Quantum divergences via convex optimization

Omar Fawzi

IPAM Entropy Inequalities, Quantum Information and Quantum Physics workshop

Based on joint works with Peter Brown and Hamza Fawzi
`arXiv:2007.12575` and `arXiv:2007.12576`

Rényi divergence

- Let $p, q \in \mathbb{R}_+^n$ and $\alpha \geq 0$, Rényi divergence

$$D_\alpha(p\|q) = \frac{1}{\alpha - 1} \log \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}.$$

- **Data-processing inequality:** $D_\alpha(Mp\|Mq) \leq D_\alpha(p\|q)$ for a channel M

  Consequence of convexity/concavity of $(a, b) \mapsto a^\alpha b^{1-\alpha}$
  concave for $\alpha \in [0, 1]$ and convex for $\alpha > 1$

- **How to extend D_α to positive semidefinite operators $\rho, \sigma \geq 0$?**
  Requirement: Data-processing inequality (DPI)
  Optional: Retain properties of classical D_α

Quantum Rényi divergences

Recall: $D_\alpha(p\|q) = \frac{1}{\alpha-1}\log Q_\alpha(p\|q)$ with $Q_\alpha(p\|q) = \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}$

- [Petz, 1986] Petz: $\overline{Q}_\alpha(\rho\|\sigma) = \mathrm{tr}\left(\rho^\alpha \sigma^{1-\alpha}\right)$
  DPI for $\alpha \in [0,2]$
  Uses: Hypothesis testing [MH, 2011]
- [MDSFT, WWY 2013] sandwiched: $\widetilde{Q}_\alpha(\rho\|\sigma) = \mathrm{tr}\left(\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right)$
  DPI for $\alpha \in [\frac{1}{2},\infty]$
  Uses: Hypothesis testing [MO, 2014], Entropy accumulation [DFR, 2016]
- [Matsumoto, 2013] geometric: $\widehat{Q}_\alpha(\rho\|\sigma) = \mathrm{tr}\left(\sigma \#_\alpha \rho\right)$
  DPI for $\alpha \in [0,2]$
  Uses: Channel capacity bounds [FF, 2019]
- [Donald, 1986] measured: $D_\alpha^{\mathbb{M}}(\rho\|\sigma) = \max_{\text{measurement } \mathcal{M}} D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma))$
  DPI for $\alpha \in [0,\infty]$
  Uses: Recoverability bounds [BHOS, 2014]

This talk: there are more!

The Rényi divergence as a convex program

For the rest of the talk $\alpha > 1$

Observation: $$Q_\alpha(p\|q) = \inf_{a \geq 0} \left\{ \sum_{i=1}^{n} a_i \; : \; q_i^{1-\frac{1}{\alpha}} a_i^{\frac{1}{\alpha}} \geq p_i \; \forall i \right\}$$

$\{(a, x, y) \in \mathbb{R}_+^3 : x^{1-\frac{1}{\alpha}} a^{\frac{1}{\alpha}} \geq y\}$ is a convex set
$\implies Q_\alpha(p\|q)$ is jointly convex
$\implies D_\alpha(p\|q)$ satisfies DPI

The Rényi divergence as a convex program

For the rest of the talk $\alpha > 1$

Observation: $$Q_\alpha(p\|q) = \inf_{a \geq 0} \left\{ \sum_{i=1}^{n} a_i \ : \ q_i^{1-\frac{1}{\alpha}} a_i^{\frac{1}{\alpha}} \geq p_i \, \forall i \right\}$$

$\{(a, x, y) \in \mathbb{R}_+^3 : x^{1-\frac{1}{\alpha}} a^{\frac{1}{\alpha}} \geq y\}$ is a convex set
$\Longrightarrow Q_\alpha(p\|q)$ is jointly convex
$\Longrightarrow D_\alpha(p\|q)$ satisfies DPI

Quantum generalization: Use matrix geometric mean in place of $x^{1-\frac{1}{\alpha}} a^{\frac{1}{\alpha}}$
For $A, B > 0$, the β-geometric mean is [Kubo-Ando]

$$A \#_\beta B = A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^\beta A^{1/2}$$

For $\beta \in [0, 1]$, $\{(A, X, Y) \in \mathrm{Pos}(\mathbb{C}^n)^3 : X \#_\beta A \geq Y\}$ is a convex set

$$Q_\alpha^{\#}(\rho\|\sigma) \ = \ \inf_{A \geq 0} \{ \mathrm{tr}\,(A) \ : \ \sigma \#_{1/\alpha} A \geq \rho \}$$

The $\#$-divergence

$$D_\alpha^\#(\rho\|\sigma) := \frac{1}{\alpha-1}\log Q_\alpha^\#(\rho\|\sigma) \qquad Q_\alpha^\#(\rho\|\sigma) \;:=\; \inf_{A\geq 0}\{\mathrm{tr}\,(A) \;:\; \sigma\#_{1/\alpha}A \geq \rho\}$$

- Convex optimization problem. Semidefinite program when α rational
- $Q_\alpha^\#(\rho\|\sigma)$ jointly convex in $(\rho,\sigma) \Rightarrow$ DPI
- If ρ,σ commute $Q_\alpha^\#(\rho\|\sigma) = Q_\alpha(\rho\|\sigma)$
- $D_\alpha^\#(\rho\|\sigma) \geq 0$ with equality iff $\rho=\sigma$

The $\#$-divergence

$$D_\alpha^\#(\rho\|\sigma) := \frac{1}{\alpha-1}\log Q_\alpha^\#(\rho\|\sigma) \qquad Q_\alpha^\#(\rho\|\sigma) \ := \ \inf_{A\geq 0}\{\mathrm{tr}\,(A) \ : \ \sigma\#_{1/\alpha}A \geq \rho\}$$

- Convex optimization problem. Semidefinite program when α rational
- $Q_\alpha^\#(\rho\|\sigma)$ jointly convex in $(\rho,\sigma) \Rightarrow$ DPI
- If ρ,σ commute $Q_\alpha^\#(\rho\|\sigma) = Q_\alpha(\rho\|\sigma)$
- $D_\alpha^\#(\rho\|\sigma) \geq 0$ with equality iff $\rho = \sigma$

Interesting properties:

- Dual form useful for computing conditional entropies for quantum correlations (NPA hierarchy)

$$Q_\alpha^\#(\rho\|\sigma) = \max_{C,D}\Big\{\mathrm{tr}\,(C\rho) + \mathrm{tr}\,(D\sigma) \quad \text{s.t. non-commutative poly. constraints on } C \text{ and } D\Big\}$$

- Chain rule and additivity properties for quantum channels
- Regularizes to the sandwiched Rényi divergence

Comparison with geometric $\widehat{D}_\alpha$

We always have $D_\alpha^{\#}(\rho\|\sigma) \leq \widehat{D}_\alpha(\rho\|\sigma)$ $= \frac{1}{\alpha-1}\log \mathrm{tr}\,(\sigma\#_\alpha\rho)$

$$\begin{aligned}\mathrm{tr}\,(\sigma\#_\alpha\rho) &= \min\{\mathrm{tr}\,(A) \;:\; A \geq \sigma\#_\alpha\rho\} \\ &\geq \min\{\mathrm{tr}\,(A) \;:\; \sigma\#_{1/\alpha}A \geq \rho\}\end{aligned}$$

Comparison with geometric $\widehat{D}_\alpha$

We always have $D_\alpha^\#(\rho\|\sigma) \le \widehat{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log \mathrm{tr}\,(\sigma\#_\alpha\rho)$

$$
\begin{aligned}
\mathrm{tr}\,(\sigma\#_\alpha\rho) &= \min\{\mathrm{tr}\,(A)\ :\ A \ge \sigma\#_\alpha\rho\} \\
&\ge \min\{\mathrm{tr}\,(A)\ :\ \sigma\#_{1/\alpha}A \ge \rho\} \\
&= Q_\alpha^\#(\rho\|\sigma)
\end{aligned}
$$

x^α is *not* operator monotone
Inequality can be strict

Comparison with geometric $\widehat{D}_\alpha$

We always have $D^{\#}_\alpha(\rho\|\sigma) \le \widehat{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log \operatorname{tr}(\sigma\#_\alpha\rho)$

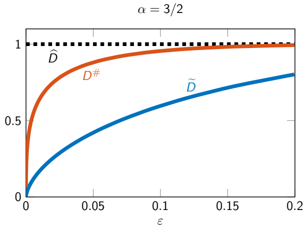
$$\begin{aligned}\operatorname{tr}\,(\sigma\#_\alpha\rho) &= \min\{\operatorname{tr}(A) \;:\; A \ge \sigma\#_\alpha\rho\}\\ &\ge \min\{\operatorname{tr}(A) \;:\; \sigma\#_{1/\alpha}A \ge \rho\}\\ &= Q^{\#}_\alpha(\rho\|\sigma)\end{aligned}$$

x^α is *not* operator monotone
Inequality can be strict

Difference can be large:

- For pure $\rho = |\phi\rangle\langle\phi|$, $\widehat{D}_\alpha(\rho\|\sigma) = \log(\langle\phi|\sigma^{-1}\phi\rangle) = D_{\max}(\rho\|\sigma)$.
- Let $\rho = |\phi\rangle\langle\phi|$ with $|\phi\rangle = \sqrt{\varepsilon}|00\rangle_{XY} + \sqrt{1-\varepsilon}|11\rangle_{XY}$, and $\sigma = I_X \otimes \rho_Y$

$$\widehat{D}_\alpha(\rho\|\sigma) = \begin{cases} 1 & \varepsilon > 0 \\ 0 & \varepsilon = 0\end{cases}$$



Comparison with sandwiched $\widetilde{D}_\alpha$

Remark: $D^{\#}_\alpha$ is *not* additive under tensor product, but it is subadditive

$$D^{\#}_\alpha(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) \leq D^{\#}_\alpha(\rho_1\|\sigma_1) + D^{\#}_\alpha(\rho_2\|\sigma_2)$$

Regularization:

$$\lim_{n\to\infty} \frac{1}{n} D^{\#}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) \;=\; \widetilde{D}_\alpha(\rho\|\sigma)$$

Proof: Use measured divergence $D^{\mathbb{M}}_\alpha(\rho\|\sigma) = \max_{\mathcal{M}} D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma))$ and $\lim_{n\to\infty} \frac{1}{n} D^{\mathbb{M}}_\alpha(\rho\|\sigma) = \widetilde{D}_\alpha(\rho\|\sigma)$

$$D^{\mathbb{M}}_\alpha(\rho\|\sigma) \;\leq\; D^{\#}_\alpha(\rho\|\sigma) \;\leq\; D^{\mathbb{M}}_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha-1}\log|\mathrm{spec}(\sigma)|.$$

Comparison with sandwiched $\widetilde{D}_\alpha$

Remark: $D_\alpha^\#$ is *not* additive under tensor product, but it is subadditive

$$D_\alpha^\#(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) \leq D_\alpha^\#(\rho_1\|\sigma_1) + D_\alpha^\#(\rho_2\|\sigma_2)$$

Regularization: $$\lim_{n\to\infty} \frac{1}{n} D_\alpha^\#(\rho^{\otimes n}\|\sigma^{\otimes n}) = \widetilde{D}_\alpha(\rho\|\sigma)$$

Proof: Use measured divergence $D_\alpha^{\mathbb{M}}(\rho\|\sigma) = \max_{\mathcal{M}} D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma))$ and $\lim_{n\to\infty} \frac{1}{n} D_\alpha^{\mathbb{M}}(\rho\|\sigma) = \widetilde{D}_\alpha(\rho\|\sigma)$

$$D_\alpha^{\mathbb{M}}(\rho\|\sigma) \leq D_\alpha^\#(\rho\|\sigma) \leq D_\alpha^{\mathbb{M}}(\rho\|\sigma) + \frac{\alpha}{\alpha-1}\log|\mathrm{spec}(\sigma)|.$$

- $D_\alpha^\#(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \leq D_\alpha^\#(\rho\|\sigma)$ for all $\mathcal{M}$. Thus $D_\alpha^{\mathbb{M}}(\rho\|\sigma) \leq D_\alpha^\#(\rho\|\sigma)$

Comparison with sandwiched $\widetilde{D}_\alpha$

Remark: $D_\alpha^{\#}$ is *not* additive under tensor product, but it is subadditive

$$D_\alpha^{\#}(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) \leq D_\alpha^{\#}(\rho_1\|\sigma_1) + D_\alpha^{\#}(\rho_2\|\sigma_2)$$

Regularization:

$$\lim_{n\to\infty} \frac{1}{n} D_\alpha^{\#}(\rho^{\otimes n}\|\sigma^{\otimes n}) \quad = \quad \widetilde{D}_\alpha(\rho\|\sigma)$$

Proof: Use measured divergence $D_\alpha^{\mathbb{M}}(\rho\|\sigma) = \max_{\mathcal{M}} D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma))$ and $\lim_{n\to\infty} \frac{1}{n} D_\alpha^{\mathbb{M}}(\rho\|\sigma) = \widetilde{D}_\alpha(\rho\|\sigma)$

$$D_\alpha^{\mathbb{M}}(\rho\|\sigma) \quad \leq \quad D_\alpha^{\#}(\rho\|\sigma) \quad \leq \quad D_\alpha^{\mathbb{M}}(\rho\|\sigma) + \frac{\alpha}{\alpha-1} \log|\mathrm{spec}(\sigma)|.$$

- $D_\alpha^{\#}(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \leq D_\alpha^{\#}(\rho\|\sigma)$ for all $\mathcal{M}$. Thus $D_\alpha^{\mathbb{M}}(\rho\|\sigma) \leq D_\alpha^{\#}(\rho\|\sigma)$
- Key fact: if $\rho \leq \rho'$ then $D_\alpha^{\#}(\rho\|\sigma) \leq D_\alpha^{\#}(\rho'\|\sigma)$.

Comparison with sandwiched $\widetilde{D}_\alpha$

Remark: $D_\alpha^\#$ is *not* additive under tensor product, but it is subadditive

$$D_\alpha^\#(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) \le D_\alpha^\#(\rho_1\|\sigma_1) + D_\alpha^\#(\rho_2\|\sigma_2)$$

Regularization: $$\lim_{n\to\infty} \frac{1}{n} D_\alpha^\#(\rho^{\otimes n}\|\sigma^{\otimes n}) \quad = \quad \widetilde{D}_\alpha(\rho\|\sigma)$$

Proof: Use measured divergence $D_\alpha^{\mathbb{M}}(\rho\|\sigma) = \max_{\mathcal{M}} D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma))$ and $\lim_{n\to\infty} \frac{1}{n} D_\alpha^{\mathbb{M}}(\rho\|\sigma) = \widetilde{D}_\alpha(\rho\|\sigma)$

$$D_\alpha^{\mathbb{M}}(\rho\|\sigma) \quad \le \quad D_\alpha^\#(\rho\|\sigma) \quad \le \quad D_\alpha^{\mathbb{M}}(\rho\|\sigma) + \frac{\alpha}{\alpha-1} \log|\mathrm{spec}(\sigma)|.$$

- $D_\alpha^\#(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \le D_\alpha^\#(\rho\|\sigma)$ for all $\mathcal{M}$. Thus $D_\alpha^{\mathbb{M}}(\rho\|\sigma) \le D_\alpha^\#(\rho\|\sigma)$
- Key fact: if $\rho \le \rho'$ then $D_\alpha^\#(\rho\|\sigma) \le D_\alpha^\#(\rho'\|\sigma)$.

  Assume $\sigma = \sum_\lambda \lambda P_\lambda$ spectral decomposition. Consider pinching map

  $$\mathcal{P}_\sigma(X) = \sum_{\lambda\in\mathrm{spec}(\sigma)} P_\lambda X P_\lambda.$$

  Using pinching inequality $\rho \le |\mathrm{spec}(\sigma)|\mathcal{P}_\sigma(\rho)$:

  $$\begin{aligned} Q_\alpha^\#(\rho\|\sigma) &\le Q_\alpha^\#(|\mathrm{spec}(\sigma)|\mathcal{P}_\sigma(\rho)\|\sigma) \\ &= |\mathrm{spec}(\sigma)|^\alpha Q_\alpha^\#(\mathcal{P}_\sigma(\rho)\|\sigma) \\ &= |\mathrm{spec}(\sigma)|^\alpha Q_\alpha^\#(\mathcal{P}_\sigma(\rho)\|\mathcal{P}_\sigma(\sigma)) \le |\mathrm{spec}(\sigma)|^\alpha Q_\alpha^{\mathbb{M}}(\rho\|\sigma). \end{aligned}$$

Extension to channels

Definition of divergence for channels: $\mathcal{N}_{X'\to Y}, \mathcal{M}_{X'\to Y}$ completely positive trace-preserving maps

$$D_\alpha^{\#}(\mathcal{N}\|\mathcal{M}) := \sup_{\rho\in\mathcal{D}(X\otimes X')} D_\alpha^{\#}\Big((\mathcal{N}_{X'\to Y}\otimes\mathcal{I}_{X\to X})(\rho)\|(\mathcal{M}_{X'\to Y}\otimes\mathcal{I}_{X\to X})(\rho)\Big)$$
$$= \sup_{\omega_X\in\mathcal{D}(X)} D_\alpha^{\#}(\omega_X^{\frac{1}{2}} J_{XY}^{\mathcal{N}} \omega_X^{\frac{1}{2}} \| \omega_X^{\frac{1}{2}} J_{XY}^{\mathcal{M}} \omega_X^{\frac{1}{2}}) \ ,$$

$J_{XY}^{\mathcal{N}} = (\mathcal{I}_{X\to X}\otimes\mathcal{N}_{X'\to Y})(|\Phi\rangle\langle\Phi|_{XX'})$ where $|\Phi\rangle_{XX'} = \sum_i |i\rangle_X \otimes |i\rangle_{X'}$

Can be written as a convex optimization problem:

$$D_\alpha^{\#}(\mathcal{N}\|\mathcal{M}) = \frac{1}{\alpha-1}\log\left\{\min_{A_{XY}\geq 0} \|\mathrm{tr}\,_Y(A_{XY})\|_\infty \ \text{ s.t. } \ J_{XY}^{\mathcal{M}} \#_{1/\alpha} A_{XY} \geq J_{XY}^{\mathcal{N}}\right\}$$

Main properties

1. Sub-additive

$$D_\alpha^{\#}(\mathcal{N}_1 \otimes \mathcal{N}_2 \| \mathcal{M}_1 \otimes \mathcal{M}_2) \leq D_\alpha^{\#}(\mathcal{N}_1 \| \mathcal{M}_1) + D_\alpha^{\#}(\mathcal{N}_2 \| \mathcal{M}_2)$$

2. Regularization coincides with that of $\widetilde{D}_\alpha$:

$$\inf_n \frac{1}{n} D_\alpha^{\#}(\mathcal{N}^{\otimes n} \| \mathcal{M}^{\otimes n}) = \sup_n \frac{1}{n} \widetilde{D}_\alpha(\mathcal{N}^{\otimes n} \| \mathcal{M}^{\otimes n}) = \widetilde{D}_\alpha^{\mathrm{reg}}(\mathcal{N} \| \mathcal{M})$$

Remark: $\widetilde{D}_\alpha^{\mathrm{reg}}(\mathcal{N} \| \mathcal{M}) > \widetilde{D}_\alpha(\mathcal{N} \| \mathcal{M})$ in general

3. Chain rule:

$$D_\alpha^{\#}(\mathcal{N}(\rho) \| \mathcal{M}(\sigma)) \ \leq \ D_\alpha^{\#}(\mathcal{N} \| \mathcal{M}) \ + \ D_\alpha^{\#}(\rho \| \sigma).$$

Application:
Channel discrimination, adaptive strategies do not help in some settings
See David Sutter's talk on Thursday

Application: computation of $\widetilde{D}^{\mathrm{reg}}(\mathcal{N}\|\mathcal{M})$

$\widetilde{D}_\alpha^{\mathrm{reg}}(\mathcal{N}\|\mathcal{M}) = \lim_{n\to\infty} \frac{1}{n}\widetilde{D}_\alpha(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n})$ can be approximated efficiently

For any $n \geq 1$

$$\frac{D_\alpha^{\#}(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n})}{n} \ - \ \epsilon(\alpha, d, n) \ \leq \ \widetilde{D}_\alpha^{\mathrm{reg}}(\mathcal{N}\|\mathcal{M}) \ \leq \ \frac{D_\alpha^{\#}(\mathcal{N}^{\otimes n}\|\mathcal{M}^{\otimes n})}{n}$$

where

$$\epsilon(\alpha, d, n) \ = \ \frac{1}{n}\frac{\alpha}{\alpha-1}(d^2+d)\log(n+d)$$

and $d = \dim X \dim Y$ and $\alpha > 1$

As $n \to \infty$ converging hierarchy of upper bound on $\widetilde{D}_\alpha^{\mathrm{reg}}(\mathcal{N}\|\mathcal{M})$

Conclusion

- Divergences defined in terms of a convex optimization program
- Has complementary properties (chain rule, "nice" variational formulation,...) and useful as a proof tool
- Applications to Shannon theory `arXiv:2007.12576` and bounding rates for device-independent cryptographic protocols `arXiv:2007.12575`

Open questions:

- $\lim_{\alpha \to 1} D_\alpha^{\#}$?
- Other divergences via optimization? $\alpha < 1$?