Chain rules for quantum relative entropies and their applications

Kun Fang[1]

[1909.05826, PRL] & [1909.05758]

Joint work with Hamza Fawzi[1],
Omar Fawzi[2], Renato Renner[3] and David Sutter[3]

[1] DAMTP, University of Cambridge
[2] Laboratoire de l'Informatique du Parallélisme, ENS de Lyon
[3] Institute for Theoretical Physics, ETH Zurich

Beyond IID 2020

The chain rule

Entropy of a **large system** = **sum** of entropies of individual **subsystems**

For example, given a tripartite quantum state ρ_{ABC} we have

$$H(AB|C) = H(A|C) + H(B|AC)$$

$$H(A|B) := H(AB) - H(B)$$

$$H(A) := -\operatorname{Tr} \rho_A \log \rho_A$$

$$D(\rho\|\sigma) := \operatorname{Tr} \rho(\log \rho - \log \sigma)$$

Recall that $H(A|B) = -D(\rho_{AB} \| \mathrm{id}_A \otimes \rho_B)$

Question: Is there a chain rule for the quantum relative entropy?

The chain rule

How does the chain rule look like for quantum relative entropy?

For classical probability distributions we have

$$D(P_{XY}\|Q_{XY}) = D(P_X\|Q_X) + \underline{\sum_x P_X(x) D(P_{Y|X=x}\|Q_{Y|X=x})}$$

Less ambitious statement

$$D(P_{XY}\|Q_{XY}) \leq D(P_X\|Q_X) + \underline{\max_x D(P_{Y|X=x}\|Q_{Y|X=x})}$$

How to model a conditional distribution (a channel) quantumly?

Relative entropy for quantum channels $\mathcal{E}_{A\to B}$ and $\mathcal{F}_{A\to B}$

Consider the worst-case scenario $D(\mathcal{E}\|\mathcal{F}) := \sup_{\rho_{RA}} D(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\rho_{RA}))$

$$D(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) \overset{?}{\leq} D(\rho_{RA}\|\sigma_{RA}) + D(\mathcal{E}\|\mathcal{F})$$

Outline for the rest of the talk

- Umegaki relative entropy **D**

  $$D(\rho\|\sigma) := \operatorname{Tr}\rho(\log\rho - \log\sigma)$$

  Chain rule and its applications **[1909.05826]**

- Belavkin-Staszewski relative entropy **$\widehat{D}$**

  $$\widehat{D}(\rho\|\sigma) = \operatorname{Tr}\rho\log[\rho^{1/2}\sigma^{-1}\rho^{1/2}]$$

  Chain rule and its applications **[1909.05758]**

- Summary and discussions

Umegaki Relative Entropy

D

Non-additivity of channel relative entropy

A fact: The channel relative entropy is **not additive** under tensor product.

There exists quantum channels $\mathcal{E}$ and $\mathcal{F}$ such that

$$D(\mathcal{E} \otimes \mathcal{E} \| \mathcal{F} \otimes \mathcal{F}) > 2D(\mathcal{E} \| \mathcal{F})$$

In sharp contrast with the relative entropy of quantum states

---

Let $\mathcal{E}$ and $\mathcal{F}$ be two different qubit **generalized amplitude damping channels** with Choi matrices $J_{\mathcal{E}}$ and $J_{\mathcal{F}}$

Using the covariance symmetry of these channels, we find

$$D(\mathcal{E} \| \mathcal{F}) = \max_{\rho = \mathrm{diag}(p, 1-p)} D(\sqrt{\rho} J_{\mathcal{E}} \sqrt{\rho} \| \sqrt{\rho} J_{\mathcal{F}} \sqrt{\rho})$$

For some clever choice of ρ we find

$$D(\mathcal{E} \otimes \mathcal{E} \| \mathcal{F} \otimes \mathcal{F}) \geq D(\mathcal{E}^{\otimes 2}(\rho) \| \mathcal{F}^{\otimes 2}(\rho)) > 2D(\mathcal{E} \| \mathcal{F})$$

Non-additivity leads to the definition of regularized channel relative entropy

$$D^{\mathrm{reg}}(\mathcal{E} \| \mathcal{F}) := \lim_{n \to \infty} \frac{1}{n} D(\mathcal{E}^{\otimes n} \| \mathcal{F}^{\otimes n})$$

The naive chain rule conjecture is false

$$D(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) \leq D(\rho_{RA}\|\sigma_{RA}) + D(\mathcal{E}\|\mathcal{F})$$

There exists states ρ_{RA},σ_{RA} and channels $\mathcal{E}_{A\to B}$,$\mathcal{F}_{A\to B}$ such that

$$D(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) > D(\rho_{RA}\|\sigma_{RA}) + D(\mathcal{E}\|\mathcal{F})$$

Recall channel relative entropy $D(\mathcal{E}\|\mathcal{F}) := \sup_{\rho_{RA}} D(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\rho_{RA}))$

The **amortized channel relative entropy** is defined as

$$D^A(\mathcal{E}\|\mathcal{F}) := \sup_{\rho_{RA},\sigma_{RA}} [D(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) - D(\rho_{RA}\|\sigma_{RA})]$$

It is known [Wang-Wilde-19] that

$$D^A(\mathcal{E}\|\mathcal{F}) \geq D^{\text{reg}}(\mathcal{E}\|\mathcal{F})$$

Hence there exist channels $\mathcal{E}$ and $\mathcal{F}$ such that

$$D^A(\mathcal{E}\|\mathcal{F}) \geq D^{\text{reg}}(\mathcal{E}\|\mathcal{F}) > D(\mathcal{E}\|\mathcal{F})$$

Chain rule for Umegaki relative entropy

$$D(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) \leq D(\rho_{RA}\|\sigma_{RA}) + D(\mathcal{E}\|\mathcal{F})$$

Chain Rule

For any quantum states ρ_{RA} , σ_{RA} and quantum channels $\mathcal{E}_{A\to B}$, $\mathcal{F}_{A\to B}$

$$D(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) \leq D(\rho_{RA}\|\sigma_{RA}) + D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F})$$

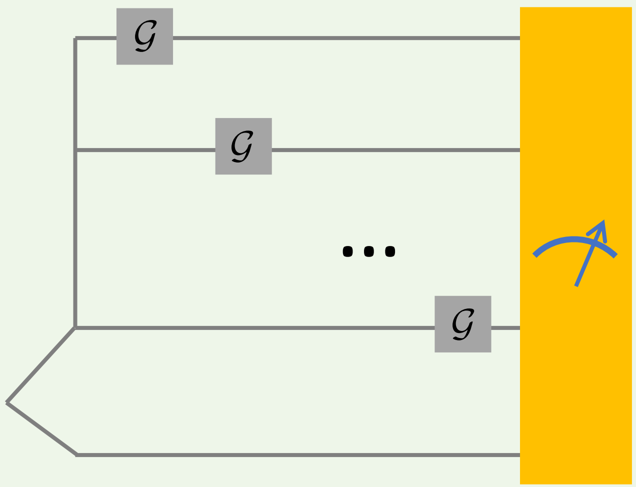
Some remarks:

- For any channel pair $(\mathcal{E}, \mathcal{F})$, there exists ρ and σ such that the chain rule holds with equality, i.e., $D^{A}(\mathcal{E}\|\mathcal{F}) = D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F})$
- $D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) = D(\mathcal{E}\|\mathcal{F})$ for specific channels
  - Classical-quantum channels
  - Covariant channels w.r.t. unitary group
  - $\mathcal{E}$ arbitrary and $\mathcal{F}$ a replacer channel
- For $\mathcal{E} = \mathcal{F}$ we recover the data-processing inequality $D(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq D(\rho\|\sigma)$

Application: channel discrimination

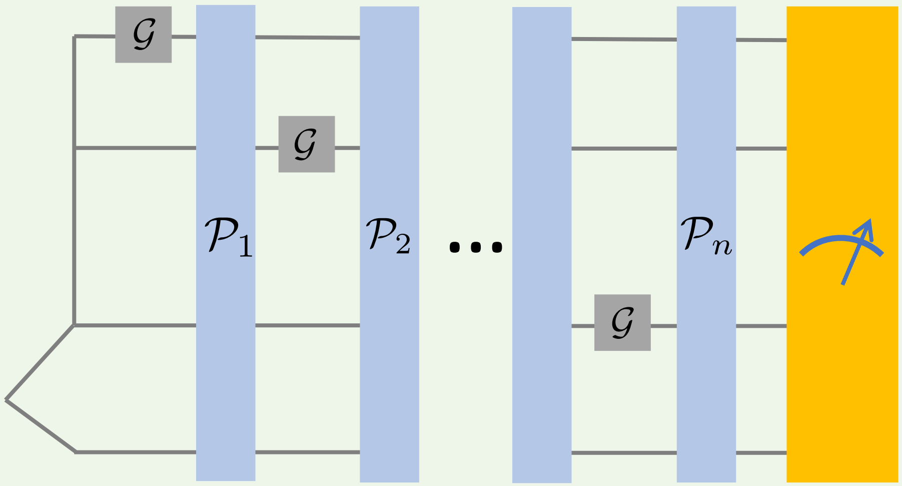
Given a quantum channel $\mathcal{G} \in \{\mathcal{E}, \mathcal{F}\}$

Using $\mathcal{G}$ n-times the task is to determine if $\mathcal{G} = \mathcal{E}$ or $\mathcal{G} = \mathcal{F}$



non-adaptive strategy

$D^{\text{reg}}(\mathcal{E}\|\mathcal{F})$ [Wang-Wilde-19]



adaptive strategy

$D^{A}(\mathcal{E}\|\mathcal{F})$ [Wang-Wilde-19]

Because non-adaptive strategies are a special case of adaptive strategies

$$D^{\text{reg}}(\mathcal{E}\|\mathcal{F}) \leq D^{A}(\mathcal{E}\|\mathcal{F})$$

But the new chain rule says

$$\boxed{D^{\text{reg}}(\mathcal{E}\|\mathcal{F}) = D^{A}(\mathcal{E}\|\mathcal{F})}$$!!!

Adaptive strategies are no more powerful than non-adaptive ones!

Open questions

- Do we have a chain rule for sandwiched/Petz Rényi relative entropy

$$D_\alpha(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) \overset{?}{\le} D_\alpha(\rho_{RA}\|\sigma_{RA}) + D_\alpha^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) \qquad \alpha \in (1/2, 1) \cup (1, \infty)$$

Sandwiched chain rule with $\alpha > 1$ is solved by [Fawzi-Fawzi-2020, see Wednesday's talk]

- Single-letterize more quantities as we did for $D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F})$ with $D^A(\mathcal{E}\|\mathcal{F})$

For example: Capacity formula?

$$Q(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} I_c(\mathcal{N}^{\otimes n}) =: I_c^{\mathrm{reg}}(\mathcal{N}) = \quad ?$$

- Extreme non-additivity of channel relative entropy? (Mark Wilde's Twitter)

Is there **a universal n** such that

$$D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) = \frac{1}{n} D(\mathcal{E}^{\otimes n}\|\mathcal{F}^{\otimes n}) \qquad \text{for all channels}$$

Analogous to a result by [Cubbit et.al, 1408.5115] that
the channel coherent information is extremely non-additive.

Belavkin-Staszewski Relative Entropy

$$\hat{D}$$

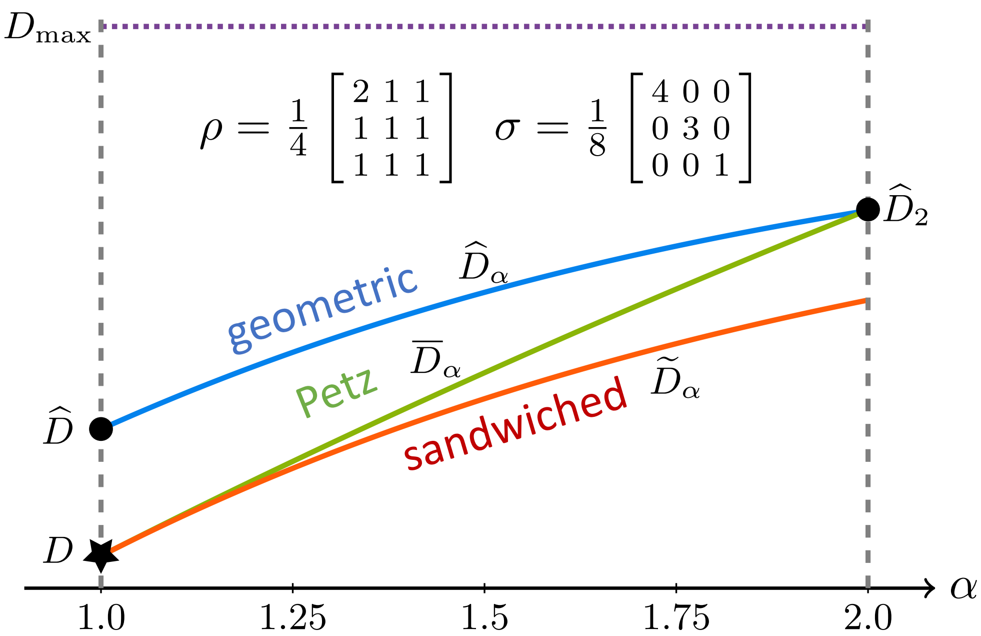
Definitions: geometric Rényi divergence

Geometric Rényi divergence:

$$\widehat{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \operatorname{Tr} G_{1-\alpha}(\rho, \sigma)$$

Matrix geometric mean (an operator connects X and Y):

$$G_\alpha(X, Y) := X^{\frac{1}{2}} \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right)^\alpha X^{\frac{1}{2}}$$



- Also called the *maximal Rényi divergence* [Matsumoto-15]: the largest quantum Rényi divergence which satisfies data-processing inequality
- Converge to Belavkin-Staszewski when alpha = 1

  $$\widehat{D}(\rho\|\sigma) := \operatorname{Tr} \rho \log[\rho^{1/2}\sigma^{-1}\rho^{1/2}]$$

- Nicer properties than the widely-used Petz and sandwiched ones

Chain rule and other basic properties

Chain Rule

For any quantum states ρ_{RA} , σ_{RA} and quantum channels $\mathcal{E}_{A\to B}$, $\mathcal{F}_{A\to B}$

$$\widehat{D}_\alpha(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) \leq \widehat{D}_\alpha(\rho_{RA}\|\sigma_{RA}) + \widehat{D}_\alpha(\mathcal{E}\|\mathcal{F}) \qquad \alpha \in (1,2]$$

Single-letter

Other nice properties:

Closed-form $\widehat{D}_\alpha(\mathcal{E}\|\mathcal{F}) = \frac{1}{\alpha - 1} \log \left\| \operatorname{Tr}_B G_{1-\alpha}(J_{RB}^{\mathcal{E}}, J_{RB}^{\mathcal{F}}) \right\|_\infty$

Additivity $\widehat{D}_\alpha(\mathcal{E}_1 \otimes \mathcal{E}_2\|\mathcal{F}_1 \otimes \mathcal{F}_2) = \widehat{D}_\alpha(\mathcal{E}_1\|\mathcal{F}_1) + \widehat{D}_\alpha(\mathcal{E}_2\|\mathcal{F}_2)$

Sub-additivity $\widehat{D}_\alpha(\mathcal{E}_2 \circ \mathcal{E}_1\|\mathcal{F}_2 \circ \mathcal{F}_1) \leq \widehat{D}_\alpha(\mathcal{E}_1\|\mathcal{F}_1) + \widehat{D}_\alpha(\mathcal{E}_2\|\mathcal{F}_2)$

SDP $\inf_{\mathcal{F}\in C} \widehat{D}_\alpha(\mathcal{E}\|\mathcal{F})$ is SDP computable if C is given by SDP conditions

Remarks:

- These properties will empower a wide range of applications.
- Similar properties hold for α in (0,1) [Katariya-Wilde-2020].
- Umegaki relative entropy does not satisfy these properties.

Application: channel capacity

Task: quantum comm. over a noisy channel with free classical comm. assistance

Channel capacity: the capability of a channel to reliably transmit information

Notoriously hard to evaluate and we aim to find an upper bound as tight as possible

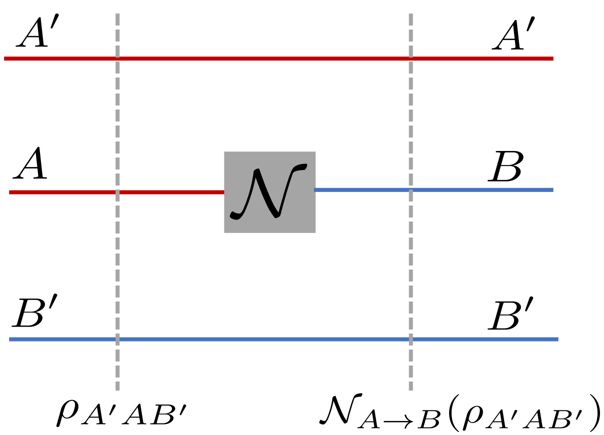
Rains entanglement measure

$$\widehat{R}_\alpha(\rho_{AB}) := \inf_{\sigma_{AB} \in \mathrm{PPT}'(A:B)} \widehat{D}_\alpha(\rho_{AB} \| \sigma_{AB}) \qquad \mathrm{PPT}'(A:B) := \{\sigma_{AB} \geq 0 : \|\sigma_{AB}^{T_B}\|_1 \leq 1\}$$

Rains channel information

$$\widehat{R}_\alpha(\mathcal{N}) := \inf_{\mathcal{M} \in \mathcal{V}(A:B)} \widehat{D}_\alpha(\mathcal{N} \| \mathcal{M}) \qquad \mathcal{V}(A:B) := \{\mathcal{M} \in \mathrm{CP} : \|\Theta_B \circ \mathcal{M}_{A\to B}\|_\diamond \leq 1\}$$

Chain rule immediately implies $\widehat{R}_\alpha(\mathcal{N}_{A\to B}(\rho_{A'AB})) - \widehat{R}_\alpha(\rho_{A'AB}) \leq \widehat{R}_\alpha(\mathcal{N})$



Q: What does this mean?

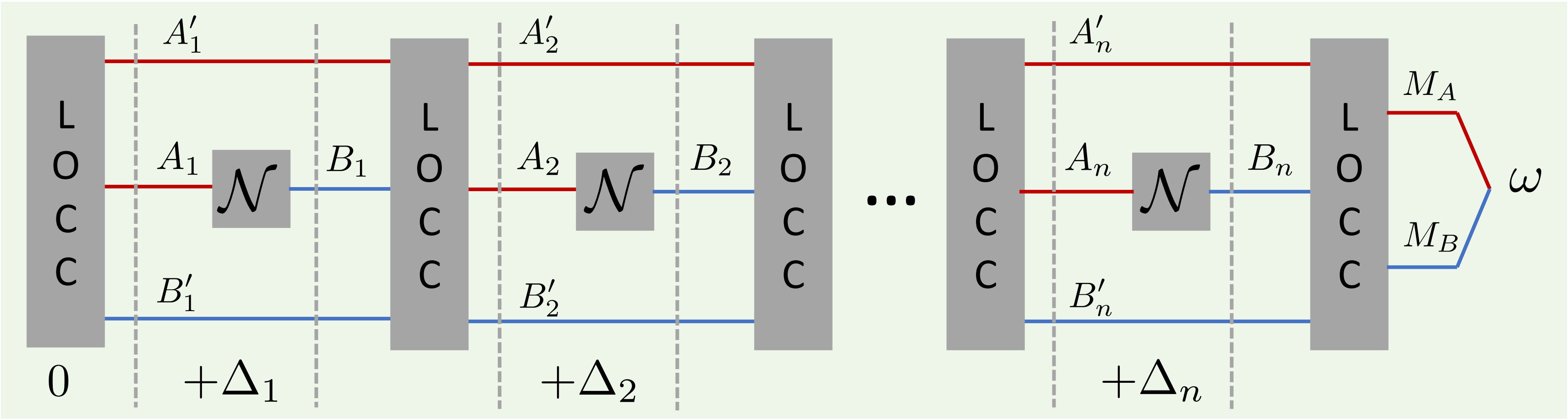
A: Net entanglement generated via the channel $\mathcal{N}$ will be no greater than $\widehat{R}_\alpha(\mathcal{N})$

Q: Why should we care about this?

A: This is a sub-module in quantum communication.

Application: channel capacity

LOCC (local operation and classical comm.) assisted quantum communication protocol



Goal: establish maximally entangled state

+ free classical communication $\xrightarrow{\text{teleportation}}$ transmit quantum info.

Using channel n times we have $\widehat{R}_\alpha(\omega) \leq \Delta_1 + \Delta_2 + \cdots + \Delta_n \leq n \cdot \widehat{R}_\alpha(\mathcal{N})$

On average, entanglement generated is no greater than $\widehat{R}_\alpha(\mathcal{N})$

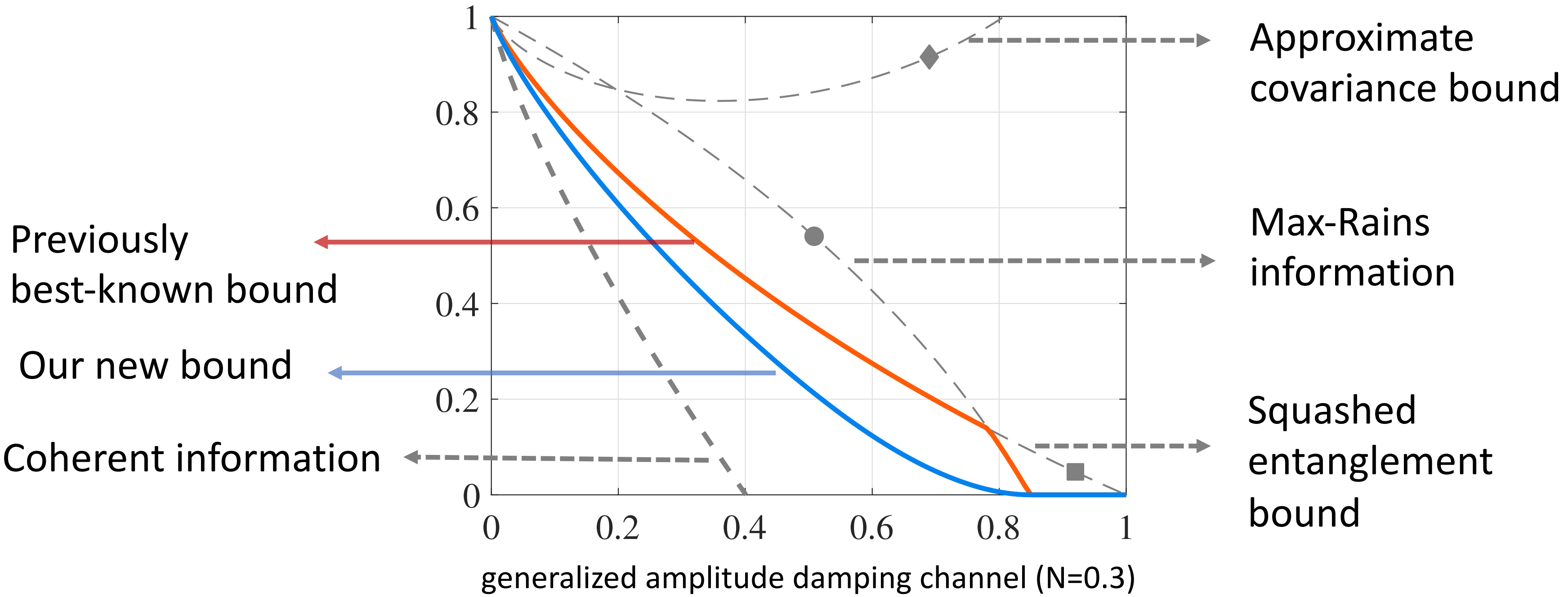
Thus we have an improved bound $Q^{\leftrightarrow}(\mathcal{N}) \leq \widehat{R}_\alpha(\mathcal{N}) \leq R_{\max}(\mathcal{N})$

Previously best-known bound [Wang-Fang-Duan-18, Berta-Wilde-18; QIP’18 talk]

Example

$$Q^{\leftrightarrow}(\mathcal{N}) \leq \widehat{R}_{\alpha}(\mathcal{N}) \leq R_{\max}(\mathcal{N})$$

- $\widehat{R}_{\alpha}(\mathcal{N})$ is tighter than $R_{\max}(\mathcal{N})$ in general.
- The improvement is significant for almost all channels.
- The new bound cannot be trivially pushed further to Umegaki's relative entropy D as a single-letter bound.



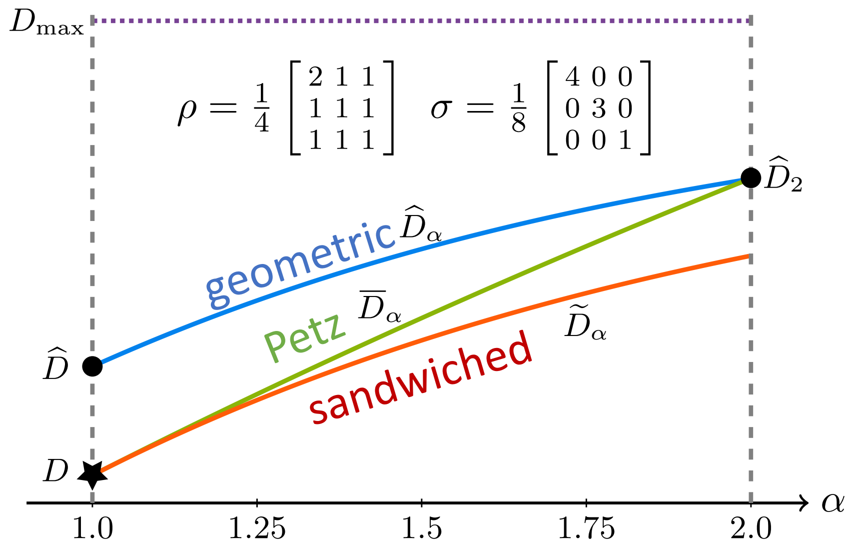
Other applications

The BS/geometric Rényi divergence can also found applications in:

- Classical/private/magic generating capacities
- Point-to-point/bidirectional channels
- Assisted/unassisted communication scenario
- Channel discrimination $D^{\mathrm{reg}}(\mathcal{N}\|\mathcal{M}) \le \widehat{D}_\alpha(\mathcal{N}\|\mathcal{M})$

See 1909.05758 for more details

Potential improvements in quantum network theory, quantum repeaters, quantum key distribution, quantum games … (basically anywhere that involves $D_{\max}$)



Open question

Belavkin-Staszewski relative entropy/Geometric Rényi divergence admits **nice mathematical properties**. **But** what are their operational meanings? Do they naturally show up in certain tasks?

For example:

- *Umegaki relative entropy*: optimal error exponent in the hypothesis testing problem [Hiai-Petz-1991]
- *Petz Rényi divergence*: quantum generalization of Chernoff's bound on the success probability in binary hypothesis testing [Audenaert et al.-2007]
- *Sandwiched Rényi divergence*: strong converse regime of asymmetric binary hypothesis testing [Mosonyi-Ogawa-2015]

Summary

Summary

- Chain rule for Umegaki relative entropy

$$D(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) \leq D(\rho_{RA}\|\sigma_{RA}) + D^{\text{reg}}(\mathcal{E}\|\mathcal{F})$$

- Adaptive strategies are no more powerful than non-adaptive ones for channel discrimination $D^{\text{reg}}(\mathcal{E}\|\mathcal{F}) = D^{A}(\mathcal{E}\|\mathcal{F})$
- Robust version of data-processing inequality
- See arXiv 1909.05826 for a slightly more general version

- Chain rule for Belavkin-Staszewski/geometric Rényi relative entropy

$$\widehat{D}_\alpha(\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})) \leq \widehat{D}_\alpha(\rho_{RA}\|\sigma_{RA}) + \widehat{D}_\alpha(\mathcal{E}\|\mathcal{F})$$

- Other nice properties: closed-form formula, additive under tensor product, sub-additive under composition, easy to do optimization
- Improved upper bounds for quantum/private/classical/magic state generation capacities /channel discrimination
- Potential applications in quantum network theory, quantum repeaters, quantum key distribution, quantum games…

Thanks for your attention!

See arXiv for more details

[1909.05826, PRL] & [1909.05758]