Relative entropy optimization in quantum information

Omar Fawzi

ENS DE LYON

ICMP 2018, Montréal

Quantum relative entropie**s**

For **classical states** (i.e., prob. distributions) P and Q on $\mathcal{X}$

$$D(P\|Q) := \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}$$

Quantum relative entropies

For **classical states** (i.e., prob. distributions) P and Q on $\mathcal{X}$

$$D(P\|Q) := \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}$$

For **quantum states** ρ and σ on $\mathbb{C}^d$, **multiple choices:**

1. Matrix logarithm [Umegaki, 1962]

$$D(\rho\|\sigma) := \mathrm{tr}[\rho\log\rho] - \mathrm{tr}[\rho\log\sigma]$$

2. Matrix logarithm in a different way [Belavkin, Stasewski, 1982]

$$D^{BS}(\rho\|\sigma) := \mathrm{tr}\left[\rho\log\left(\rho^{1/2}\sigma^{-1}\rho^{1/2}\right)\right]$$

3. Optimize over all measurements [Donald, 1986]

$$D_{\mathbb{M}}(\rho\|\sigma) := \sup_{\{M_x\}_{x\in\mathcal{X}}\ \mathrm{PSD}, \sum_x M_x = \mathrm{id}} \sum_{x\in\mathcal{X}} \mathrm{tr}[M_x\rho]\log\frac{\mathrm{tr}[M_x\rho]}{\mathrm{tr}[M_x\sigma]}$$

Most common is Umegaki's: hypothesis testing interpretation [Hiai, Petz, 1991, The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability]

Quantum relative entropies

For **classical states** (i.e., prob. distributions) P and Q on $\mathcal{X}$

$$D(P\|Q) := \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}$$

For **quantum states** ρ and σ on $\mathbb{C}^d$, **multiple choices:**

1. Matrix logarithm [Umegaki, 1962]

$$D(\rho\|\sigma) := \mathrm{tr}[\rho\log\rho] - \mathrm{tr}[\rho\log\sigma]$$

2. Matrix logarithm in a different way [Belavkin, Stasewski, 1982]

$$D^{BS}(\rho\|\sigma) := \mathrm{tr}\left[\rho\log\left(\rho^{1/2}\sigma^{-1}\rho^{1/2}\right)\right]$$

3. Optimize over all measurements [Donald, 1986]

$$D_{\mathbb{M}}(\rho\|\sigma) := \sup_{\substack{\{M_x\}_{x\in\mathcal{X}} \\ \mathrm{PSD}, \sum_x M_x = \mathrm{id}}} \sum_{x\in\mathcal{X}} \mathrm{tr}[M_x\rho]\log\frac{\mathrm{tr}[M_x\rho]}{\mathrm{tr}[M_x\sigma]}$$

Most common is Umegaki's: hypothesis testing interpretation [Hiai, Petz, 1991, The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability]

... but others can also be useful too.

Quantum relative entropies

$$D_{\mathbb{M}}(\rho\|\sigma) \leq D(\rho\|\sigma) \leq D^{BS}(\rho\|\sigma)$$

Most important property: **Joint convexity**

$$\mathbb{D}((1-t)\rho_0 + t\rho_1\|(1-t)\sigma_0 + t\sigma_1) \leq (1-t)\mathbb{D}(\rho_0\|\sigma_0) + t\mathbb{D}(\rho_1\|\sigma_1)$$

- Classical relative entropy $D(P\|Q)$: simple application of convexity of $x \mapsto x \log x$
- **Quantum** relative entropies:
  - D: consequence of Lieb's concavity theorem [Lieb, 1973]
  - D^{BS}: consequence of concavity of matrix geometric mean [Fujii, Kamei, 1989]
  - $D_{\mathbb{M}}$: follows easily from the classical case as sup of convex functions

Operational consequence: Data processing inequality, for $\mathcal{N}$ completely positive trace preserving map

$$\mathbb{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq \mathbb{D}(\rho\|\sigma)$$

Another appealing consequence: For $\mathcal{S}$ convex, $\min_{(\rho,\sigma)\in\mathcal{S}} \mathbb{D}(\rho\|\sigma)$ is a convex problem

Quantities based on relative entropy optimization

1. Relative entropy of entanglement

$$E_R(\rho_{AB}) = \min_{\sigma_{AB} \in \mathrm{Sep}_{AB}} D(\rho_{AB} \| \sigma_{AB})$$

   More generally, relative entropy of resource $E(\rho) = \min_{\sigma \in \mathcal{F}} D(\rho\|\sigma)$ in a resource theory where $\mathcal{F}$ are the free states
   Quantifies amount of resource in state ρ

2. Quantum channel capacities, e.g., entanglement assisted capacity $\mathcal{N}(\rho) = \mathrm{tr}_E(U\rho U^*)$ with U isometry $A \to B \otimes E$

$$\begin{aligned} C_{ea}(\mathcal{N}) = \max \quad & -D(\sigma_{BE}\|\sigma_B \otimes \mathrm{id}_E) - D(\sigma_B\|\mathrm{id}_B) \\ \text{s.t.} \quad & \sigma_{BE} = \mathcal{N}(\rho_A),\ \rho_A \in \mathrm{D}(A) \end{aligned}$$

3. $\mathbb{D}$ of recovery of ρ_{ABC}: quantifies how well C can be locally recovered

$$\min_{\mathcal{R}:\mathrm{L}(B)\to\mathrm{L}(BC)\ \mathrm{CPTP}} \mathbb{D}(\rho_{ABC}\|(\mathcal{I}_A \otimes \mathcal{R})(\rho_{AB}))$$

Running example: **recoverability**

Recoverability

$$I(A:C|B)_\rho = D(\rho_{AB}\|\mathrm{id}_A \otimes \rho_B) - D(\rho_{ABC}\|\mathrm{id}_A \otimes \rho_{BC})$$

Motivation:

Operational properties of states ρ_{ABC} with $I(A:C|B)_\rho \leq \epsilon$

near-saturation of data processing inequality for D
"approximate quantum Markov chains"

Surprisingly, there is a state ρ_{ABC} with $I(A:C|B)_\rho \leq \frac{1}{d}$ and ρ_{ABC} is $\frac{1}{4}$-far from exact Markov states [Ibinson, Linden, Winter, 2006] and [Christandl, Schuch, Winter, 2012]

Recoverability

$$I(A:C|B)_\rho = D(\rho_{AB}\|\mathrm{id}_A \otimes \rho_B) - D(\rho_{ABC}\|\mathrm{id}_A \otimes \rho_{BC})$$

Motivation:

Operational properties of states ρ_{ABC} with $I(A:C|B)_\rho \leq \epsilon$

near-saturation of data processing inequality for D
"approximate quantum Markov chains"

Surprisingly, there is a state ρ_{ABC} with $I(A:C|B)_\rho \leq \frac{1}{d}$ and ρ_{ABC} is $\frac{1}{4}$-far from exact Markov states [Ibinson, Linden, Winter, 2006] and [Christandl, Schuch, Winter, 2012]

But, the state ρ_{ABC} is **approximately recoverable** [Fawzi, Renner, 2014] building on [Li, Winter, 2012], ..., [Berta, Seshadreesan, Wilde, 2014]:

$$\min_{\mathcal{R}:\mathrm{L}(B)\to\mathrm{L}(BC)\ \mathrm{CPTP}} \mathbb{D}(\rho_{ABC}\|(\mathcal{I}_A \otimes \mathcal{R})(\rho_{AB})) \leq \epsilon$$

for $\mathbb{D} = -2\log F$ (aka sandwiched Rényi divergence of order $\frac{1}{2}$)

Recoverability

Let $\mathbb{D}^{\mathrm{rec}}(\rho_{ABC}) = \min_{\mathcal{R}:\mathsf{L}(B)\to\mathsf{L}(BC) \text{ CPTP}} \mathbb{D}(\rho_{ABC} \| (\mathcal{I}_A \otimes \mathcal{R})(\rho_{AB}))$

We saw that

$$\mathbb{D}^{\mathrm{rec}}(\rho_{ABC}) \leq I(A:C|B)_\rho \quad \text{for } \mathbb{D} = -2\log F$$ [Fawzi, Renner, 2014]

Note that $-2\log F \leq D_{\mathbb{M}} \leq D \leq D^{BS}$
The inequality is true with $\mathbb{D} = D$ classically
Can it be improved in quantum case with $\mathbb{D} = D_{\mathbb{M}}, D, D^{BS}$?

Recoverability

Let $\mathbb{D}^{\mathrm{rec}}(\rho_{ABC}) = \min_{\mathcal{R}:\mathsf{L}(B)\to\mathsf{L}(BC)\text{ CPTP}} \mathbb{D}(\rho_{ABC} \| (\mathcal{I}_A \otimes \mathcal{R})(\rho_{AB}))$

We saw that

$$\mathbb{D}^{\mathrm{rec}}(\rho_{ABC}) \leq I(A:C|B)_\rho \quad \text{for } \mathbb{D} = -2\log F \quad \text{[Fawzi, Renner, 2014]}$$

Note that $-2\log F \leq D_{\mathbb{M}} \leq D \leq D^{BS}$
The inequality is true with $\mathbb{D} = D$ classically
Can it be improved in quantum case with $\mathbb{D} = D_{\mathbb{M}}, D, D^{BS}$?

YES for $D_{\mathbb{M}}$ as shown in [Brandao, Harrow, Oppenheim, Strelchuck, 2014]
NO for D as shown in [Fawzi, Fawzi, 2017]

Recoverability

Let $\mathbb{D}^{\mathrm{rec}}(\rho_{ABC}) = \min_{\mathcal{R}:\mathsf{L}(B)\to\mathsf{L}(BC)\text{ CPTP}} \mathbb{D}(\rho_{ABC}\|(\mathcal{I}_A\otimes\mathcal{R})(\rho_{AB}))$

We saw that

$$\mathbb{D}^{\mathrm{rec}}(\rho_{ABC}) \leq I(A:C|B)_\rho \quad \text{for } \mathbb{D} = -2\log F \quad \text{[Fawzi, Renner, 2014]}$$

Note that $-2\log F \leq D_{\mathbb{M}} \leq D \leq D^{BS}$
The inequality is true with $\mathbb{D} = D$ classically
Can it be improved in quantum case with $\mathbb{D} = D_{\mathbb{M}}, D, D^{BS}$?

YES for $D_{\mathbb{M}}$ as shown in [Brandao, Harrow, Oppenheim, Strelchuck, 2014]
NO for D as shown in [Fawzi, Fawzi, 2017]

Why does $D_{\mathbb{M}}$ behave better here? $\rightarrow$ **additivity** property of $D_{\mathbb{M}}^{\mathrm{rec}}$ under tensor product, not satisfied by D^{rec}

Additivity of optimized relative entropies I

- Consider

$$\mathbb{D}^{\mathrm{opt}}(\rho) := \min_{\sigma \in \mathcal{C}} \mathbb{D}(\rho \| \sigma)\,,$$

where $\mathcal{C}$ convex set of states

Additivity of optimized relative entropies I

- Consider

$$\mathbb{D}^{\mathrm{opt}}(\rho) := \min_{\sigma \in \mathcal{C}} \mathbb{D}(\rho \| \sigma)\,,$$

where $\mathcal{C}$ convex set of states

Additivity of optimized relative entropies I

- Consider

$$\mathbb{D}^{\mathrm{opt}}(\rho) := \min_{\sigma\in\mathcal{C}} \mathbb{D}(\rho\|\sigma)\,,$$

  where $\mathcal{C}$ convex set of states

- Both $\mathbb{D} = D$ and $\mathbb{D} = D_{\mathbb{M}}$ are super-additive on tensor product states

$$\mathbb{D}(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) \geq \mathbb{D}(\rho_1\|\sigma_1) + \mathbb{D}(\rho_2\|\sigma_2)$$

- Does this property transfer to $\mathbb{D}^{\mathrm{opt}}$?

Additivity of optimized relative entropies I

- Consider

$$\mathbb{D}^{\mathrm{opt}}(\rho) := \min_{\sigma \in \mathcal{C}} \mathbb{D}(\rho \| \sigma)\,,$$

where $\mathcal{C}$ convex set of states

- Both $\mathbb{D} = D$ and $\mathbb{D} = D_{\mathbb{M}}$ are super-additive on tensor product states

$$\mathbb{D}(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) \geq \mathbb{D}(\rho_1 \| \sigma_1) + \mathbb{D}(\rho_2 \| \sigma_2)$$

- Does this property transfer to $\mathbb{D}^{\mathrm{opt}}$?
- Super-additivity of $\mathbb{D}^{\mathrm{opt}}$ on tensor product states:

$$\begin{aligned}\min_{\sigma_{12} \in \mathcal{C}_{12}} \mathbb{D}(\rho_1 \otimes \rho_2 \| \sigma_{12}) &= \mathbb{D}^{\mathrm{opt}}(\rho_1 \otimes \rho_2) \\ &\overset{?}{\geq} \mathbb{D}^{\mathrm{opt}}(\rho_1) + \mathbb{D}^{\mathrm{opt}}(\rho_2) \\ &= \min_{\sigma_1 \in \mathcal{C}_1} \mathbb{D}(\rho_1 \| \sigma_1) + \min_{\sigma_2 \in \mathcal{C}_2} \mathbb{D}(\rho_2 \| \sigma_2)\end{aligned}$$

Using variational formulas

- Idea: Use variational formulas for $\mathbb{D}$

$$D(\rho\|\sigma) = \sup_{\omega>0} \operatorname{tr}[\rho \log \omega] + 1 - \operatorname{tr} \exp(\log \sigma + \log \omega) \quad \text{[Petz, 1988]}$$

$$D_{\mathbb{M}}(\rho\|\sigma) = \sup_{\omega>0} \operatorname{tr}[\rho \log \omega] + 1 - \operatorname{tr}[\sigma\omega] \quad \text{[Hiai, Petz '93, Berta, Fawzi, Tomamichel '15]}$$

- Remarks:
  - Golden-Thompson inequality $\to D_{\mathbb{M}} \leq D$
  - The formula for $D_{\mathbb{M}} \to$ efficient computation for $D_{\mathbb{M}}$

Using variational formulas

- Idea: Use variational formulas for $\mathbb{D}$

$$D(\rho\|\sigma) = \sup_{\omega>0} \mathrm{tr}[\rho \log \omega] + 1 - \mathrm{tr} \exp(\log \sigma + \log \omega) \quad \text{[Petz, 1988]}$$

$$D_{\mathbb{M}}(\rho\|\sigma) = \sup_{\omega>0} \mathrm{tr}[\rho \log \omega] + 1 - \mathrm{tr}[\sigma\omega] \quad \text{[Hiai, Petz '93, Berta, Fawzi, Tomamichel '15]}$$

- Remarks:
  - Golden-Thompson inequality $\rightarrow D_{\mathbb{M}} \leq D$
  - The formula for $D_{\mathbb{M}}$ $\rightarrow$ efficient computation for $D_{\mathbb{M}}$
- Back to showing additivity $\mathbb{D}(\rho\|\sigma) = \sup_{\omega>0} f(\rho, \sigma, \omega)$
  Using Sion's minimax theorem:

$$\mathbb{D}^{\mathrm{opt}}(\rho) = \min_{\sigma\in\mathcal{C}} \sup_{\omega>0} f(\rho,\sigma,\omega) = \sup_{\omega>0} \min_{\sigma\in\mathcal{C}} f(\rho,\sigma,\omega)$$

Using variational formulas

- Idea: Use variational formulas for $\mathbb{D}$

$$D(\rho\|\sigma) = \sup_{\omega>0} \operatorname{tr}[\rho \log \omega] + 1 - \operatorname{tr}\exp(\log\sigma + \log\omega) \quad \text{[Petz, 1988]}$$

$$D_{\mathbb{M}}(\rho\|\sigma) = \sup_{\omega>0} \operatorname{tr}[\rho \log \omega] + 1 - \operatorname{tr}[\sigma\omega] \quad \text{[Hiai, Petz '93, Berta, Fawzi, Tomamichel '15]}$$

- Remarks:
  - Golden-Thompson inequality $\to D_{\mathbb{M}} \leq D$
  - The formula for $D_{\mathbb{M}} \to$ efficient computation for $D_{\mathbb{M}}$
- Back to showing additivity $\mathbb{D}(\rho\|\sigma) = \sup_{\omega>0} f(\rho,\sigma,\omega)$
  Using Sion's minimax theorem:

$$\mathbb{D}^{\mathrm{opt}}(\rho) = \min_{\sigma\in\mathcal{C}} \sup_{\omega>0} f(\rho,\sigma,\omega) = \sup_{\omega>0} \min_{\sigma\in\mathcal{C}} f(\rho,\sigma,\omega)$$

- For $\mathbb{D} = D_{\mathbb{M}}$,

$$\min_{\sigma\in\mathcal{C}} f(\rho,\sigma,\omega) = \min_{\sigma\in\mathcal{C}} \operatorname{tr}[\rho\log\omega] + 1 - \operatorname{tr}[\sigma\omega]$$

  Semidefinite program (if $\mathcal{C}$ is nice) $\to$ use strong duality

$$\min_{\sigma\in\mathcal{C}} f(\rho,\sigma,\omega) = \max_{\bar\sigma\in\bar{\mathcal{C}}_\omega} \bar f(\rho,\bar\sigma,\omega)$$

Using variational formulas

- Idea: Use variational formulas for $\mathbb{D}$

$$D(\rho\|\sigma) = \sup_{\omega>0} \operatorname{tr}[\rho \log \omega] + 1 - \operatorname{tr}\exp(\log\sigma + \log\omega) \quad \text{[Petz, 1988]}$$

$$D_{\mathbb{M}}(\rho\|\sigma) = \sup_{\omega>0} \operatorname{tr}[\rho\log\omega] + 1 - \operatorname{tr}[\sigma\omega] \quad \text{[Hiai, Petz '93, Berta, Fawzi, Tomamichel '15]}$$

- Remarks:
  - Golden-Thompson inequality $\to D_{\mathbb{M}} \leq D$
  - The formula for $D_{\mathbb{M}} \to$ efficient computation for $D_{\mathbb{M}}$
- Back to showing additivity $\mathbb{D}(\rho\|\sigma) = \sup_{\omega>0} f(\rho,\sigma,\omega)$
  Using Sion's minimax theorem:

$$\mathbb{D}^{\mathrm{opt}}(\rho) = \min_{\sigma\in\mathcal{C}} \sup_{\omega>0} f(\rho,\sigma,\omega) = \sup_{\omega>0} \min_{\sigma\in\mathcal{C}} f(\rho,\sigma,\omega)$$

- For $\mathbb{D} = D_{\mathbb{M}}$,

$$\min_{\sigma\in\mathcal{C}} f(\rho,\sigma,\omega) = \min_{\sigma\in\mathcal{C}} \operatorname{tr}[\rho\log\omega] + 1 - \operatorname{tr}[\sigma\omega]$$

  Semidefinite program (if $\mathcal{C}$ is nice) $\to$ use strong duality

$$\min_{\sigma\in\mathcal{C}} f(\rho,\sigma,\omega) = \max_{\bar\sigma\in\bar{\mathcal{C}}_\omega} \bar f(\rho,\bar\sigma,\omega)$$

- For $\mathbb{D} = D$, not a semidefinite program: $\sigma \mapsto \operatorname{tr}\exp(\log\sigma + \log\omega)$ is concave but no simple expression for its dual

Additivity using variational formulas

- We wrote the optimized measured relative entropy as

$$D_{\mathbb{M}}^{\rm opt}(\rho) = \sup_{\omega>0} \max_{\bar\sigma \in \bar{\mathcal{C}}_\omega} \bar f(\rho, \bar\sigma, \omega)$$

- Want to show $D_{\mathbb{M}}^{\rm opt}(\rho_1 \otimes \rho_2) \geq D_{\mathbb{M}}^{\rm opt}(\rho_1) + D_{\mathbb{M}}^{\rm opt}(\rho_2)$
- Proof: Take $\omega_1, \omega_2 > 0$ and $\bar\sigma_1 \in \bar{\mathcal{C}}_1$, $\bar\sigma_2 \in \bar{\mathcal{C}}_2$ achieving maximum. Then consider $\omega_1 \otimes \omega_2$ and $\bar\sigma_1 \otimes \bar\sigma_2 \in \bar{\mathcal{C}}_{12}$ and get

$$\begin{aligned} D_{\mathbb{M}}^{\rm opt}(\rho_1 \otimes \rho_2) \geq \bar f(\rho_1 \otimes \rho_2, \bar\sigma_1 \otimes \bar\sigma_2, \omega_1 \otimes \omega_2) &\geq \bar f(\rho_1, \bar\sigma_1, \omega_1) + \bar f(\rho_2, \bar\sigma_2, \omega_2) \\ &= D_{\mathbb{M}}^{\rm opt}(\rho_1) + D_{\mathbb{M}}^{\rm opt}(\rho_2) \end{aligned}$$

  This works provided
  - $\bar f$ is super-additive

$$\bar f(\rho_1 \otimes \rho_2, \bar\sigma_1 \otimes \bar\sigma_2, \omega_1 \otimes \omega_2) \geq \bar f(\rho_1, \bar\sigma_1, \omega_1) + \bar f(\rho_2, \bar\sigma_2, \omega_2)$$

  - the sets $\bar{\mathcal{C}}$ are closed under tensor products

$$\bar\sigma_1 \in \bar{\mathcal{C}}_1 \text{ and } \bar\sigma_2 \in \bar{\mathcal{C}}_2 \quad \text{imply that} \quad \bar\sigma_1 \otimes \bar\sigma_2 \in \bar{\mathcal{C}}_{12}$$

- For **recoverability example**:
  - $\mathcal{C} = \{(\mathcal{I}_A \otimes \mathcal{R}_{B\to BC})(\rho_{AB})\}$ and $f = \mathrm{tr}(\rho_{ABC} \log \omega_{ABC}) + 1 - \mathrm{tr}(\sigma_{ABC}\omega_{ABC})$
  - $\bar{\mathcal{C}}_\omega = \{\bar\sigma : \mathrm{id}_{BC} \otimes \bar\sigma_{AR} \geq \omega_{ABC} \otimes \mathrm{id}_R, \mathrm{tr}(\bar\sigma_{AR}\rho_{AR}) = 1\}$ and $\bar f = \mathrm{tr}(\rho_{ABC} \log \omega_{ABC})$

$\Rightarrow D_{\mathbb{M}}^{\rm rec}$ is super-additive

Relative entropy optimization: algorithms

Based on [Fawzi, Saunderson, 2015]

Example: computing $D_{\mathbb{M}}(\rho\|\sigma) = \sup_{\omega>0} \operatorname{tr}[\rho\log\omega] + 1 - \operatorname{tr}[\sigma\omega]$ for fixed ρ, σ (computing $\min_{(\rho,\sigma)\in\mathcal{S}} \mathbb{D}(\rho\|\sigma)$ slightly more complicated but uses similar ideas)

- $\log\omega \approx \frac{\omega^{2^{-k}}-1}{2^{-k}}$ for $k\to\infty$
- $k=1$: $\omega \succeq T^2$ iff $\begin{bmatrix}\omega & T\\ T & I\end{bmatrix} \succeq 0$
- $k=1$:

$$D_{\mathbb{M}}(\rho\|\sigma) \approx \max\left\{\operatorname{tr}\left[\rho\left(\frac{T-1}{1/2}\right)\right] + 1 - \operatorname{tr}[\sigma\omega] : \begin{bmatrix}\omega & T\\ T & I\end{bmatrix} \succeq 0\right\} \leftarrow \text{SDP}$$

- Recursion for $k \geq 2$

For more efficient approximation [Fawzi, Saunderson, Parrilo, 2017]
[Fawzi, Fawzi, 2017] used it to show that D^{rec} is not additive

Concluding remarks

- Multiple quantum relative entropies $\mathbb{D}$ that are jointly convex
- Use variational expressions and duality to establish additivity properties of optimized relative entropies
- Can efficiently approximate $\min_{(\rho,\sigma)\in\mathcal{S}} \mathbb{D}(\rho\|\sigma)$ using semidefinite programs

`https://github.com/hfawzi/cvxquad/`