha\in[0,1] \text{ or } \varrho^0\le\sigma^0, \\ +\infty, & \text{o.w.} \end{cases}$$

Sandwiched trace quantities:

$$Q_\alpha^*(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

# Properties of the Rényi divergences

Definition:

(Petz-type) trace quantities:

$$Q_\alpha(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\varrho^\alpha\sigma^{1-\alpha}, & \alpha\in[0,1] \text{ or } \varrho^0\le\sigma^0, \\ +\infty, & \text{o.w.} \end{cases}$$

Sandwiched trace quantities:

$$Q_\alpha^*(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

Exercise 67. They are homogenous and have the direct sum property for all $\alpha\ge 0$.

# Properties of the Rényi divergences

Definition:

(Petz-type) trace quantities:

$$Q_\alpha(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\varrho^\alpha\sigma^{1-\alpha}, & \alpha\in[0,1] \text{ or } \varrho^0\le\sigma^0, \\ +\infty, & \text{o.w.} \end{cases}$$

Sandwiched trace quantities:

$$Q_\alpha^*(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1), \\ +\infty, & \text{o.w.} \end{cases}$$

Exercise 67. They are homogenous and have the direct sum property for all $\alpha\ge 0$. Hence $Q_\alpha^\#$ jointly convex $\Longleftrightarrow$ monotone

# Properties of the Rényi divergences

Definition:

(Petz-type) trace quantities:

$$Q_\alpha(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\varrho^\alpha\sigma^{1-\alpha}, & \alpha\in[0,1] \text{ or } \varrho^0\le\sigma^0,\\ +\infty, & \text{o.w.}\end{cases}$$

Sandwiched trace quantities:

$$Q_\alpha^*(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1),\\ +\infty, & \text{o.w.}\end{cases}$$

Exercise 67. They are homogenous and have the direct sum property for all $\alpha\ge 0$. Hence $Q_\alpha^\#$ jointly convex $\Longleftrightarrow$ monotone $\Longleftrightarrow$ $D_\alpha^\#$ monotone.

# Properties of the Rényi divergences

Definition:

(Petz-type) trace quantities:

$$Q_\alpha(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\varrho^\alpha\sigma^{1-\alpha}, & \alpha\in[0,1] \text{ or } \varrho^0\le\sigma^0,\\ +\infty, & \text{o.w.}\end{cases}$$

Sandwiched trace quantities:

$$Q_\alpha^*(\varrho\|\sigma) := \begin{cases} \operatorname{sgn}(\alpha-1)\operatorname{Tr}\left(\varrho^{1/2}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{1/2}\right)^\alpha, & \varrho^0\le\sigma^0 \text{ or } \alpha\in(0,1),\\ +\infty, & \text{o.w.}\end{cases}$$

Exercise 67. They are homogenous and have the direct sum property for all $\alpha\ge 0$. Hence $Q_\alpha^\#$ jointly convex $\iff$ monotone $\iff$ $D_\alpha^\#$ monotone.

Theorem 68. $D_\alpha$ monotone $\iff \alpha\in[0,2]$,

$D_\alpha^*$ monotone $\iff \alpha\in[1/2,+\infty]$. ([Tomamichel])

# Rényi entropies

Definition 69. Rényi $\alpha$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\alpha(A)_\varrho \quad := \quad -D_\alpha(\varrho \| I_A)$$

# Rényi entropies

Definition 69. Rényi $\alpha$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$H_\alpha(A)_\varrho \quad := \quad -D_\alpha(\varrho \| I_A) = \frac{1}{1-\alpha} \log \operatorname{Tr} \varrho^\alpha$$

# Rényi entropies

**Definition 69.** Rényi $\alpha$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$\begin{aligned} H_\alpha(A)_\varrho &:= -D_\alpha(\varrho \| I_A) = \frac{1}{1-\alpha} \log \operatorname{Tr} \varrho^\alpha \\ &= -D_\alpha^*(\varrho \| I_A) =: H_\alpha^*(A)_\varrho \end{aligned}$$

# Rényi entropies

Definition 69. Rényi $\alpha$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$\begin{aligned} H_\alpha(A)_\varrho &:= -D_\alpha(\varrho \| I_A) = \frac{1}{1-\alpha} \log \operatorname{Tr} \varrho^\alpha \\ &= -D_\alpha^*(\varrho \| I_A) =: H_\alpha^*(A)_\varrho \end{aligned}$$

Exercise 70. $\lim_{\alpha \to 1} H_\alpha(A)_\varrho = H_1(A)_\varrho = -\operatorname{Tr} \varrho \log \varrho$ von Neumann entropy.

# Rényi entropies

Definition 69. Rényi $\alpha$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$\begin{aligned} H_\alpha(A)_\varrho &:= -D_\alpha(\varrho \| I_A) = \frac{1}{1-\alpha} \log \operatorname{Tr} \varrho^\alpha \\ &= -D^*_\alpha(\varrho \| I_A) =: H^*_\alpha(A)_\varrho \end{aligned}$$

Exercise 70. $\lim_{\alpha \to 1} H_\alpha(A)_\varrho = H_1(A)_\varrho = -\operatorname{Tr} \varrho \log \varrho$ von Neumann entropy.

Exercise 71. $\forall \alpha \in [0, +\infty] \quad H_\alpha$ is additive, Schur-concave,

$$0 \le H_\alpha(A)_\varrho \le \log |A|,$$

with equality for pure/maximally mixed states, respectively.

# Rényi entropies

Definition 69. Rényi $\alpha$-entropy of a state $\varrho \in \mathcal{S}(\mathcal{H}_A)$:

$$
\begin{aligned}
H_\alpha(A)_\varrho &:= -D_\alpha(\varrho \| I_A) = \frac{1}{1-\alpha} \log \operatorname{Tr} \varrho^\alpha \\
&= -D_\alpha^*(\varrho \| I_A) =: H_\alpha^*(A)_\varrho
\end{aligned}
$$

Exercise 70. $\lim_{\alpha \to 1} H_\alpha(A)_\varrho = H_1(A)_\varrho = -\operatorname{Tr} \varrho \log \varrho$ von Neumann entropy.

Exercise 71. $\forall \alpha \in [0, +\infty]$ $\quad H_\alpha$ is additive, Schur-concave,

$$
0 \leq H_\alpha(A)_\varrho \leq \log |A|,
$$

with equality for pure/maximally mixed states, respectively.

Exercise 72. $H_\alpha$ is concave $\Longleftrightarrow \alpha \in [0, 1]$.

# Quantum state compression

- Problem: How to quantify the information content of a (mixed) quantum state $\varrho$?

# Quantum state compression

- Problem: How to quantify the information content of a (mixed) quantum state $\varrho$?
- Ensemble picture: $\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$

# Quantum state compression

- Problem: How to quantify the information content of a (mixed) quantum state $\varrho$?
- Ensemble picture: $\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$

  Quantum source emits signal sequences $\psi_{i_1} \otimes \ldots \otimes \psi_{i_n}$ with probability $p_{i_1} \cdot \ldots \cdot p_{i_n}$

# Quantum state compression

- Problem: How to quantify the information content of a (mixed) quantum state $\varrho$?
- Ensemble picture: $\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$

  Quantum source emits signal sequences $\psi_{i_1} \otimes \ldots \otimes \psi_{i_n}$ with probability $p_{i_1} \cdot \ldots \cdot p_{i_n}$

  Encoding/compression: $\mathcal{E}_n : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathbb{C}^2)^{\otimes k_n}$ CPTP

  Decoding: $\mathcal{D}_n : \mathcal{B}(\mathbb{C}^2)^{\otimes k_n} \to \mathcal{B}(\mathcal{H}^{\otimes n})$ CPTP

# Quantum state compression

- Problem: How to quantify the information content of a (mixed) quantum state $\varrho$?
- Ensemble picture: $\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$

  Quantum source emits signal sequences $\psi_{i_1} \otimes \ldots \otimes \psi_{i_n}$ with probability $p_{i_1} \cdot \ldots \cdot p_{i_n}$

  Encoding/compression: $\mathcal{E}_n : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathbb{C}^2)^{\otimes k_n}$ CPTP

  Decoding: $\mathcal{D}_n : \mathcal{B}(\mathbb{C}^2)^{\otimes k_n} \to \mathcal{B}(\mathcal{H}^{\otimes n})$ CPTP

  Fidelity criterion:

$$\hat{F}_n^e := \sum_i p_i \left(1 - F(|\psi_i\rangle\langle\psi_i| \| (\mathcal{D}_n \circ \mathcal{E}_n)(|\psi_i\rangle\langle\psi_i|)\right))$$

# Quantum state compression

- Problem: How to quantify the information content of a (mixed) quantum state $\varrho$?
- Ensemble picture: $\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$

  Quantum source emits signal sequences $\psi_{i_1} \otimes \ldots \otimes \psi_{i_n}$ with probability $p_{i_1} \cdot \ldots \cdot p_{i_n}$

  Encoding/compression: $\mathcal{E}_n : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathbb{C}^2)^{\otimes k_n}$ CPTP

  Decoding: $\mathcal{D}_n : \mathcal{B}(\mathbb{C}^2)^{\otimes k_n} \to \mathcal{B}(\mathcal{H}^{\otimes n})$ CPTP

  Fidelity criterion:

  $$\hat{F}_n^e := \sum_i p_i \left(1 - F(|\psi_i\rangle\langle\psi_i| \| (\mathcal{D}_n \circ \mathcal{E}_n)(|\psi_i\rangle\langle\psi_i|)\right))$$

- Purification picture: $\varrho$ is the marginal of $|\psi_\varrho\rangle$ on $RA$

# Quantum state compression

- Problem: How to quantify the information content of a (mixed) quantum state $\varrho$?
- Ensemble picture: $\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$

  Quantum source emits signal sequences $\psi_{i_1} \otimes \ldots \otimes \psi_{i_n}$ with probability $p_{i_1} \cdot \ldots \cdot p_{i_n}$

  Encoding/compression: $\mathcal{E}_n : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathbb{C}^2)^{\otimes k_n}$ CPTP

  Decoding: $\mathcal{D}_n : \mathcal{B}(\mathbb{C}^2)^{\otimes k_n} \to \mathcal{B}(\mathcal{H}^{\otimes n})$ CPTP

  Fidelity criterion:

  $$\hat{F}_n^e := \sum_i p_i \left(1 - F(|\psi_i\rangle\langle\psi_i| \| (\mathcal{D}_n \circ \mathcal{E}_n)(|\psi_i\rangle\langle\psi_i|)\right))$$

- Purification picture: $\varrho$ is the marginal of $|\psi_\varrho\rangle$ on $RA$

  Encoding and decoding as above.

  Fidelity criterion: $\hat{F}_n^p := 1 - F\left(|\psi_\varrho\rangle\langle\psi_\varrho| \| (\mathrm{id}_R \otimes \mathcal{D}_n \circ \mathcal{E}_n)|\psi_\varrho\rangle\langle\psi_\varrho|\right)$

# Quantum state compression

- Idea: Look at the problem as state discrimination between $\varrho$ and $\pi_A$.

  Use the Neyman-Pearson projections $T_{n,c} = \{\varrho^{\otimes n} - e^{nc}\pi_A^{\otimes n} > 0\}$ to compress as

  $$\mathcal{E}_n(.) := T_{n,c}(.)T_{n,c} + |0\rangle\langle 0| \operatorname{Tr}(.)(I - T_{n,c}), \qquad \operatorname{ran} T_{n,c} \cong (\mathbb{C}^2)^{\otimes k_n}$$

  Decoding trivial.

# Quantum state compression

- Idea: Look at the problem as state discrimination between $\varrho$ and $\pi_A$.

  Use the Neyman-Pearson projections $T_{n,c} = \{\varrho^{\otimes n} - e^{nc}\pi_A^{\otimes n} > 0\}$ to compress as

  $$\mathcal{E}_n(.) := T_{n,c}(.)T_{n,c} + |0\rangle\langle 0| \operatorname{Tr}(.)(I - T_{n,c}), \qquad \operatorname{ran} T_{n,c} \cong (\mathbb{C}^2)^{\otimes k_n}$$

  Decoding trivial.

- Exercise 73. Show that $\hat{F}_n^e \le \hat{F}_n^p \le \operatorname{Tr} \varrho^{\otimes n}(I - T_{n,c}) = \alpha_n(T_{n,c})$

# Quantum state compression

- Idea: Look at the problem as state discrimination between $\varrho$ and $\pi_A$.

  Use the Neyman-Pearson projections $T_{n,c} = \{\varrho^{\otimes n} - e^{nc}\pi_A^{\otimes n} > 0\}$ to compress as

  $$\mathcal{E}_n(.) := T_{n,c}(.)T_{n,c} + |0\rangle\langle 0| \operatorname{Tr}(.)(I - T_{n,c}), \qquad \operatorname{ran} T_{n,c} \cong (\mathbb{C}^2)^{\otimes k_n}$$

  Decoding trivial.

- Exercise 73. Show that $\hat{F}_n^e \leq \hat{F}_n^p \leq \operatorname{Tr} \varrho^{\otimes n}(I - T_{n,c}) = \alpha_n(T_{n,c})$

- Corollary 74. For every coding rate $R = \lim_{n\to+\infty} \frac{k_n}{n}$,

  $$F_n^e \leq \hat{F}_n^p \leq e^{-n \sup_{0<\alpha<1} \frac{\alpha-1}{\alpha}[-R+H_\alpha(A)_\varrho]}$$

# Quantum state compression

- Idea: Look at the problem as state discrimination between $\varrho$ and $\pi_A$.

  Use the Neyman-Pearson projections $T_{n,c} = \{\varrho^{\otimes n} - e^{nc}\pi_A^{\otimes n} > 0\}$ to compress as

  $$\mathcal{E}_n(.) := T_{n,c}(.)T_{n,c} + |0\rangle\langle 0| \operatorname{Tr}(.)(I - T_{n,c}), \qquad \operatorname{ran} T_{n,c} \cong (\mathbb{C}^2)^{\otimes k_n}$$

  Decoding trivial.

- Exercise 73. Show that $\hat{F}_n^e \leq \hat{F}_n^p \leq \operatorname{Tr} \varrho^{\otimes n}(I - T_{n,c}) = \alpha_n(T_{n,c})$

- Corollary 74. For every coding rate $R = \lim_{n\to+\infty} \frac{k_n}{n}$,

  $$F_n^e \leq \hat{F}_n^p \leq e^{-n \sup_{0<\alpha<1} \frac{\alpha-1}{\alpha}[-R+H_\alpha(A)_\varrho]}$$

- Theorem 75. ([Hayashi02]) This is sharp, the exact direct error exponent is

  $$d(R) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha}\left[-R + H_\alpha(A)_\varrho\right], \qquad R > 0.$$

# Quantum state compression

- Idea: Look at the problem as state discrimination between $\varrho$ and $\pi_A$.

  Use the Neyman-Pearson projections $T_{n,c} = \{\varrho^{\otimes n} - e^{nc}\pi_A^{\otimes n} > 0\}$ to compress as

  $$\mathcal{E}_n(.) := T_{n,c}(.)T_{n,c} + |0\rangle\langle 0| \operatorname{Tr}(.)(I - T_{n,c}), \qquad \operatorname{ran} T_{n,c} \cong (\mathbb{C}^2)^{\otimes k_n}$$

  Decoding trivial.

- Exercise 73. Show that $\hat{F}_n^e \leq \hat{F}_n^p \leq \operatorname{Tr} \varrho^{\otimes n}(I - T_{n,c}) = \alpha_n(T_{n,c})$

- Corollary 74. For every coding rate $R = \lim_{n\to+\infty} \frac{k_n}{n}$,

  $$F_n^e \leq \hat{F}_n^p \leq e^{-n \sup_{0<\alpha<1} \frac{\alpha-1}{\alpha}[-R+H_\alpha(A)_\varrho]}$$

- Theorem 75. ([Hayashi02]) This is sharp, the exact direct error exponent is (Operational interpretation of $H_\alpha(A)_\varrho$)

  $$d(R) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha}\left[-R + H_\alpha(A)_\varrho\right], \qquad R > 0.$$

- Theorem 75. ([Hayashi02]) This is sharp, the exact direct error exponent is

$$d(R) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ -R + H_\alpha(A)_\varrho \right], \qquad R > 0.$$

- Theorem 75. ([Hayashi02]) This is sharp, the exact direct error exponent is

$$d(R) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ -R + H_\alpha(A)_\varrho \right], \qquad R > 0.$$

- Strong converse: $\hat{F}_n^e \to 1$ for $R > H_1(A)_\varrho$.

  Schumacher compression, operational interpretation of the von Neumann entropy.

# Quantum state compression

- Theorem 75. ([Hayashi02]) This is sharp, the exact direct error exponent is

$$d(R) = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[-R + H_\alpha(A)_\varrho\right], \qquad R > 0.$$

- Strong converse: $\hat{F}_n^e \to 1$ for $R > H_1(A)_\varrho$.

  Schumacher compression, operational interpretation of the von Neumann entropy.

- The strong converse exponent is not known.

# Entanglement concentration

- Goal: Extract $k_n$ copies of EPR pairs from $n$ copies of
$$|\psi\rangle = \sum_x \sqrt{P(x)}|x\rangle_A|x\rangle_B$$
using LOCC operations.

# Entanglement concentration

- Goal: Extract $k_n$ copies of EPR pairs from $n$ copies of

$$|\psi\rangle = \sum_x \sqrt{P(x)}|x\rangle_A|x\rangle_B$$

  using LOCC operations.

- Theorem 76. ([HKMMW02]) Direct exponent: Success probability goes to $1$ with an exponent $r$, optimal extraction rate is

$$d(r) = \sup_{\alpha>1}\left\{\frac{r}{1-\alpha} + H_\alpha(A)_{|\psi\rangle\langle\psi|}\right\}$$

# Entanglement concentration

- Goal: Extract $k_n$ copies of EPR pairs from $n$ copies of
$$|\psi\rangle = \sum_x \sqrt{P(x)}|x\rangle_A|x\rangle_B$$
using LOCC operations.

- Theorem 76. ([HKMMW02]) Direct exponent: Success probability goes to $1$ with an exponent $r$, optimal extraction rate is
$$d(r) = \sup_{\alpha>1}\left\{\frac{r}{1-\alpha} + H_\alpha(A)_{|\psi\rangle\langle\psi|}\right\}$$

  Strong converse exponent: Success probability goes to $0$ with an exponent $r$, optimal extraction rate is
$$sc(r) = \inf_{0<\alpha<1}\left\{\frac{\alpha r}{1-\alpha} + H_\alpha(A)_{|\psi\rangle\langle\psi|}\right\}$$

# Entanglement concentration

Extension to the problem of extracting $k_n$ copies of $|\phi\rangle$ from $n$ copies of $|\psi\rangle$:

$$sc(r) = \inf_{0<\alpha<1} \left\{ \frac{\alpha r + H_\alpha(A)_{|\psi\rangle\langle\psi|}}{H_\alpha(A)_{|\phi\rangle\langle\phi|}} \right\}$$

Poster 61: Asger Kjaerulff Jensen, Péter Vrana, *Asymptotic LOCC transformations and the asymptotic spectrum*

# Conditional Rényi entropy

- Reminder:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$

# Conditional Rényi entropy

- Reminder:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$

- Quantum Sibson identity: ([SW13]) $\overline{\sigma} := \frac{(\mathrm{Tr}_A\, \varrho^\alpha)^{1/\alpha}}{\mathrm{Tr}(\ldots)}$

$$D_\alpha(\varrho_{AB} \| I_A \otimes \sigma_B) = \frac{\alpha}{\alpha - 1} \log \mathrm{Tr}\left(\mathrm{Tr}_A\, \varrho^\alpha\right)^{1/\alpha} + D_\alpha(\overline{\sigma} \| \sigma_B),$$

# Conditional Rényi entropy

- Reminder:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$

- Quantum Sibson identity: ([SW13]) $\quad \overline{\sigma} := \frac{(\mathrm{Tr}_A\, \varrho^\alpha)^{1/\alpha}}{\mathrm{Tr}(\ldots)}$

$$D_\alpha(\varrho_{AB} \| I_A \otimes \sigma_B) = \frac{\alpha}{\alpha - 1} \log \mathrm{Tr}\left(\mathrm{Tr}_A\, \varrho^\alpha\right)^{1/\alpha} + D_\alpha(\overline{\sigma} \| \sigma_B),$$

- Corollary: $H_\alpha(A|B)_\varrho = \frac{\alpha}{\alpha-1} \log \mathrm{Tr}\left(\mathrm{Tr}_A\, \varrho^\alpha\right)^{1/\alpha}$

# Conditional Rényi entropy

- Reminder:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$

- Quantum Sibson identity: ([SW13]) $\quad \overline{\sigma} := \frac{(\mathrm{Tr}_A\, \varrho^\alpha)^{1/\alpha}}{\mathrm{Tr}(\dots)}$

$$D_\alpha(\varrho_{AB} \| I_A \otimes \sigma_B) = \frac{\alpha}{\alpha - 1} \log \mathrm{Tr}\left(\mathrm{Tr}_A\, \varrho^\alpha\right)^{1/\alpha} + D_\alpha(\overline{\sigma} \| \sigma_B),$$

- Corollary: $H_\alpha(A|B)_\varrho = \frac{\alpha}{\alpha-1} \log \mathrm{Tr}\left(\mathrm{Tr}_A\, \varrho^\alpha\right)^{1/\alpha}$ Additive.

# Conditional Rényi entropy

- Reminder:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$

- Quantum Sibson identity: ([SW13]) $\overline{\sigma} := \frac{(\operatorname{Tr}_A \varrho^\alpha)^{1/\alpha}}{\operatorname{Tr}(\ldots)}$

$$D_\alpha(\varrho_{AB} \| I_A \otimes \sigma_B) = \frac{\alpha}{\alpha - 1} \log \operatorname{Tr} \left(\operatorname{Tr}_A \varrho^\alpha\right)^{1/\alpha} + D_\alpha(\overline{\sigma} \| \sigma_B),$$

- Corollary: $H_\alpha(A|B)_\varrho = \frac{\alpha}{\alpha-1} \log \operatorname{Tr} \left(\operatorname{Tr}_A \varrho^\alpha\right)^{1/\alpha}$ Additive.

- No such explicit expression for the conditional $H_\alpha^*$.

# Conditional Rényi entropy

- Reminder:

$$H_\Delta(A|B)_\varrho := -\inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \Delta(\varrho_{AB} \| I_A \otimes \sigma_B)$$

- Quantum Sibson identity: ([SW13]) $\overline{\sigma} := \frac{(\mathrm{Tr}_A\, \varrho^\alpha)^{1/\alpha}}{\mathrm{Tr}(\ldots)}$

$$D_\alpha(\varrho_{AB} \| I_A \otimes \sigma_B) = \frac{\alpha}{\alpha - 1} \log \mathrm{Tr}\, (\mathrm{Tr}_A\, \varrho^\alpha)^{1/\alpha} + D_\alpha(\overline{\sigma} \| \sigma_B),$$

- Corollary: $H_\alpha(A|B)_\varrho = \frac{\alpha}{\alpha-1} \log \mathrm{Tr}\, (\mathrm{Tr}_A\, \varrho^\alpha)^{1/\alpha}$ Additive.

- No such explicit expression for the conditional $H_\alpha^*$. Additivity?

# Conditional Rényi entropy

- Duality relations: ([Tomamichel])

  $H^*_\alpha$ and $H^*_\beta$ are dual for $\frac{1}{\alpha} + \frac{1}{\beta} = 2$, $\alpha, \beta \in [1/2, +\infty]$

# Conditional Rényi entropy

- Duality relations: ([Tomamichel])

  $H_\alpha^*$ and $H_\beta^*$ are dual for $\frac{1}{\alpha} + \frac{1}{\beta} = 2,\ \alpha, \beta \in [1/2, +\infty]$

  $H_\alpha^\downarrow$ and $H_\beta^\downarrow$ are dual for $\alpha + \beta = 2,\ \ \alpha, \beta \in [0, 2]$

# Conditional Rényi entropy

- Duality relations: ([Tomamichel])

$H_\alpha^*$ and $H_\beta^*$ are dual for $\frac{1}{\alpha} + \frac{1}{\beta} = 2,\ \alpha, \beta \in [1/2, +\infty]$

$H_\alpha^\downarrow$ and $H_\beta^\downarrow$ are dual for $\alpha + \beta = 2,\ \ \alpha, \beta \in [0, 2]$

$H_\alpha$ and $H_\beta^{*\downarrow}$ are dual for $\alpha\beta = 1,\ \ \alpha, \beta \in [0, +\infty]$

# Conditional Rényi entropy

- Duality relations: ([Tomamichel])

  $H_\alpha^*$ and $H_\beta^*$ are dual for $\frac{1}{\alpha} + \frac{1}{\beta} = 2,\ \alpha, \beta \in [1/2, +\infty]$

  $H_\alpha^\downarrow$ and $H_\beta^\downarrow$ are dual for $\alpha + \beta = 2,\ \ \alpha, \beta \in [0, 2]$

  $H_\alpha$ and $H_\beta^{*\downarrow}$ are dual for $\alpha\beta = 1,\ \ \alpha, \beta \in [0, +\infty]$

- Corollary: Conditional $H_\alpha^*$ is additive for $\alpha \in [1/2, +\infty]$.

# Conditional Rényi entropy

- Duality relations: ([Tomamichel])

$$H^*_\alpha \text{ and } H^*_\beta \quad \text{are dual for} \quad \frac{1}{\alpha}+\frac{1}{\beta}=2,\ \alpha,\beta\in[1/2,+\infty]$$

$$H^\downarrow_\alpha \text{ and } H^\downarrow_\beta \quad \text{are dual for} \quad \alpha+\beta=2,\ \ \alpha,\beta\in[0,2]$$

$$H_\alpha \text{ and } H^{*\downarrow}_\beta \quad \text{are dual for} \quad \alpha\beta=1,\ \ \alpha,\beta\in[0,+\infty]$$

- Corollary: Conditional $H^*_\alpha$ is additive for $\alpha\in[1/2,+\infty]$.
- Corollary: Alternative proof for the additivity of the conditional $H_\alpha$.

# Conditional Rényi entropy

- Duality relations: ([Tomamichel])

  $H_\alpha^*$ and $H_\beta^*$ are dual for $\frac{1}{\alpha}+\frac{1}{\beta}=2,\ \alpha,\beta\in[1/2,+\infty]$

  $H_\alpha^\downarrow$ and $H_\beta^\downarrow$ are dual for $\alpha+\beta=2,\ \ \alpha,\beta\in[0,2]$

  $H_\alpha$ and $H_\beta^{*\downarrow}$ are dual for $\alpha\beta=1,\ \ \alpha,\beta\in[0,+\infty]$

- Corollary: Conditional $H_\alpha^*$ is additive for $\alpha\in[1/2,+\infty]$.
- Corollary: Alternative proof for the additivity of the conditional $H_\alpha$.
- Additivity of the Rényi mutual informations $I_\alpha$ and $I_\alpha^*$ can be established by similar techniques ([HT16]).

# Conditional Rényi entropy

Operational intrepretation of the conditional Rényi entropies: ([HT16])

State discrimination

$$\varrho_{AB}^{\otimes n} \qquad \text{v.s.} \qquad \pi_A^{\otimes n} \otimes \sigma, \qquad \sigma \in \mathcal{S}(\mathcal{H}^{\otimes n})$$

# Conditional Rényi entropy

Operational intrepretation of the conditional Rényi entropies: ([HT16])

State discrimination

$$\varrho_{AB}^{\otimes n} \qquad \text{v.s.} \qquad \pi_A^{\otimes n} \otimes \sigma, \qquad \sigma \in \mathcal{S}(\mathcal{H}^{\otimes n})$$

direct exponent:

$$d_r = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[r - \log|A| + H_\alpha(A|B)_\varrho\right]$$

strong converse exponent:

$$sc_r = \sup_{1<\alpha} \frac{\alpha - 1}{\alpha} \left[r - \log|A| + H_\alpha^*(A|B)_\varrho\right]$$

# Rényi mutual information

Operational intrepretation of the Rényi mutual informations: ([HT16])

State discrimination

$$\varrho_{AB}^{\otimes n} \qquad \text{v.s.} \qquad \varrho_A^{\otimes n} \otimes \sigma, \qquad \sigma \in \mathcal{S}(\mathcal{H}^{\otimes n})$$

direct exponent:

$$d_r = \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[ r - I_\alpha(A:B)_\varrho \right]$$

strong converse exponent:

$$sc_r = \sup_{1<\alpha} \frac{\alpha - 1}{\alpha} \left[ r - I_\alpha^*(A:B)_\varrho \right]$$

# Rényi divergence radii

- Classical-quantum channel: $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$

# Rényi divergence radii

- Classical-quantum channel: $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$
- Weighted Rényi divergence radii: $P \in \mathcal{P}_f(\mathcal{X})$

$$R^{\#}_{\alpha,P}(\operatorname{ran} W) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sum_{x \in \mathcal{X}} P(x) D^{\#}_{\alpha}(W(x) \| \sigma)$$

# Rényi divergence radii

- Classical-quantum channel: $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$
- Weighted Rényi divergence radii: $P \in \mathcal{P}_f(\mathcal{X})$

$$R^{\#}_{\alpha,P}(\operatorname{ran} W) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sum_{x \in \mathcal{X}} P(x) D^{\#}_{\alpha}(W(x)\|\sigma)$$

- Theorem 77. ([CGH18, MO18]) $\overline{\sigma}$ is a minimizer

for $R_{\alpha,P}$ iff $\overline{\sigma} = \sum_x P(x) \frac{1}{Q_\alpha(W(x)\|\sigma)} \sigma^{\frac{1-\alpha}{2}} W(x)^\alpha \sigma^{\frac{1-\alpha}{2}}$

and for $R^*_{\alpha,P}$ iff

$$\overline{\sigma} = \sum_x P(x) \frac{1}{Q^*_\alpha(W(x)\|\sigma)} \sigma^{\frac{1-\alpha}{2\alpha}} W(x)^{1/2} \left[ W(x)^{1/2} \sigma^{\frac{1-\alpha}{\alpha}} W(x)^{1/2} \right]^{\alpha-1} W(x)^{1/2} \sigma^{\frac{1-\alpha}{2\alpha}}$$

# Rényi divergence radii

- Classical-quantum channel: $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$
- Weighted Rényi divergence radii: $P \in \mathcal{P}_f(\mathcal{X})$

$$R^{\#}_{\alpha,P}(\operatorname{ran} W) := \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sum_{x\in\mathcal{X}} P(x) D^{\#}_{\alpha}(W(x)\|\sigma)$$

- Theorem 77. ([CGH18, MO18]) $\overline{\sigma}$ is a minimizer

for $R_{\alpha,P}$ iff $$\overline{\sigma} = \sum_x P(x) \frac{1}{Q_\alpha(W(x)\|\sigma)} \sigma^{\frac{1-\alpha}{2}} W(x)^\alpha \sigma^{\frac{1-\alpha}{2}}$$

and for $R^*_{\alpha,P}$ iff

$$\overline{\sigma} = \sum_x P(x) \frac{1}{Q^*_\alpha(W(x)\|\sigma)} \sigma^{\frac{1-\alpha}{2\alpha}} W(x)^{1/2} \left[W(x)^{1/2} \sigma^{\frac{1-\alpha}{\alpha}} W(x)^{1/2}\right]^{\alpha-1} W(x)^{1/2} \sigma^{\frac{1-\alpha}{2\alpha}}$$

- Corollary 78. Additivity: $R^{\#}_{\alpha,P^{\otimes n}}(\operatorname{ran} W^{\otimes n}) = n R^{\#}_{\alpha,P}(\operatorname{ran} W)$.

# Rényi divergence radii

- Constant composition coding: messages $1, \ldots, M_n$

  encoding: $k \mapsto \underline{x}^{(k)} \in \mathcal{X}^n$

  decoding: POVM on $\mathcal{H}^{\otimes n}$

# Rényi divergence radii

- Constant composition coding: messages $1, \ldots, M_n$

  encoding: $k \mapsto \underline{x}^{(k)} \in \mathcal{X}^n$

  decoding: POVM on $\mathcal{H}^{\otimes n}$

  for all $k$: $\frac{1}{n}\#\{i : x_i^{(k)} = x\} = P_n(x)$, $\quad P_n \to P$

# Rényi divergence radii

- Constant composition coding: messages $1, \ldots, M_n$

  encoding: $k \mapsto \underline{x}^{(k)} \in \mathcal{X}^n$

  decoding: POVM on $\mathcal{H}^{\otimes n}$

  for all $k$: $\frac{1}{n}\#\{i : x_i^{(k)} = x\} = P_n(x)$, $\quad P_n \to P$

- Theorem 79. ([MO18]) For all rates $R = \lim_n \frac{1}{n} \log M_n > R_{1,P}(\operatorname{ran} W)$, the average success probability decays with an exponent

$$sc(R, W, P) = \sup_{\alpha > 1} \frac{\alpha - 1}{\alpha} \left[R - R^*_{\alpha,P}(\operatorname{ran} W)\right]$$

Operational interpretation of the weighted sandwiched Rényi divergence radii for $\alpha > 1$.

# Rényi divergence radii

- Constant composition coding: messages $1, \ldots, M_n$

  encoding: $k \mapsto \underline{x}^{(k)} \in \mathcal{X}^n$

  decoding: POVM on $\mathcal{H}^{\otimes n}$

  for all $k$: $\frac{1}{n}\#\{i : x_i^{(k)} = x\} = P_n(x)$, $\quad P_n \to P$

- Theorem 79. ([MO18]) For all rates $R = \lim_n \frac{1}{n} \log M_n > R_{1,P}(\operatorname{ran} W)$, the average success probability decays with an exponent

$$sc(R, W, P) = \sup_{\alpha > 1} \frac{\alpha - 1}{\alpha} \left[R - R^*_{\alpha,P}(\operatorname{ran} W)\right]$$

  Operational interpretation of the weighted sandwiched Rényi divergence radii for $\alpha > 1$.

  Additivity of the weighted divergence radius is crucial for the proof.

# Rényi divergence radii

- Constant composition coding: messages $1, \ldots, M_n$

  encoding: $k \mapsto \underline{x}^{(k)} \in \mathcal{X}^n$

  decoding: POVM on $\mathcal{H}^{\otimes n}$

  for all $k$: $\frac{1}{n}\#\{i : x_i^{(k)} = x\} = P_n(x)$, $\quad P_n \to P$

# Rényi divergence radii

- Constant composition coding: messages $1, \ldots, M_n$

  encoding: $k \mapsto \underline{x}^{(k)} \in \mathcal{X}^n$

  decoding: POVM on $\mathcal{H}^{\otimes n}$

  for all $k$: $\frac{1}{n}\#\{i : x_i^{(k)} = x\} = P_n(x)$, $\quad P_n \to P$

- Theorem 80. ([MO17]) Without composition constraint:

$$sc(R, W) = \sup_{\alpha > 1} \frac{\alpha - 1}{\alpha} [R - R_\alpha^*(\operatorname{ran} W)]$$

Operational interpretation of the sandwiched Rényi divergence radii for $\alpha > 1$.

Reminder:

$$R_\alpha^*(\operatorname{ran} W) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_x D_\alpha^*(W(x)\|\sigma) = \sup_{P \in \mathcal{P}_f(\mathcal{X})} R_{\alpha,P}^*(\operatorname{ran} W)$$

# Rényi divergence radii

- What about the direct exponent and Rényi divergence radii for $\alpha < 1$?

# Rényi divergence radii

- What about the direct exponent and Rényi divergence radii for $\alpha < 1$?
- Constant composition coding with rates

  $$R = \lim_n \frac{1}{n} \log M_n < R_{1,P}(\operatorname{ran} W):$$

  The error probability goes to zero with a rate $d(R, W, P)$.

# Rényi divergence radii

- What about the direct exponent and Rényi divergence radii for $\alpha < 1$?
- Constant composition coding with rates

  $R = \lim_n \frac{1}{n} \log M_n < R_{1,P}(\operatorname{ran} W)$:

  The error probability goes to zero with a rate $d(R, W, P)$.

  Classical-quantum sphere packing bound: ([DW17])

$$d(R, W, P) \leq \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[R - R_{\alpha,P}(\operatorname{ran} W)\right]$$

# Rényi divergence radii

- What about the direct exponent and Rényi divergence radii for $\alpha < 1$?
- Constant composition coding with rates

  $$R = \lim_n \frac{1}{n} \log M_n < R_{1,P}(\operatorname{ran} W):$$

  The error probability goes to zero with a rate $d(R, W, P)$.

  Classical-quantum sphere packing bound: ([DW17])

  $$d(R, W, P) \leq \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} [R - R_{\alpha,P}(\operatorname{ran} W)]$$

- Becomes $+\infty$ at $R_{0,P}(\operatorname{ran} W)$.

# Rényi divergence radii

- What about the direct exponent and Rényi divergence radii for $\alpha < 1$?
- Constant composition coding with rates

  $R = \lim_n \frac{1}{n} \log M_n < R_{1,P}(\operatorname{ran} W)$:

  The error probability goes to zero with a rate $d(R, W, P)$.

  Classical-quantum sphere packing bound: ([DW17])

$$d(R, W, P) \leq \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} \left[R - R_{\alpha,P}(\operatorname{ran} W)\right]$$

- Becomes $+\infty$ at $R_{0,P}(\operatorname{ran} W)$.
- Optimizing over all channels with the same confusability graph gives Marton's weighted version of the Lovász $\theta$-function:

$$\theta(G, P) = \inf_W R_{0,P}(\operatorname{ran} W).$$

# Rényi divergence radii

- What about the direct exponent and Rényi divergence radii for $\alpha < 1$?
- Constant composition coding with rates

  $R = \lim_n \frac{1}{n} \log M_n < R_{1,P}(\operatorname{ran} W)$:

  The error probability goes to zero with a rate $d(R, W, P)$.

  Classical-quantum sphere packing bound: ([DW17])

$$d(R, W, P) \leq \sup_{0<\alpha<1} \frac{\alpha - 1}{\alpha} [R - R_{\alpha,P}(\operatorname{ran} W)]$$

- Becomes $+\infty$ at $R_{0,P}(\operatorname{ran} W)$.
- Optimizing over all channels with the same confusability graph gives Marton's weighted version of the Lovász $\theta$-function:

$$\theta(G, P) = \inf_W R_{0,P}(\operatorname{ran} W).$$

- Taking the maximum over all $P$ gives the Lovász $\theta$-function.

# Channel divergences

- Let $\mathcal{M},\mathcal{N}: A \to B$ be CPTP, $\Delta$ a divergence. Channel $\Delta$ divergence:

$$\Delta(\mathcal{M}\|\mathcal{N}) := \sup_{\varrho_{RA}} \Delta\left((\mathrm{id}_R \otimes M)\varrho_{RA} \| (\mathrm{id}_R \otimes N)\varrho_{RA}\right)$$

# Channel divergences

- Let $\mathcal{M}, \mathcal{N} : A \to B$ be CPTP, $\Delta$ a divergence. Channel $\Delta$ divergence:

$$\Delta(\mathcal{M}\|\mathcal{N}) := \sup_{\varrho_{RA}} \Delta\left((\mathrm{id}_R \otimes M)\varrho_{RA} \| (\mathrm{id}_R \otimes N)\varrho_{RA}\right)$$

- When $\mathcal{M}(.) = \varrho\,\mathrm{Tr}(.)$ and $\mathcal{N}(.) = \sigma\,\mathrm{Tr}(.)$ are replacer channels then

$$\Delta(\mathcal{M}\|\mathcal{N}) = \Delta(\varrho\|\sigma)$$

for every stable $\Delta$.

# Channel divergences

- Let $\mathcal{M}, \mathcal{N} : A \to B$ be CPTP, $\Delta$ a divergence. Channel $\Delta$ divergence:

$$\Delta(\mathcal{M}\|\mathcal{N}) := \sup_{\varrho_{RA}} \Delta\left((\mathrm{id}_R \otimes M)\varrho_{RA} \| (\mathrm{id}_R \otimes N)\varrho_{RA}\right)$$

- When $\mathcal{M}(.) = \varrho \operatorname{Tr}(.)$ and $\mathcal{N}(.) = \sigma \operatorname{Tr}(.)$ are replacer channels then

$$\Delta(\mathcal{M}\|\mathcal{N}) = \Delta(\varrho\|\sigma)$$

  for every stable $\Delta$.

- The Rényi channel divergences give the direct and strong converse exponents in quantum channel discrimination if only product strategies are allowed.

# Channel divergences

- Theorem 81. ([CMW16, BHKW18]) For channel discrimination with adaptive strategies

$$sc_r(\mathcal{M}\|\mathcal{N}) = \sup_{\alpha>1} \frac{\alpha-1}{\alpha} \left[r - D^*_\alpha(\mathcal{M}\|\mathcal{N})\right]$$

  if $\mathcal{N}$ is a replacer channel, or both channels are classical-quantum.

  Moreover, in these cases Stein's lemma holds with $D_1(\mathcal{M}\|\mathcal{N})$ as exponent.

# Channel divergences

- Theorem 81. ([CMW16, BHKW18]) For channel discrimination with adaptive strategies

$$sc_r(\mathcal{M}\|\mathcal{N}) = \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[r - D^*_\alpha(\mathcal{M}\|\mathcal{N})\right]$$

  if $\mathcal{N}$ is a replacer channel, or both channels are classical-quantum.

  Moreover, in these cases Stein's lemma holds with $D_1(\mathcal{M}\|\mathcal{N})$ as exponent.

- Other types of information measures for channels?

# Channel divergences

- Theorem 81. ([CMW16, BHKW18]) For channel discrimination with adaptive strategies

$$sc_r(\mathcal{M}\|\mathcal{N}) = \sup_{\alpha>1} \frac{\alpha - 1}{\alpha} \left[r - D^*_\alpha(\mathcal{M}\|\mathcal{N})\right]$$

  if $\mathcal{N}$ is a replacer channel, or both channels are classical-quantum.

  Moreover, in these cases Stein's lemma holds with $D_1(\mathcal{M}\|\mathcal{N})$ as exponent.

- Other types of information measures for channels?

  Poster 2: Gilad Gour, Mark Wilde, *Entropy of a quantum channel*

# Further topics

- Entropy and thermodynamics: Characterization of possible state transitions in terms of Rényi entropies/ divergences.

  Poster 182: Paul Boes, Jens Eisert, Rodrigo Gallego, Markus Mueller and Henrik Wilming, *Von Neumann entropy from unitarity*

# Further topics

- Entropy and thermodynamics: Characterization of possible state transitions in terms of Rényi entropies/ divergences.

  Poster 182: Paul Boes, Jens Eisert, Rodrigo Gallego, Markus Mueller and Henrik Wilming, *Von Neumann entropy from unitarity*

- Smooth entropies, applications in information theory, cryptography, thermodynamics.

  Poster 94: Anurag Anshu, Mario Berta, Kun Fang, Rahul Jain, Marco Tomamichel and Xin Wang, *Smooth entropies for quantum channels and multipartite states*

AM14 Koenraad M. R. Audenaert and Milán Mosonyi, *Upper bounds on the error probabilities and asymptotic error exponents in quantum multiple state discrimination* J. Math. Phys., 55, 102201, 2014, arXiv:1401.7658

Beigi13 Salman Beigi, *Sandwiched Rényi Divergence Satisfies Data Processing Inequality*, Journal of Mathematical Physics, 54:12, 122202, 2013

BSW15 Mario Berta, Kaushik P. Seshadreesan, Mark M. Wilde, *Rényi generalizations of the conditional quantum mutual information*, Journal of Mathematical Physics, 56:2, 022205, 2015

BHKW18 Mario Berta, Christoph Hirche, Eneet Kaur, Mark M. Wilde, *Amortized Channel Divergence for Asymptotic Quantum Channel Discrimination*, arXiv:1808.01498, 2018

Bhatia Rajendra Bhatia, *Matrix Analysis*, Springer, 169, 1997

CGH18 Hao-Chung Cheng, Li Gao, Min-Hsiu Hsieh, *Properties of Noncommutative Rényi and Augustin Information*, arXiv:1811.04218, 2018

CCYZ12 Patrick J. Coles, Roger Colbeck, Li Yu and Michael Zwolak, *Uncertainty Relations from Simple Entropic Properties*, Phys. Rev. Lett., 108, 210405, 2012

CBTW17 Patrick J. Coles, Mario Berta, Marco Tomamichel, Stephanie Wehner, *Entropic uncertainty relations and their applications*, Rev. Mod. Phys., 89, 015002, 2017

CMW16 Tom Cooney, Milán Mosonyi, Mark M. Wilde, *Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication*, Communications in Mathematical Physics, 344, 3, 797–829, 2016

CsK Imre Csiszár and János Körner, *Information theory: coding theorems for discrete memoryless channels*, 2nd ed., Cambridge University Press, 2011

Csiszar95 Imre Csiszár, *Generalized Cutoff Rates and Rényi's Information Measures*, IEEE Transactions on Information Theory, 41:1, 26–34 1995

DW17 Marco Dalai, Andreas Winter, *Constant Compositions in the Sphere Packing Bound for Classical-Quantum Channels*, IEEE Transactions on Information Theory, 63:9, 5603–5617, 2017

Datta09 Nilanjana Datta, *Min- and Max-Relative Entropies and a New Entanglement Monotone*, IEEE Transactions on Information Theory, 55:6, 2816-2826, 2009

FL13 Rupert L. Frank and Elliott H. Lieb, *Monotonicity of a relative Rényi entropy*, Journal of Mathematical Physics, 54:12, 122201, 2013

# Bibliography


FR15 Omar Fawzi and Renato Renner, *Quantum conditional mutual information and approximate Markov chains*, Communications in Mathematical Physics, 340, 575-611, 2015

Hayashi02 Masahito Hayashi, *Exponents of quantum fixed-length pure state source coding*, Physical Review A, 66, 032321, 2002

HKMMW02 Masahito Hayashi, Masato Koashi, Keiji Matsumoto, Fumiaki Morikoshi, Andreas Winter, *Error exponents for entanglement concentration*, Journal of Physics A: Mathematical and General, 36:2, 527, 2002,

HT16 Masahito Hayashi, Marco Tomamichel, *Correlation detection and an operational interpretation of the Rényi mutual information*, Journal of Mathematical Physics, 57, 102201, 2016

# Bibliography


HJPW04 P. Hayden, R. Jozsa, D. Petz and A. Winter, *Structure of states which satisfy strong subadditivity of quantum entropy with equality*, Communications in Mathematical Physics, 246, 359–374, 2004

HP91 Fumio Hiai and Dénes Petz, *The proper formula for relative entropy and its asymptotics in quantum probability*, Communications in Mathematical Physics, 143:1, 99-114, 1991

Hiai10 F. Hiai, *Matrix Analysis: Matrix Monotone Functions, Matrix Means, and Majorization*, Interdisciplinary Information Sciences, 16, 139-248, 2010

HOW05 Michał Horodecki, Jonathan Oppenheim and Andreas Winter, *Partial quantum information*, Nature, 436, 673–676, 2005

HOW07 Michał Horodecki, Jonathan Oppenheim and Andreas Winter, *Quantum state merging and negative information*, Communications in Mathematical Physics, 269, 1, 107-136, 2007

# Bibliography


IRS17 Raban Iten, Joseph M. Renes, David Sutter, *Pretty Good Measures in Quantum Information Theory*, IEEE Transactions on Information Theory, 63:2, 1270–1279, 2017

KRS09 Robert König, Renato Renner and Christian Schaffner, *The Operational Meaning of Min- and Max-Entropy*, IEEE Transactions on Information Theory, 55:9, 4337-4347, 2009

KW09 Robert König and Stephanie Wehner, *A strong converse for classical channel coding using entangled inputs*, Physical Review Letters, 103:7, 070504, 2009

LS09 Debbie Leung, Graeme Smith, *Continuity of Quantum Channel Capacities*, Commun. Math. Phys., 292, 201–215, 2009

LR73 Elliott H. Lieb, Mary Beth Ruskai, *A Fundamental Property of Quantum-Mechanical Entropy*, Phys. Rev. Lett., 30, 434, 1973

# Bibliography


MH11 Milán Mosonyi and Fumio Hiai, *On the quantum Rényi relative entropies and related capacity formulas*, IEEE Transactions on Information Theory 57:4, 2474–2487, 2011

MO15 Milán Mosonyi and Tomohiro Ogawa, *Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies*, Communications in Mathematical Physics, 334:1, 1617–1648, 2015

MO17 Milán Mosonyi, Tomohiro Ogawa, *Strong converse exponent for classical-quantum channel coding*, Communications in Mathematical Physics, 355:1, 373-426, 2017

MO18 Milán Mosonyi and Tomohiro Ogawa, *Divergence radii and the strong converse exponent of classical-quantum channel coding with constant compositions*, arXiv:1811.10599, 2018

# Bibliography


MLDSzFT13 Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, Marco Tomamichel, *On quantum Rényi entropies: A new generalization and some properties*, Journal of Mathematical Physics, 54:12, 122203, 2013,

Nagaoka06 Hiroshi Nagaoka, *The converse part of the theorem for Quantum Hoeffding bound*, arXiv:quant-ph/0611289, 2006,

NC Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000

NSz09 Michael Nussbaum, Arleta Szkoła, *The Chernoff Lower Bound for Symmetric Quantum Hypothesis Testing*, The Annals of Statistics, 37, 1040-1057, 2009,

ON00 Tomohiro Ogawa and Hiroshi Nagaoka, *Strong Converse and Stein's Lemma in Quantum Hypothesis Testing*, IEEE Transactions on Information Theory, 46:7, 2428–2433, 2000

Renner05 Renato Renner, *Security of Quantum Key Distribution*, Swiss Federal Institute of Technology Zurich, Diss. ETH No. 16242, 2005

SW13 Naresh Sharma and Naqueeb Ahmad Warsi, *Fundamental Bound on the Reliability of Quantum Information Transmission*, Phys. Rev. Lett., 110, 080501, 2013,

Tomamichel Marco Tomamichel, *Quantum Information Processing with Finite Resources*, SpringerBriefs in Mathematical Physics, 2016

Umegaki62 H. Umegaki, *Conditional expectation in an operator algebra*, Kodai Math. Sem. Rep., 14, 59-85, 1962

WWY13 Mark M. Wilde, Andreas Winter, Dong Yang, *Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy*, Communications in Mathematical Physics, 331:2, 593–622, 2014

# Bibliography


Winter16 Andreas Winter, *Tight Uniform Continuity Bounds for Quantum Entropies: Conditional Entropy, Relative Entropy Distance and Energy Constraints*, Commun. Math. Phys., 45:7, 291–313, 2016

YD09 Jon T. Yard, Igor Devetak, *Optimal Quantum Source Coding With Quantum Side Information at the Encoder and Decoder*, IEEE Transactions on Information Theory, 55:11, 5339–5351, 2009

YKL75 H.P. Yuen and R.S. Kennedy and M. Lax, *Optimum Testing of Multiple Hypotheses in Quantum Detection Theory*, IEEE Trans. Inform. Theory, 21:2, 125–134, 1975