Equality condition in the data processing inequality for the quantum relative entropy

Felix Leditzky

CAKE Seminar

17 February 2016

Table of Contents

Table of Contents

Motivation: Kullback-Leibler divergence

Let P, Q be probability distributions on a discrete probability space $\mathcal{X}$, and define the **Kullback-Leibler divergence** $D_{\mathrm{KL}}(P\|Q)$:

$$D_{\mathrm{KL}}(P\|Q) := \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

This relative entropy is a *premetric*:

$$D_{\mathrm{KL}}(P\|Q) \geq 0 \qquad \text{and} \qquad D_{\mathrm{KL}}(P\|Q) = 0 \text{ iff } P = Q.$$

Motivation: Kullback-Leibler divergence

Let P, Q be probability distributions on a discrete probability space $\mathcal{X}$, and define the **Kullback-Leibler divergence** $D_{\mathrm{KL}}(P \| Q)$:

$$D_{\mathrm{KL}}(P \| Q) := \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

This relative entropy is a *premetric*:

$$D_{\mathrm{KL}}(P \| Q) \geq 0 \qquad \text{and} \qquad D_{\mathrm{KL}}(P \| Q) = 0 \text{ iff } P = Q.$$

Motivation: Kullback-Leibler divergence

Operational interpretation: Binary hypothesis testing

- Assume that we are given n independent and identically distributed (i.i.d.) copies of one of two probability distributions P or Q.

- **Goal:** Determine whether we have P (*null hypothesis* H_P) or Q (*alternative hypothesis* H_Q).

- Two possible errors:
  - **Type-I error**: We falsely reject H_P.
  - **Type-II error**: We falsely accept H_P.

Motivation: Kullback-Leibler divergence

Operational interpretation: Binary hypothesis testing

- Assume that we are given n independent and identically distributed (i.i.d.) copies of one of two probability distributions P or Q.
- **Goal:** Determine whether we have P (*null hypothesis* H_P) or Q (*alternative hypothesis* H_Q).
- Two possible errors:
  - **Type-I error**: We falsely reject H_P.
  - **Type-II error**: We falsely accept H_P.

Motivation: Kullback-Leibler divergence

Operational interpretation: Binary hypothesis testing

- Assume that we are given n independent and identically distributed (i.i.d.) copies of one of two probability distributions P or Q.
- **Goal:** Determine whether we have P (*null hypothesis* H_P) or Q (*alternative hypothesis* H_Q).
- Two possible errors:
  - **Type-I error**: We falsely reject H_P.
  - **Type-II error**: We falsely accept H_P.

Motivation: Kullback-Leibler divergence

Operational interpretation: Binary hypothesis testing

- Assume that we are given n independent and identically distributed (i.i.d.) copies of one of two probability distributions P or Q.
- **Goal:** Determine whether we have P (*null hypothesis* H_P) or Q (*alternative hypothesis* H_Q).
- Two possible errors:
  - **Type-I error**: We falsely reject H_P.
  - **Type-II error**: We falsely accept H_P.

Motivation: Kullback-Leibler divergence

Operational interpretation: Binary hypothesis testing

- Assume that we are given n independent and identically distributed (i.i.d.) copies of one of two probability distributions P or Q.
- **Goal:** Determine whether we have P (*null hypothesis* H_P) or Q (*alternative hypothesis* H_Q).
- Two possible errors:
  - **Type-I error**: We falsely reject H_P.
  - **Type-II error**: We falsely accept H_P.

Motivation: Kullback-Leibler divergence

Operational interpretation: Binary hypothesis testing

- In general: Trade-off between these errors.
- One possibility: Try to minimize both at the same time
  $\longrightarrow$ symmetric hypothesis testing, Chernoff bound
- Another one:

$$\text{minimize type-II error s.t. type-I error } \leq \epsilon$$

- Optimal exponent in the limit $n \to \infty$ given by $D_{\mathrm{KL}}(P \| Q)$:

$$\text{type-II error } \approx \exp(-nD_{\mathrm{KL}}(P \| Q)) \quad \text{for large } n.$$

Motivation: Kullback-Leibler divergence

Operational interpretation: Binary hypothesis testing

- In general: Trade-off between these errors.
- One possibility: Try to minimize both at the same time
  $\longrightarrow$ symmetric hypothesis testing, Chernoff bound
- Another one:

$$\text{minimize type-II error s.t. type-I error } \leq \epsilon$$

- Optimal exponent in the limit $n \to \infty$ given by $D_{\mathrm{KL}}(P\|Q)$:

$$\text{type-II error } \approx \exp(-nD_{\mathrm{KL}}(P\|Q)) \quad \text{for large } n.$$

Motivation: Kullback-Leibler divergence

Operational interpretation: Binary hypothesis testing

- In general: Trade-off between these errors.
- One possibility: Try to minimize both at the same time
  $\longrightarrow$ symmetric hypothesis testing, Chernoff bound
- Another one:

$$\text{minimize type-II error s.t. type-I error } \leq \epsilon$$

- Optimal exponent in the limit $n \to \infty$ given by $D_{\mathrm{KL}}(P\|Q)$:

$$\text{type-II error } \approx \exp(-nD_{\mathrm{KL}}(P\|Q)) \quad \text{for large } n.$$

Motivation: Kullback-Leibler divergence

Operational interpretation: Binary hypothesis testing

- In general: Trade-off between these errors.
- One possibility: Try to minimize both at the same time
  $\longrightarrow$ symmetric hypothesis testing, Chernoff bound
- Another one:

$$\text{minimize type-II error s.t. type-I error } \leq \epsilon$$

- Optimal exponent in the limit $n \to \infty$ given by $D_{\mathrm{KL}}(P\|Q)$:

$$\text{type-II error } \approx \exp(-nD_{\mathrm{KL}}(P\|Q)) \quad \text{for large } n.$$

Motivation: Kullback-Leibler divergence

KL-divergence satisfies the **data processing inequality** (DPI):

- Let P_X, Q_X be probability distributions on $\mathcal{X}$, and let $\Gamma_{Y|X}: X \to Y \in \mathcal{X}$ be a *classical channel*.
- Denote by P_Y, Q_Y the resulting distributions, that is, $P_Y(x) := \sum_{z\in\mathcal{X}} P_X(z)\Gamma_{Y|X}(z|x)$ and similar for Q_Y.
- Data processing inequality:

$$D_{\mathrm{KL}}(P_X \| Q_X) \geq D_{\mathrm{KL}}(P_Y \| Q_Y)$$

- Consequence: Transformations $\Gamma_{Y|X}$ make it *harder* to discriminate between P_X and Q_X.

Motivation: Kullback-Leibler divergence

KL-divergence satisfies the **data processing inequality** (DPI):

- Let P_X, Q_X be probability distributions on $\mathcal{X}$, and let $\Gamma_{Y|X} \colon X \to Y \in \mathcal{X}$ be a *classical channel*.
- Denote by P_Y, Q_Y the resulting distributions, that is, $P_Y(x) := \sum_{z\in\mathcal{X}} P_X(z)\Gamma_{Y|X}(z|x)$ and similar for Q_Y.
- Data processing inequality:

$$D_{\mathrm{KL}}(P_X \| Q_X) \geq D_{\mathrm{KL}}(P_Y \| Q_Y)$$

- Consequence: Transformations $\Gamma_{Y|X}$ make it *harder* to discriminate between P_X and Q_X.

Motivation: Kullback-Leibler divergence

KL-divergence satisfies the **data processing inequality** (DPI):

- Let P_X, Q_X be probability distributions on $\mathcal{X}$, and let $\Gamma_{Y|X}\colon X \to Y \in \mathcal{X}$ be a *classical channel*.
- Denote by P_Y, Q_Y the resulting distributions, that is, $P_Y(x) := \sum_{z\in\mathcal{X}} P_X(z)\Gamma_{Y|X}(z|x)$ and similar for Q_Y.
- Data processing inequality:

$$D_{\mathrm{KL}}(P_X\|Q_X) \geq D_{\mathrm{KL}}(P_Y\|Q_Y)$$

- Consequence: Transformations $\Gamma_{Y|X}$ make it *harder* to discriminate between P_X and Q_X.

Motivation: Kullback-Leibler divergence

KL-divergence satisfies the **data processing inequality** (DPI):

- Let P_X, Q_X be probability distributions on $\mathcal{X}$, and let $\Gamma_{Y|X}: X \to Y \in \mathcal{X}$ be a *classical channel*.
- Denote by P_Y, Q_Y the resulting distributions, that is, $P_Y(x) := \sum_{z\in\mathcal{X}} P_X(z)\Gamma_{Y|X}(z|x)$ and similar for Q_Y.
- Data processing inequality:

$$D_{\mathrm{KL}}(P_X\|Q_X) \geq D_{\mathrm{KL}}(P_Y\|Q_Y)$$

- Consequence: Transformations $\Gamma_{Y|X}$ make it *harder* to discriminate between P_X and Q_X.

Motivation: Kullback-Leibler divergence

Importance of data processing inequality

$$D_{\mathrm{KL}}(P_X \| Q_X) \geq D_{\mathrm{KL}}(P_Y \| Q_Y)$$

- Phrase an information-theoretic task in terms of transformations (e.g. encoding, decoding, ...).
- Characterize the task by entropic quantities based on relative entropies such as $D_{\mathrm{KL}}(\cdot \| \cdot)$.
- Data processing inequality then allows us to derive bounds on the **optimal rate** of the task.
- Same principle in Classical and Quantum Information Theory!

Motivation: Kullback-Leibler divergence

Importance of data processing inequality

$$D_{\text{KL}}(P_X \| Q_X) \geq D_{\text{KL}}(P_Y \| Q_Y)$$

- Phrase an information-theoretic task in terms of transformations (e.g. encoding, decoding, ...).
- Characterize the task by entropic quantities based on relative entropies such as $D_{\text{KL}}(\cdot\|\cdot)$.
- Data processing inequality then allows us to derive bounds on the **optimal rate** of the task.
- Same principle in Classical and Quantum Information Theory!

Motivation: Kullback-Leibler divergence

Importance of data processing inequality

$$D_{\mathrm{KL}}(P_X \| Q_X) \geq D_{\mathrm{KL}}(P_Y \| Q_Y)$$

- Phrase an information-theoretic task in terms of transformations (e.g. encoding, decoding, ...).
- Characterize the task by entropic quantities based on relative entropies such as $D_{\mathrm{KL}}(\cdot \| \cdot)$.
- Data processing inequality then allows us to derive bounds on the **optimal rate** of the task.
- Same principle in Classical and Quantum Information Theory!

Motivation: Kullback-Leibler divergence

Importance of data processing inequality

$$D_{\mathrm{KL}}(P_X \| Q_X) \geq D_{\mathrm{KL}}(P_Y \| Q_Y)$$

- Phrase an information-theoretic task in terms of transformations (e.g. encoding, decoding, ...).
- Characterize the task by entropic quantities based on relative entropies such as $D_{\mathrm{KL}}(\cdot \| \cdot)$.
- Data processing inequality then allows us to derive bounds on the **optimal rate** of the task.
- Same principle in Classical and Quantum Information Theory!

Table of Contents

Mathematics of Quantum Mechanics 101

How do we "make things quantum"?

- Replace the discrete probability space $\mathcal{X}$ by a Hilbert space $\mathcal{H}$ of dimension $|\mathcal{X}| < \infty$ (that is, $\mathcal{H} \cong \mathbb{C}^{|\mathcal{X}|}$).
- **Density operator** (or mixed state) is an operator ρ acting on $\mathcal{H}$ that is
  1. **positive:** $\rho \geq 0$ (that is, $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$)
  2. **normalized:** $\operatorname{Tr} \rho = 1$
- Eigenvalues of a density matrix form a probability distribution!
- However, for a unitary U the operators ρ and $U\rho U^\dagger$ have the same eigenvalues.

Mathematics of Quantum Mechanics 101

How do we "make things quantum"?

- Replace the discrete probability space $\mathcal{X}$ by a Hilbert space $\mathcal{H}$ of dimension $|\mathcal{X}| < \infty$ (that is, $\mathcal{H} \cong \mathbb{C}^{|\mathcal{X}|}$).
- **Density operator** (or mixed state) is an operator ρ acting on $\mathcal{H}$ that is
  1. **positive:** $\rho \geq 0$ (that is, $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$)
  2. **normalized:** $\mathrm{Tr}\, \rho = 1$
- Eigenvalues of a density matrix form a probability distribution!
- However, for a unitary U the operators ρ and $U\rho U^\dagger$ have the same eigenvalues.

Mathematics of Quantum Mechanics 101

How do we "make things quantum"?

- Replace the discrete probability space $\mathcal{X}$ by a Hilbert space $\mathcal{H}$ of dimension $|\mathcal{X}| < \infty$ (that is, $\mathcal{H} \cong \mathbb{C}^{|\mathcal{X}|}$).
- **Density operator** (or mixed state) is an operator ρ acting on $\mathcal{H}$ that is
  1. **positive:** $\rho \geq 0$ (that is, $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$)
  2. **normalized:** $\operatorname{Tr} \rho = 1$
- Eigenvalues of a density matrix form a probability distribution!
- However, for a unitary U the operators ρ and $U\rho U^\dagger$ have the same eigenvalues.

Mathematics of Quantum Mechanics 101

How do we "make things quantum"?

- Replace the discrete probability space $\mathcal{X}$ by a Hilbert space $\mathcal{H}$ of dimension $|\mathcal{X}| < \infty$ (that is, $\mathcal{H} \cong \mathbb{C}^{|\mathcal{X}|}$).
- **Density operator** (or mixed state) is an operator ρ acting on $\mathcal{H}$ that is
  1. **positive:** $\rho \geq 0$ (that is, $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$)
  2. **normalized:** $\operatorname{Tr} \rho = 1$
- Eigenvalues of a density matrix form a probability distribution!
- However, for a unitary U the operators ρ and $U\rho U^\dagger$ have the same eigenvalues.

Mathematics of Quantum Mechanics 101

How do we "make things quantum"?

- Replace the discrete probability space $\mathcal{X}$ by a Hilbert space $\mathcal{H}$ of dimension $|\mathcal{X}| < \infty$ (that is, $\mathcal{H} \cong \mathbb{C}^{|\mathcal{X}|}$).
- **Density operator** (or mixed state) is an operator ρ acting on $\mathcal{H}$ that is
  1. **positive:** $\rho \geq 0$ (that is, $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$)
  2. **normalized:** $\operatorname{Tr} \rho = 1$
- Eigenvalues of a density matrix form a probability distribution!
- However, for a unitary U the operators ρ and $U\rho U^\dagger$ have the same eigenvalues.

Mathematics of Quantum Mechanics 101

How do we "make things quantum"?

- Replace the discrete probability space $\mathcal{X}$ by a Hilbert space $\mathcal{H}$ of dimension $|\mathcal{X}| < \infty$ (that is, $\mathcal{H} \cong \mathbb{C}^{|\mathcal{X}|}$).
- **Density operator** (or mixed state) is an operator ρ acting on $\mathcal{H}$ that is
  1. **positive:** $\rho \geq 0$ (that is, $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$)
  2. **normalized:** $\operatorname{Tr} \rho = 1$
- Eigenvalues of a density matrix form a probability distribution!
- However, for a unitary U the operators ρ and $U\rho U^{\dagger}$ have the same eigenvalues.

Mathematics of Quantum Mechanics 101

- Interpretation: Assume that the **pure state** of a system is described by a normalized (column) vector $|\psi\rangle \in \mathcal{H}$.
- **Mixed state** $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ describes a system that is in the pure state ψ_i with probability p_i.

  (in general, $|\psi_i\rangle \not\perp |\psi_j\rangle$ for $i \neq j$)
- Spectral decomposition of ρ:

$$\rho = \sum_i \lambda_i |e_i\rangle\langle e_i| \quad \text{with } \langle e_i|e_j\rangle = \delta_{ij}$$

  where $|e_i\rangle$ is an eigenvector of ρ with eigenvalue $\lambda_i \geq 0$.
- "Quantumness": In general, $[\rho, \sigma] \neq 0$ for two states ρ,σ, that is, ρ and σ have different eigenbases.

Mathematics of Quantum Mechanics 101

- Interpretation: Assume that the **pure state** of a system is described by a normalized (column) vector $|\psi\rangle \in \mathcal{H}$.
- **Mixed state** $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ describes a system that is in the pure state ψ_i with probability p_i.

  (in general, $|\psi_i\rangle \not\perp |\psi_j\rangle$ for $i \neq j$)
- Spectral decomposition of ρ:

$$\rho = \sum_i \lambda_i |e_i\rangle\langle e_i| \quad \text{with } \langle e_i|e_j\rangle = \delta_{ij}$$

  where $|e_i\rangle$ is an eigenvector of ρ with eigenvalue $\lambda_i \geq 0$.
- "Quantumness": In general, $[\rho, \sigma] \neq 0$ for two states ρ,σ, that is, ρ and σ have different eigenbases.

Mathematics of Quantum Mechanics 101

- Interpretation: Assume that the **pure state** of a system is described by a normalized (column) vector $|\psi\rangle \in \mathcal{H}$.
- **Mixed state** $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ describes a system that is in the pure state ψ_i with probability p_i.

  (in general, $|\psi_i\rangle \not\perp |\psi_j\rangle$ for $i \neq j$)
- Spectral decomposition of ρ:

$$\rho = \sum_i \lambda_i |e_i\rangle\langle e_i| \quad \text{with } \langle e_i|e_j\rangle = \delta_{ij}$$

  where $|e_i\rangle$ is an eigenvector of ρ with eigenvalue $\lambda_i \geq 0$.
- "Quantumness": In general, $[\rho, \sigma] \neq 0$ for two states ρ,σ, that is, ρ and σ have different eigenbases.

Mathematics of Quantum Mechanics 101

- Interpretation: Assume that the **pure state** of a system is described by a normalized (column) vector $|\psi\rangle \in \mathcal{H}$.
- **Mixed state** $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ describes a system that is in the pure state ψ_i with probability p_i.

  (in general, $|\psi_i\rangle \not\perp |\psi_j\rangle$ for $i \neq j$)
- Spectral decomposition of ρ:

$$\rho = \sum_i \lambda_i |e_i\rangle\langle e_i| \quad \text{with } \langle e_i|e_j\rangle = \delta_{ij}$$

  where $|e_i\rangle$ is an eigenvector of ρ with eigenvalue $\lambda_i \geq 0$.
- "Quantumness": In general, $[\rho, \sigma] \neq 0$ for two states ρ,σ, that is, ρ and σ have different eigenbases.

Mathematics of Quantum Mechanics 101

A bit more abstract:

- Density operators correspond to positive, normalized elements of the C^*-algebra $\mathcal{B}(\mathcal{H})$ of linear bounded operators acting on a Hilbert space $\mathcal{H}$ (for us $\dim \mathcal{H} < \infty$).
- The $*$-map is given by the adjoint $^\dagger : A \mapsto A^\dagger$, and $\|A^\dagger A\| = \|A\|^2$ where $\|\cdot\|$ is the operator norm.
- Note that $A \geq 0 \Rightarrow A^\dagger = A$ (pos. elements are Hermitian).
- We equip $\mathcal{B}(\mathcal{H})$ with the **Hilbert-Schmidt inner product**:

$$\langle X, Y \rangle := \mathrm{Tr}(X^\dagger Y)$$

Mathematics of Quantum Mechanics 101

A bit more abstract:

- Density operators correspond to positive, normalized elements of the C^*-algebra $\mathcal{B}(\mathcal{H})$ of linear bounded operators acting on a Hilbert space $\mathcal{H}$ (for us $\dim \mathcal{H} < \infty$).
- The $*$-map is given by the adjoint $^\dagger : A \mapsto A^\dagger$, and $\|A^\dagger A\| = \|A\|^2$ where $\|\cdot\|$ is the operator norm.
- Note that $A \geq 0 \Rightarrow A^\dagger = A$ (pos. elements are Hermitian).
- We equip $\mathcal{B}(\mathcal{H})$ with the **Hilbert-Schmidt inner product**:

$$\langle X, Y \rangle := \mathrm{Tr}(X^\dagger Y)$$

Mathematics of Quantum Mechanics 101

A bit more abstract:

- Density operators correspond to positive, normalized elements of the C^*-algebra $\mathcal{B}(\mathcal{H})$ of linear bounded operators acting on a Hilbert space $\mathcal{H}$ (for us $\dim \mathcal{H} < \infty$).
- The $*$-map is given by the adjoint $^\dagger : A \mapsto A^\dagger$, and $\|A^\dagger A\| = \|A\|^2$ where $\|\cdot\|$ is the operator norm.
- Note that $A \geq 0 \Rightarrow A^\dagger = A$ (pos. elements are Hermitian).
- We equip $\mathcal{B}(\mathcal{H})$ with the **Hilbert-Schmidt inner product**:

$$\langle X, Y \rangle := \mathrm{Tr}(X^\dagger Y)$$

Mathematics of Quantum Mechanics 101

A bit more abstract:

- Density operators correspond to positive, normalized elements of the C^*-algebra $\mathcal{B}(\mathcal{H})$ of linear bounded operators acting on a Hilbert space $\mathcal{H}$ (for us $\dim \mathcal{H} < \infty$).
- The $*$-map is given by the adjoint $^\dagger : A \mapsto A^\dagger$, and $\|A^\dagger A\| = \|A\|^2$ where $\| \cdot \|$ is the operator norm.
- Note that $A \geq 0 \Rightarrow A^\dagger = A$ (pos. elements are Hermitian).
- We equip $\mathcal{B}(\mathcal{H})$ with the **Hilbert-Schmidt inner product**:

$$\langle X, Y \rangle := \mathrm{Tr}(X^\dagger Y)$$

Mathematics of Quantum Mechanics 101

Dynamical evolution of a quantum system:

- A **quantum channel** (or quantum operation) is a map $\Lambda\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ that is
  1. **trace-preserving (TP):** $\mathrm{Tr}(\Lambda(X)) = \mathrm{Tr}\, X$ for all $X \in \mathcal{B}(\mathcal{H})$.
  2. **completely positive (CP):** The map

$$\Lambda \otimes \mathrm{id}_n \colon \mathcal{B}(\mathcal{H}) \otimes M_n(\mathbb{C}) \longrightarrow \mathcal{B}(\mathcal{K}) \otimes M_n(\mathbb{C})$$

is positive for all $n \in \mathbb{N}$. $(X \geq 0 \Rightarrow \Lambda \otimes \mathrm{id}_n(X) \geq 0)$

- Define the adjoint map $\Lambda^\dagger\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$ through

$$\langle \Lambda^\dagger(Y), X\rangle = \langle Y, \Lambda(X)\rangle \quad \text{for all } X \in \mathcal{B}(\mathcal{H}), Y \in \mathcal{B}(\mathcal{K}).$$

- Then Λ is TP iff $\Lambda^\dagger$ is unital, i.e. $\Lambda^\dagger(\mathbb{1}) = \mathbb{1}$.

Mathematics of Quantum Mechanics 101

Dynamical evolution of a quantum system:

- A **quantum channel** (or quantum operation) is a map $\Lambda\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ that is
  1. **trace-preserving (TP):** $\mathrm{Tr}(\Lambda(X)) = \mathrm{Tr}\, X$ for all $X \in \mathcal{B}(\mathcal{H})$.
  2. **completely positive (CP):** The map

$$\Lambda \otimes \mathrm{id}_n\colon \mathcal{B}(\mathcal{H}) \otimes M_n(\mathbb{C}) \longrightarrow \mathcal{B}(\mathcal{K}) \otimes M_n(\mathbb{C})$$

     is positive for all $n \in \mathbb{N}$. $(X \geq 0 \Rightarrow \Lambda \otimes \mathrm{id}_n(X) \geq 0)$

- Define the adjoint map $\Lambda^\dagger\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$ through

$$\langle \Lambda^\dagger(Y), X\rangle = \langle Y, \Lambda(X)\rangle \quad \text{for all } X \in \mathcal{B}(\mathcal{H}), Y \in \mathcal{B}(\mathcal{K}).$$

- Then Λ is TP iff $\Lambda^\dagger$ is unital, i.e. $\Lambda^\dagger(\mathbb{1}) = \mathbb{1}$.

Mathematics of Quantum Mechanics 101

Dynamical evolution of a quantum system:

- A **quantum channel** (or quantum operation) is a map $\Lambda\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ that is
  1. **trace-preserving (TP):** $\operatorname{Tr}(\Lambda(X)) = \operatorname{Tr} X$ for all $X \in \mathcal{B}(\mathcal{H})$.
  2. **completely positive (CP):** The map

     $$\Lambda \otimes \mathrm{id}_n\colon \mathcal{B}(\mathcal{H}) \otimes M_n(\mathbb{C}) \longrightarrow \mathcal{B}(\mathcal{K}) \otimes M_n(\mathbb{C})$$

     is positive for all $n \in \mathbb{N}$. $(X \geq 0 \Rightarrow \Lambda \otimes \mathrm{id}_n(X) \geq 0)$
- Define the adjoint map $\Lambda^\dagger\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$ through

  $$\langle \Lambda^\dagger(Y), X \rangle = \langle Y, \Lambda(X) \rangle \quad \text{for all } X \in \mathcal{B}(\mathcal{H}), Y \in \mathcal{B}(\mathcal{K}).$$
- Then Λ is TP iff $\Lambda^\dagger$ is unital, i.e. $\Lambda^\dagger(\mathbb{1}) = \mathbb{1}$.

Mathematics of Quantum Mechanics 101

Dynamical evolution of a quantum system:

- A **quantum channel** (or quantum operation) is a map $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ that is
  1. **trace-preserving (TP):** $\mathrm{Tr}(\Lambda(X)) = \mathrm{Tr}\, X$ for all $X \in \mathcal{B}(\mathcal{H})$.
  2. **completely positive (CP):** The map

$$\Lambda \otimes \mathrm{id}_n : \mathcal{B}(\mathcal{H}) \otimes M_n(\mathbb{C}) \longrightarrow \mathcal{B}(\mathcal{K}) \otimes M_n(\mathbb{C})$$

is positive for all $n \in \mathbb{N}$. $(X \geq 0 \Rightarrow \Lambda \otimes \mathrm{id}_n(X) \geq 0)$

- Define the adjoint map $\Lambda^\dagger : \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$ through

$$\langle \Lambda^\dagger(Y), X \rangle = \langle Y, \Lambda(X) \rangle \quad \text{for all } X \in \mathcal{B}(\mathcal{H}), Y \in \mathcal{B}(\mathcal{K}).$$

- Then Λ is TP iff $\Lambda^\dagger$ is unital, i.e. $\Lambda^\dagger(\mathbb{1}) = \mathbb{1}$.

Mathematics of Quantum Mechanics 101

Dynamical evolution of a quantum system:

- A **quantum channel** (or quantum operation) is a map $\Lambda\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ that is

  1. **trace-preserving (TP):** $\operatorname{Tr}(\Lambda(X)) = \operatorname{Tr} X$ for all $X \in \mathcal{B}(\mathcal{H})$.
  2. **completely positive (CP):** The map

  $$\Lambda \otimes \mathrm{id}_n\colon \mathcal{B}(\mathcal{H}) \otimes M_n(\mathbb{C}) \longrightarrow \mathcal{B}(\mathcal{K}) \otimes M_n(\mathbb{C})$$

  is positive for all $n \in \mathbb{N}$. $(X \geq 0 \Rightarrow \Lambda \otimes \mathrm{id}_n(X) \geq 0)$

- Define the adjoint map $\Lambda^\dagger\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$ through

  $$\langle \Lambda^\dagger(Y), X\rangle = \langle Y, \Lambda(X)\rangle \quad \text{for all } X \in \mathcal{B}(\mathcal{H}), Y \in \mathcal{B}(\mathcal{K}).$$

- Then Λ is TP iff $\Lambda^\dagger$ is unital, i.e. $\Lambda^\dagger(\mathbb{1}) = \mathbb{1}$.

Mathematics of Quantum Mechanics 101

Canonical example of quantum channel: Partial trace

- Consider two Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$.

- Define a linear map $\mathrm{Tr}_1 : \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_2)$ by

$$\mathrm{Tr}_1(X \otimes Y) = \mathrm{Tr}(X)Y$$

  for arbitrary $X \in \mathcal{B}(\mathcal{H}_1)$, $Y \in \mathcal{B}(\mathcal{H}_2)$.

- Trace-preserving: $\mathrm{Tr}(\mathrm{Tr}_1(X \otimes Y)) = \mathrm{Tr}(X)\,\mathrm{Tr}(Y) = \mathrm{Tr}(X \otimes Y)$

- Completely positive: $\mathrm{Tr}_1 \otimes \mathrm{id}_n$ is positive for all $n \in \mathbb{N}$.

Mathematics of Quantum Mechanics 101

Canonical example of quantum channel: Partial trace

- Consider two Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$.

- Define a linear map $\mathrm{Tr}_1 : \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_2)$ by

$$\mathrm{Tr}_1(X \otimes Y) = \mathrm{Tr}(X)Y$$

  for arbitrary $X \in \mathcal{B}(\mathcal{H}_1)$, $Y \in \mathcal{B}(\mathcal{H}_2)$.

- Trace-preserving: $\mathrm{Tr}(\mathrm{Tr}_1(X \otimes Y)) = \mathrm{Tr}(X)\,\mathrm{Tr}(Y) = \mathrm{Tr}(X \otimes Y)$

- Completely positive: $\mathrm{Tr}_1 \otimes \mathrm{id}_n$ is positive for all $n \in \mathbb{N}$.

Mathematics of Quantum Mechanics 101

Canonical example of quantum channel: Partial trace

- Consider two Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$.
- Define a linear map $\mathrm{Tr}_1 : \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_2)$ by

$$\mathrm{Tr}_1(X \otimes Y) = \mathrm{Tr}(X)Y$$

for arbitrary $X \in \mathcal{B}(\mathcal{H}_1)$, $Y \in \mathcal{B}(\mathcal{H}_2)$.

- Trace-preserving: $\mathrm{Tr}(\mathrm{Tr}_1(X \otimes Y)) = \mathrm{Tr}(X)\,\mathrm{Tr}(Y) = \mathrm{Tr}(X \otimes Y)$
- Completely positive: $\mathrm{Tr}_1 \otimes \mathrm{id}_n$ is positive for all $n \in \mathbb{N}$.

Mathematics of Quantum Mechanics 101

Canonical example of quantum channel: Partial trace

- Consider two Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$.

- Define a linear map $\text{Tr}_1 : \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_2)$ by

$$\text{Tr}_1(X \otimes Y) = \text{Tr}(X)Y$$

  for arbitrary $X \in \mathcal{B}(\mathcal{H}_1)$, $Y \in \mathcal{B}(\mathcal{H}_2)$.

- Trace-preserving: $\text{Tr}(\text{Tr}_1(X \otimes Y)) = \text{Tr}(X)\,\text{Tr}(Y) = \text{Tr}(X \otimes Y)$

- Completely positive: $\text{Tr}_1 \otimes \text{id}_n$ is positive for all $n \in \mathbb{N}$.

Mathematics of Quantum Mechanics 101

Importance of partial trace

1. **Stinespring's representation theorem** [Stinespring 1955]:
   Let $\Lambda\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a quantum channel, then there is an isometry $V\colon \mathcal{H} \to \mathcal{K} \otimes \mathcal{K}'$ s.t. $\Lambda(X) = \mathrm{Tr}_2(VXV^\dagger)$.
   (*Every* quantum ch. looks like the partial trace in some space.)

2. **Purification:** Let $\rho \in \mathcal{B}(\mathcal{H})$ be a mixed state, then there is a Hilbert space $\mathcal{H}'$ (we may take $\dim \mathcal{H}' = \dim \mathcal{H}$) and a pure state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}'$ such that $\rho = \mathrm{Tr}_2 |\psi\rangle\langle\psi|$.
   (*Every* state looks like a pure state in some space.)

Mathematics of Quantum Mechanics 101

Importance of partial trace

1. **Stinespring's representation theorem** [Stinespring 1955]:
   Let $\Lambda\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a quantum channel, then there is an isometry $V\colon \mathcal{H} \to \mathcal{K} \otimes \mathcal{K}'$ s.t. $\Lambda(X) = \mathrm{Tr}_2(VXV^\dagger)$.
   (*Every* quantum ch. looks like the partial trace in some space.)

2. **Purification:** Let $\rho \in \mathcal{B}(\mathcal{H})$ be a mixed state, then there is a Hilbert space $\mathcal{H}'$ (we may take $\dim \mathcal{H}' = \dim \mathcal{H}$) and a pure state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}'$ such that $\rho = \mathrm{Tr}_2 |\psi\rangle\langle\psi|$.
   (*Every* state looks like a pure state in some space.)

Mathematics of Quantum Mechanics 101

Operators and functional calculus

- We write $A \geq B$ if $A - B \geq 0$.
- Let $A \geq 0$ with spectral decomposition $A = \sum_i \lambda_i |e_i\rangle\langle e_i|$, and let $f\colon \mathbb{R}^+ \to \mathbb{R}$, then we define $f(A) := \sum_i f(\lambda_i)|e_i\rangle\langle e_i|$.
- f is operator monotone: $A \geq B$ implies $f(A) \geq f(B)$.
- f is operator convex: For $\lambda \in (0, 1)$ and $A_1, A_2 \geq 0$,

$$f(\lambda A_1 + (1-\lambda)A_2) \leq \lambda f(A_1) + (1-\lambda)f(A_2)$$

- Jensen's operator inequality: f is operator convex iff

$$f(V^\dagger AV) \leq V^\dagger f(A)V$$

for all $A \geq 0$ and V with $\|V\| \leq 1$.

Mathematics of Quantum Mechanics 101

Operators and functional calculus

- We write $A \geq B$ if $A - B \geq 0$.
- Let $A \geq 0$ with spectral decomposition $A = \sum_i \lambda_i |e_i\rangle\langle e_i|$, and let $f\colon \mathbb{R}^+ \to \mathbb{R}$, then we define $f(A) := \sum_i f(\lambda_i)|e_i\rangle\langle e_i|$.
- f is operator monotone: $A \geq B$ implies $f(A) \geq f(B)$.
- f is operator convex: For $\lambda \in (0,1)$ and $A_1, A_2 \geq 0$,

$$f(\lambda A_1 + (1-\lambda)A_2) \leq \lambda f(A_1) + (1-\lambda)f(A_2)$$

- Jensen's operator inequality: f is operator convex iff

$$f(V^\dagger A V) \leq V^\dagger f(A) V$$

for all $A \geq 0$ and V with $\|V\| \leq 1$.

Mathematics of Quantum Mechanics 101

Operators and functional calculus

- We write $A \geq B$ if $A - B \geq 0$.
- Let $A \geq 0$ with spectral decomposition $A = \sum_i \lambda_i |e_i\rangle\langle e_i|$, and let $f\colon \mathbb{R}^+ \to \mathbb{R}$, then we define $f(A) := \sum_i f(\lambda_i)|e_i\rangle\langle e_i|$.
- f is operator monotone: $A \geq B$ implies $f(A) \geq f(B)$.
- f is operator convex: For $\lambda \in (0,1)$ and $A_1, A_2 \geq 0$,

$$f(\lambda A_1 + (1-\lambda)A_2) \leq \lambda f(A_1) + (1-\lambda)f(A_2)$$

- Jensen's operator inequality: f is operator convex iff

$$f(V^\dagger A V) \leq V^\dagger f(A) V$$

for all $A \geq 0$ and V with $\|V\| \leq 1$.

Mathematics of Quantum Mechanics 101

Operators and functional calculus

- We write $A \geq B$ if $A - B \geq 0$.
- Let $A \geq 0$ with spectral decomposition $A = \sum_i \lambda_i |e_i\rangle\langle e_i|$, and let $f\colon \mathbb{R}^+ \to \mathbb{R}$, then we define $f(A) := \sum_i f(\lambda_i)|e_i\rangle\langle e_i|$.
- f is operator monotone: $A \geq B$ implies $f(A) \geq f(B)$.
- f is operator convex: For $\lambda \in (0,1)$ and $A_1, A_2 \geq 0$,

$$f(\lambda A_1 + (1-\lambda)A_2) \leq \lambda f(A_1) + (1-\lambda)f(A_2)$$

- Jensen's operator inequality: f is operator convex iff

$$f(V^\dagger A V) \leq V^\dagger f(A) V$$

for all $A \geq 0$ and V with $\|V\| \leq 1$.

Mathematics of Quantum Mechanics 101

Operators and functional calculus

- We write $A \geq B$ if $A - B \geq 0$.
- Let $A \geq 0$ with spectral decomposition $A = \sum_i \lambda_i |e_i\rangle\langle e_i|$, and let $f\colon \mathbb{R}^+ \to \mathbb{R}$, then we define $f(A) := \sum_i f(\lambda_i)|e_i\rangle\langle e_i|$.
- f is operator monotone: $A \geq B$ implies $f(A) \geq f(B)$.
- f is operator convex: For $\lambda \in (0,1)$ and $A_1, A_2 \geq 0$,

$$f(\lambda A_1 + (1-\lambda)A_2) \leq \lambda f(A_1) + (1-\lambda) f(A_2)$$

- Jensen's operator inequality: f is operator convex iff

$$f(V^\dagger A V) \leq V^\dagger f(A) V$$

for all $A \geq 0$ and V with $\|V\| \leq 1$.

Table of Contents

Quantum relative entropy

Let $X, Y \in \mathcal{B}(\mathcal{H})$, $X, Y \geq 0$ with $\operatorname{supp} X \subseteq \operatorname{supp} Y$, then

$$D(X \| Y) := \operatorname{Tr}[X(\log X - \log Y)]$$

Properties:

- If X and Y are states, then $D(X\|Y) \geq 0$, and $= 0$ iff $X = Y$.
- Correct quantum generalization of KL-divergence:
  - $D(\hat{P}\|\hat{Q}) = D_{KL}(P\|Q)$ for *classical* states

    $$\hat{P} = \sum_{x \in \mathcal{X}} P(x)|x\rangle\langle x| \qquad \hat{Q} = \sum_{x \in \mathcal{X}} Q(x)|x\rangle\langle x|$$

  - error exponent in *quantum hypothesis testing*

    (just as $D_{KL}(\cdot\|\cdot)$ in classical hyp. testing)

Quantum relative entropy

Let $X, Y \in \mathcal{B}(\mathcal{H})$, $X, Y \geq 0$ with $\operatorname{supp} X \subseteq \operatorname{supp} Y$, then

$$D(X\|Y) := \mathrm{Tr}[X(\log X - \log Y)]$$

Properties:

- If X and Y are states, then $D(X\|Y) \geq 0$, and $= 0$ iff $X = Y$.
- Correct quantum generalization of KL-divergence:
  - $D(\hat{P}\|\hat{Q}) = D_{\mathrm{KL}}(P\|Q)$ for *classical* states

$$\hat{P} = \sum_{x \in \mathcal{X}} P(x)|x\rangle\langle x| \qquad \hat{Q} = \sum_{x \in \mathcal{X}} Q(x)|x\rangle\langle x|.$$

  - error exponent in *quantum hypothesis testing*

    (just as $D_{\mathrm{KL}}(\cdot\|\cdot)$ in classical hyp. testing)

Quantum relative entropy

Let $X, Y \in \mathcal{B}(\mathcal{H})$, $X, Y \geq 0$ with $\operatorname{supp} X \subseteq \operatorname{supp} Y$, then

$$D(X\|Y) := \operatorname{Tr}[X(\log X - \log Y)]$$

Properties:

- If X and Y are states, then $D(X\|Y) \geq 0$, and $= 0$ iff $X = Y$.
- Correct quantum generalization of KL-divergence:
  - $D(\hat{P}\|\hat{Q}) = D_{\mathrm{KL}}(P\|Q)$ for *classical* states

    $$\hat{P} = \sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x| \qquad \hat{Q} = \sum_{x\in\mathcal{X}} Q(x)|x\rangle\langle x|.$$
  - error exponent in *quantum hypothesis testing*

    (just as $D_{\mathrm{KL}}(\cdot\|\cdot)$ in classical hyp. testing)

Quantum relative entropy

Let $X, Y \in \mathcal{B}(\mathcal{H})$, $X, Y \geq 0$ with $\operatorname{supp} X \subseteq \operatorname{supp} Y$, then

$$D(X\|Y) := \operatorname{Tr}[X(\log X - \log Y)]$$

Properties:

- If X and Y are states, then $D(X\|Y) \geq 0$, and $= 0$ iff $X = Y$.
- Correct quantum generalization of KL-divergence:
  - $D(\hat{P}\|\hat{Q}) = D_{\mathrm{KL}}(P\|Q)$ for *classical* states

$$\hat{P} = \sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x| \qquad \hat{Q} = \sum_{x\in\mathcal{X}} Q(x)|x\rangle\langle x|.$$

  - error exponent in *quantum hypothesis testing*

    (just as $D_{\mathrm{KL}}(\cdot\|\cdot)$ in classical hyp. testing)

Quantum relative entropy

Let $X, Y \in \mathcal{B}(\mathcal{H})$, $X, Y \geq 0$ with $\operatorname{supp} X \subseteq \operatorname{supp} Y$, then

$$D(X\|Y) := \mathrm{Tr}[X(\log X - \log Y)]$$

Properties:

- If X and Y are states, then $D(X\|Y) \geq 0$, and $= 0$ iff $X = Y$.
- Correct quantum generalization of KL-divergence:
  - $D(\hat{P}\|\hat{Q}) = D_{\mathrm{KL}}(P\|Q)$ for *classical* states

    $$\hat{P} = \sum_{x \in \mathcal{X}} P(x)|x\rangle\langle x| \qquad \hat{Q} = \sum_{x \in \mathcal{X}} Q(x)|x\rangle\langle x|.$$
  - error exponent in *quantum hypothesis testing*

    (just as $D_{\mathrm{KL}}(\cdot\|\cdot)$ in classical hyp. testing)

Quantum relative entropy

- Let $\Phi \colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be TP and 2-positive map.

  2-positive: For $A_i \in \mathcal{B}(\mathcal{H}), i = 1, \ldots, 4$ we have

$$\begin{pmatrix} \Phi(A_1) & \Phi(A_2) \\ \Phi(A_3) & \Phi(A_4) \end{pmatrix} \geq 0 \quad \text{if} \quad \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} \geq 0.$$

- Then $D(\cdot\|\cdot)$ satisfies the **data processing inequality**

$$D(X\|Y) \geq D(\Phi(X)\|\Phi(Y)).$$

- This holds in particular for every CPTP map Λ.

  Recall: Φ CP $:\Leftrightarrow$ Φ is n-positive for all n

- Recent result: DPI holds for every positive map.

  [Müller-Hermes and Reeb 2015]

Quantum relative entropy

- Let $\Phi\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be TP and 2-positive map.

  2-positive: For $A_i \in \mathcal{B}(\mathcal{H}), i = 1, \dots, 4$ we have

$$\begin{pmatrix} \Phi(A_1) & \Phi(A_2) \\ \Phi(A_3) & \Phi(A_4) \end{pmatrix} \geq 0 \quad \text{if} \quad \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} \geq 0.$$

- Then $D(\cdot\|\cdot)$ satisfies the **data processing inequality**

$$D(X\|Y) \geq D(\Phi(X)\|\Phi(Y)).$$

- This holds in particular for every CPTP map Λ.

  Recall: Φ CP $:\Leftrightarrow$ Φ is n-positive for all n

- Recent result: DPI holds for every positive map.

  [Müller-Hermes and Reeb 2015]

Quantum relative entropy

- Let $\Phi\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be TP and 2-positive map.

  2-positive: For $A_i \in \mathcal{B}(\mathcal{H}), i = 1, \ldots, 4$ we have

$$\begin{pmatrix} \Phi(A_1) & \Phi(A_2) \\ \Phi(A_3) & \Phi(A_4) \end{pmatrix} \geq 0 \quad \text{if} \quad \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} \geq 0.$$

- Then $D(\cdot\|\cdot)$ satisfies the **data processing inequality**

$$D(X\|Y) \geq D(\Phi(X)\|\Phi(Y)).$$

- This holds in particular for every CPTP map Λ.

  Recall: Φ CP $:\Leftrightarrow$ Φ is n-positive for all n

- Recent result: DPI holds for every positive map.

  [Müller-Hermes and Reeb 2015]

Quantum relative entropy

- Let $\Phi\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be TP and 2-positive map.

  2-positive: For $A_i \in \mathcal{B}(\mathcal{H}), i = 1, \ldots, 4$ we have

$$\begin{pmatrix} \Phi(A_1) & \Phi(A_2) \\ \Phi(A_3) & \Phi(A_4) \end{pmatrix} \geq 0 \quad \text{if} \quad \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} \geq 0.$$

- Then $D(\cdot\|\cdot)$ satisfies the **data processing inequality**

$$D(X\|Y) \geq D(\Phi(X)\|\Phi(Y)).$$

- This holds in particular for every CPTP map Λ.

  Recall: Φ CP $:\Leftrightarrow$ Φ is n-positive for all n

- Recent result: DPI holds for every positive map.

  [Müller-Hermes and Reeb 2015]

Table of Contents

Equality in data processing inequality

Theorem (Petz 1988)

Let $\Phi \colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ *be a 2-positive TP map, and let* $X, Y \in \mathcal{B}(\mathcal{H})$ *be invertible density operators. Then we have*

$$D(X\|Y) = D(\Phi(X)\|\Phi(Y))$$

if and only if for all $t \in \mathbb{R}$

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it}.$$

Remark: Assumption of invertible X, Y can be relaxed to $\operatorname{supp} X \subseteq \operatorname{supp} Y$.

Equality in data processing inequality

Theorem (Petz 1988)

Let $\Phi\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ *be a 2-positive TP map, and let* $X, Y \in \mathcal{B}(\mathcal{H})$ *be invertible density operators. Then we have*

$$D(X\|Y) = D(\Phi(X)\|\Phi(Y))$$

if and only if for all $t \in \mathbb{R}$

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it}.$$

Remark: Assumption of invertible X, Y can be relaxed to $\operatorname{supp} X \subseteq \operatorname{supp} Y$.

Equality in data processing inequality

$D(X\|Y) = D(\Phi(X)\|\Phi(Y))$ if and only if

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it} \quad \text{for all } t \in \mathbb{R}.$$

- Algebraic condition on map and operators.
- Equivalent formulation: There exists a **recovery map**

$$\mathcal{R}_{\Phi,Y}(\cdot) = Y^{1/2}\Phi^\dagger\left(\Phi(Y)^{-1/2} \cdot \Phi(Y)^{-1/2}\right) Y^{1/2}$$

  such that $\mathcal{R}_{\Phi,Y}(\Phi(X)) = X$ and $\mathcal{R}_{\Phi,Y}(\Phi(Y)) = Y$.
- $\mathcal{R}_{\Phi,Y}$ recovers X, Y by *reverting* the action of Φ.
- Recovery on X:

$$X = Y^{1/2}\Phi^\dagger\left(\Phi(Y)^{-1/2}\Phi(X)\Phi(Y)^{-1/2}\right) Y^{1/2}$$

Equality in data processing inequality

$D(X\|Y) = D(\Phi(X)\|\Phi(Y))$ if and only if

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it} \quad \text{for all } t \in \mathbb{R}.$$

- Algebraic condition on map and operators.
- Equivalent formulation: There exists a **recovery map**

$$\mathcal{R}_{\Phi,Y}(\cdot) = Y^{1/2}\Phi^\dagger\left(\Phi(Y)^{-1/2} \cdot \Phi(Y)^{-1/2}\right) Y^{1/2}$$

such that $\mathcal{R}_{\Phi,Y}(\Phi(X)) = X$ and $\mathcal{R}_{\Phi,Y}(\Phi(Y)) = Y$.

- $\mathcal{R}_{\Phi,Y}$ recovers X, Y by *reverting* the action of Φ.
- Recovery on X:

$$X = Y^{1/2}\Phi^\dagger\left(\Phi(Y)^{-1/2}\Phi(X)\Phi(Y)^{-1/2}\right) Y^{1/2}$$

Equality in data processing inequality

$D(X\|Y) = D(\Phi(X)\|\Phi(Y))$ if and only if

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it} \quad \text{for all } t \in \mathbb{R}.$$

- Algebraic condition on map and operators.
- Equivalent formulation: There exists a **recovery map**

$$\mathcal{R}_{\Phi,Y}(\cdot) = Y^{1/2}\Phi^\dagger\left(\Phi(Y)^{-1/2} \cdot \Phi(Y)^{-1/2}\right) Y^{1/2}$$

  such that $\mathcal{R}_{\Phi,Y}(\Phi(X)) = X$ and $\mathcal{R}_{\Phi,Y}(\Phi(Y)) = Y$.
- $\mathcal{R}_{\Phi,Y}$ recovers X, Y by *reverting* the action of Φ.
- Recovery on X:

$$X = Y^{1/2}\Phi^\dagger\left(\Phi(Y)^{-1/2}\Phi(X)\Phi(Y)^{-1/2}\right) Y^{1/2}$$

Equality in data processing inequality

$D(X\|Y) = D(\Phi(X)\|\Phi(Y))$ if and only if

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it} \quad \text{for all } t \in \mathbb{R}.$$

- Algebraic condition on map and operators.
- Equivalent formulation: There exists a **recovery map**

$$\mathcal{R}_{\Phi,Y}(\cdot) = Y^{1/2}\Phi^\dagger\left(\Phi(Y)^{-1/2} \cdot \Phi(Y)^{-1/2}\right) Y^{1/2}$$

  such that $\mathcal{R}_{\Phi,Y}(\Phi(X)) = X$ and $\mathcal{R}_{\Phi,Y}(\Phi(Y)) = Y$.
- $\mathcal{R}_{\Phi,Y}$ recovers X, Y by *reverting* the action of Φ.
- Recovery on X:

$$X = Y^{1/2}\Phi^\dagger\left(\Phi(Y)^{-1/2}\Phi(X)\Phi(Y)^{-1/2}\right) Y^{1/2}$$

Proof of the main theorem

- We first analyze a proof of the data processing inequality via **relative modular operators**.
- Consider the multiplication operators $L_A(T) := AT$ and $R_B(T) = TB$, satisfying
  - $L_A \circ R_B = R_B \circ L_A$.
  - $L_{A^{-1}} = L_A^{-1}$ if A is invertible, likewise for R_B.
  - L_A, R_B are self-adjoint, and positive if $A, B \geq 0$.
  - For analytic $f \colon \mathbb{R}^+ \to \mathbb{R}$ we have $f(L_A) = L_{f(A)}$ and likewise for R_B if $A, B \geq 0$.

Proof of the main theorem

- We first analyze a proof of the data processing inequality via **relative modular operators**.
- Consider the multiplication operators $L_A(T) := AT$ and $R_B(T) = TB$, satisfying
  - $L_A \circ R_B = R_B \circ L_A$.
  - $L_{A^{-1}} = L_A^{-1}$ if A is invertible, likewise for R_B.
  - L_A, R_B are self-adjoint, and positive if $A, B \geq 0$.
  - For analytic $f\colon \mathbb{R}^+ \to \mathbb{R}$ we have $f(L_A) = L_{f(A)}$ and likewise for R_B if $A, B \geq 0$.

Proof of the main theorem

- We first analyze a proof of the data processing inequality via **relative modular operators**.
- Consider the multiplication operators $L_A(T) := AT$ and $R_B(T) = TB$, satisfying
  - $L_A \circ R_B = R_B \circ L_A$.
  - $L_{A^{-1}} = L_A^{-1}$ if A is invertible, likewise for R_B.
  - L_A, R_B are self-adjoint, and positive if $A, B \geq 0$.
  - For analytic $f\colon \mathbb{R}^+ \to \mathbb{R}$ we have $f(L_A) = L_{f(A)}$ and likewise for R_B if $A, B \geq 0$.

Proof of the main theorem

- We first analyze a proof of the data processing inequality via **relative modular operators**.
- Consider the multiplication operators $L_A(T) := AT$ and $R_B(T) = TB$, satisfying
  - $L_A \circ R_B = R_B \circ L_A$.
  - $L_{A^{-1}} = L_A^{-1}$ if A is invertible, likewise for R_B.
  - L_A, R_B are self-adjoint, and positive if $A, B \geq 0$.
  - For analytic $f \colon \mathbb{R}^+ \to \mathbb{R}$ we have $f(L_A) = L_{f(A)}$ and likewise for R_B if $A, B \geq 0$.

Proof of the main theorem

- We first analyze a proof of the data processing inequality via **relative modular operators**.
- Consider the multiplication operators $L_A(T) := AT$ and $R_B(T) = TB$, satisfying
  - $L_A \circ R_B = R_B \circ L_A$.
  - $L_{A^{-1}} = L_A^{-1}$ if A is invertible, likewise for R_B.
  - L_A, R_B are self-adjoint, and positive if $A, B \geq 0$.
  - For analytic $f \colon \mathbb{R}^+ \to \mathbb{R}$ we have $f(L_A) = L_{f(A)}$ and likewise for R_B if $A, B \geq 0$.

Proof of the main theorem

- We first analyze a proof of the data processing inequality via **relative modular operators**.
- Consider the multiplication operators $L_A(T) := AT$ and $R_B(T) = TB$, satisfying
  - $L_A \circ R_B = R_B \circ L_A$.
  - $L_{A^{-1}} = L_A^{-1}$ if A is invertible, likewise for R_B.
  - L_A, R_B are self-adjoint, and positive if $A, B \geq 0$.
  - For analytic $f \colon \mathbb{R}^+ \to \mathbb{R}$ we have $f(L_A) = L_{f(A)}$ and likewise for R_B if $A, B \geq 0$.

Proof of the main theorem

- Define the **relative modular operator** $\Delta_{Y|X} = L_Y \circ R_{X^{-1}}$.
- Then $\log \Delta_{Y|X} = L_{\log Y} - R_{\log X}$, and

$$D(X\|Y) = \mathrm{Tr}[X(\log X - \log Y)] = -\langle X^{1/2}, \log \Delta_{Y|X}(X^{1/2})\rangle.$$

- Assume now that $\Phi(X)$ is also invertible, and set

$$\Delta \equiv \Delta_{Y|X} \qquad\qquad \Delta_\Phi \equiv \Delta_{\Phi(Y)|\Phi(X)}$$

such that

$$D(X\|Y) = -\langle X^{1/2}, \log \Delta(X^{1/2})\rangle$$

$$D(\Phi(X)\|\Phi(Y)) = -\langle \Phi(X)^{1/2}, \log \Delta_\Phi(\Phi(X)^{1/2})\rangle.$$

Proof of the main theorem

- Define the **relative modular operator** $\Delta_{Y|X} = L_Y \circ R_{X^{-1}}$.
- Then $\log \Delta_{Y|X} = L_{\log Y} - R_{\log X}$, and

$$D(X\|Y) = \mathrm{Tr}[X(\log X - \log Y)] = -\langle X^{1/2}, \log \Delta_{Y|X}(X^{1/2})\rangle.$$

- Assume now that $\Phi(X)$ is also invertible, and set

$$\Delta \equiv \Delta_{Y|X} \qquad \Delta_\Phi \equiv \Delta_{\Phi(Y)|\Phi(X)}$$

such that

$$D(X\|Y) = -\langle X^{1/2}, \log \Delta(X^{1/2})\rangle$$

$$D(\Phi(X)\|\Phi(Y)) = -\langle \Phi(X)^{1/2}, \log \Delta_\Phi(\Phi(X)^{1/2})\rangle.$$

Proof of the main theorem

- Define the **relative modular operator** $\Delta_{Y|X} = L_Y \circ R_{X^{-1}}$.
- Then $\log \Delta_{Y|X} = L_{\log Y} - R_{\log X}$, and

$$D(X\|Y) = \mathrm{Tr}[X(\log X - \log Y)] = -\langle X^{1/2}, \log \Delta_{Y|X}(X^{1/2})\rangle.$$

- Assume now that $\Phi(X)$ is also invertible, and set

$$\Delta \equiv \Delta_{Y|X} \qquad\qquad \Delta_\Phi \equiv \Delta_{\Phi(Y)|\Phi(X)}$$

such that

$$D(X\|Y) = -\langle X^{1/2}, \log \Delta(X^{1/2})\rangle$$
$$D(\Phi(X)\|\Phi(Y)) = -\langle \Phi(X)^{1/2}, \log \Delta_\Phi(\Phi(X)^{1/2})\rangle.$$

Proof of the main theorem

- Consider the integral representation

$$\log x = \int_0^\infty \frac{1}{1+t} - \frac{1}{x+t}\, dt.$$

- We can then write

$$D(X\|Y) = \int_0^\infty \langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2})\rangle - (1+t)^{-1}\, dt$$

$$D(\Phi(X)\|\Phi(Y)) = \int_0^\infty \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle - (1+t)^{-1}\, dt$$

and focus on the integrands written in terms of the **resolvents** $(\Delta_{(\Phi)} + t)^{-1}$.

Proof of the main theorem

- Consider the integral representation

$$\log x = \int_0^\infty \frac{1}{1+t} - \frac{1}{x+t}\, dt.$$

- We can then write

$$D(X\|Y) = \int_0^\infty \langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2})\rangle - (1+t)^{-1}\, dt$$

$$D(\Phi(X)\|\Phi(Y)) = \int_0^\infty \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle - (1+t)^{-1}\, dt$$

and focus on the integrands written in terms of the **resolvents** $(\Delta_{(\Phi)} + t)^{-1}$.

Proof of the main theorem

- Define a linear map $V\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$:

$$V(A) := \Phi^\dagger\left(A\Phi(X)^{-1/2}\right) X^{1/2}.$$

- $\Phi^\dagger$ is unital: $V(\Phi(X)^{1/2}) = \Phi^\dagger(\mathbb{1})X^{1/2} = X^{1/2}$.
- V is a *contraction*: $\|V(A)\|^2 \leq \|A\|$ for all A.
- To show this, use the Schwarz inequality

$$\Phi^\dagger(A^\dagger A) \geq \Phi^\dagger(A^\dagger)\Phi^\dagger(A).$$

2-positive TP maps largest class of maps for which SI holds!

- V relates the two modular operators:

$$V^\dagger \Delta V \leq \Delta_\Phi,$$

again by application of the Schwarz inequality.

Proof of the main theorem

- Define a linear map $V\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$:

$$V(A) := \Phi^\dagger\left(A\Phi(X)^{-1/2}\right) X^{1/2}.$$

- $\Phi^\dagger$ is unital: $V(\Phi(X)^{1/2}) = \Phi^\dagger(\mathbb{1})X^{1/2} = X^{1/2}$.
- V is a *contraction*: $\|V(A)\|^2 \leq \|A\|$ for all A.
- To show this, use the Schwarz inequality

$$\Phi^\dagger(A^\dagger A) \geq \Phi^\dagger(A^\dagger)\Phi^\dagger(A).$$

2-positive TP maps largest class of maps for which SI holds!

- V relates the two modular operators:

$$V^\dagger \Delta V \leq \Delta_\Phi,$$

again by application of the Schwarz inequality.

Proof of the main theorem

- Define a linear map $V\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$:

$$V(A) := \Phi^\dagger\left(A\Phi(X)^{-1/2}\right) X^{1/2}.$$

- $\Phi^\dagger$ is unital: $V(\Phi(X)^{1/2}) = \Phi^\dagger(\mathbb{1})X^{1/2} = X^{1/2}$.
- V is a *contraction*: $\|V(A)\|^2 \leq \|A\|$ for all A.
- To show this, use the Schwarz inequality

$$\Phi^\dagger(A^\dagger A) \geq \Phi^\dagger(A^\dagger)\Phi^\dagger(A).$$

  2-positive TP maps largest class of maps for which SI holds!

- V relates the two modular operators:

$$V^\dagger \Delta V \leq \Delta_\Phi,$$

  again by application of the Schwarz inequality.

Proof of the main theorem

- Define a linear map $V\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$:

$$V(A) := \Phi^\dagger\left(A\Phi(X)^{-1/2}\right) X^{1/2}.$$

- $\Phi^\dagger$ is unital: $V(\Phi(X)^{1/2}) = \Phi^\dagger(\mathbb{1})X^{1/2} = X^{1/2}$.
- V is a *contraction*: $\|V(A)\|^2 \leq \|A\|$ for all A.
- To show this, use the Schwarz inequality

$$\Phi^\dagger(A^\dagger A) \geq \Phi^\dagger(A^\dagger)\Phi^\dagger(A).$$

2-positive TP maps largest class of maps for which SI holds!

- V relates the two modular operators:

$$V^\dagger \Delta V \leq \Delta_\Phi,$$

again by application of the Schwarz inequality.

Proof of the main theorem

- Define a linear map $V\colon \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$:

$$V(A) := \Phi^\dagger\left(A\Phi(X)^{-1/2}\right) X^{1/2}.$$

- $\Phi^\dagger$ is unital: $V(\Phi(X)^{1/2}) = \Phi^\dagger(\mathbb{1})X^{1/2} = X^{1/2}$.
- V is a *contraction*: $\|V(A)\|^2 \leq \|A\|$ for all A.
- To show this, use the Schwarz inequality

$$\Phi^\dagger(A^\dagger A) \geq \Phi^\dagger(A^\dagger)\Phi^\dagger(A).$$

2-positive TP maps largest class of maps for which SI holds!

- V relates the two modular operators:

$$V^\dagger \Delta V \leq \Delta_\Phi,$$

again by application of the Schwarz inequality.

Proof of the main theorem

- V is a contraction, and $V^\dagger \Delta V \leq \Delta_\Phi$.
- $y \mapsto (y+t)^{-1}$ is operator monotone (decreasing) and operator convex:

$$(\Delta_\Phi + t)^{-1} \leq (V^\dagger \Delta V + t)^{-1} \leq V^\dagger (\Delta + t)^{-1} V$$

(second $\leq$ follows from Jensen's operator inequality)

- Hence:

$$\begin{aligned}\langle X^{1/2}, (\Delta+t)^{-1}(X^{1/2})\rangle &= \langle V\Phi(X)^{1/2}, (\Delta+t)^{1/2} V(\Phi(X)^{1/2})\rangle \\ &= \langle \Phi(X)^{1/2}, V^\dagger (\Delta+t)^{1/2} V(\Phi(X)^{1/2})\rangle \\ &\geq \langle \Phi(X)^{1/2}, (\Delta_\Phi+t)^{-1}(\Phi(X)^{1/2})\rangle.\end{aligned}$$

Proof of the main theorem

- V is a contraction, and $V^\dagger \Delta V \leq \Delta_\Phi$.
- $y \mapsto (y+t)^{-1}$ is operator monotone (decreasing) and operator convex:

$$(\Delta_\Phi + t)^{-1} \leq (V^\dagger \Delta V + t)^{-1} \leq V^\dagger (\Delta + t)^{-1} V$$

(second $\leq$ follows from Jensen's operator inequality)

- Hence:

$$\begin{aligned} \langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2}) \rangle &= \langle V\Phi(X)^{1/2}, (\Delta + t)^{1/2} V(\Phi(X)^{1/2}) \rangle \\ &= \langle \Phi(X)^{1/2}, V^\dagger (\Delta + t)^{1/2} V(\Phi(X)^{1/2}) \rangle \\ &\geq \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2}) \rangle. \end{aligned}$$

Proof of the main theorem

- V is a contraction, and $V^\dagger \Delta V \leq \Delta_\Phi$.
- $y \mapsto (y+t)^{-1}$ is operator monotone (decreasing) and operator convex:

$$(\Delta_\Phi + t)^{-1} \leq (V^\dagger \Delta V + t)^{-1} \leq V^\dagger (\Delta + t)^{-1} V$$

(second $\leq$ follows from Jensen's operator inequality)

- Hence:

$$\begin{aligned}\langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2})\rangle &= \langle V\Phi(X)^{1/2}, (\Delta + t)^{1/2} V(\Phi(X)^{1/2})\rangle \\ &= \langle \Phi(X)^{1/2}, V^\dagger (\Delta + t)^{1/2} V(\Phi(X)^{1/2})\rangle \\ &\geq \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle.\end{aligned}$$

Proof of the main theorem

- Recall integral representations:

$$D(X\|Y) = \int_0^\infty \langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2})\rangle - (1+t)^{-1}\, dt$$

$$D(\Phi(X)\|\Phi(Y)) = \int_0^\infty \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle - (1+t)^{-1}\, dt$$

- Just proved:

$$\langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2})\rangle \geq \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle.$$

- Insert this in the integral representations:

$$D(X\|Y) \geq D(\Phi(X)\|\Phi(Y))$$

Proof of the main theorem

- Recall integral representations:

$$D(X\|Y) = \int_0^\infty \langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2})\rangle - (1+t)^{-1}\, dt$$

$$D(\Phi(X)\|\Phi(Y)) = \int_0^\infty \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle - (1+t)^{-1}\, dt$$

- Just proved:

$$\langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2})\rangle \geq \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle.$$

- Insert this in the integral representations:

$$D(X\|Y) \geq D(\Phi(X)\|\Phi(Y))$$

Proof of the main theorem

- Recall integral representations:

$$D(X\|Y) = \int_0^\infty \langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2})\rangle - (1+t)^{-1}\, dt$$

$$D(\Phi(X)\|\Phi(Y)) = \int_0^\infty \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle - (1+t)^{-1}\, dt$$

- Just proved:

$$\langle X^{1/2}, (\Delta + t)^{-1}(X^{1/2})\rangle \geq \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle.$$

- Insert this in the integral representations:

$$D(X\|Y) \geq D(\Phi(X)\|\Phi(Y))$$

Proof of the main theorem

- Equality in DPI if and only if for all $t > 0$

$$\langle \Phi(X)^{1/2}, V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2}) \rangle = \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2}) \rangle.$$

- Let $A \geq B$, then $\langle v|A|v\rangle = \langle v|B|v\rangle$ implies $A|v\rangle = B|v\rangle$.
- Hence, $V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2}) = (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})$.
- It follows by an easy calculation that

$$\|V^\dagger(\Delta + t)^{-1}(X^{1/2})\|^2 = \|V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2})\|^2 = \|(\Delta + t)^{-1}(X^{1/2})\|^2.$$

Proof of the main theorem

- Equality in DPI if and only if for all $t > 0$

$$\langle \Phi(X)^{1/2}, V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2}) \rangle = \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2}) \rangle.$$

- Let $A \geq B$, then $\langle v|A|v\rangle = \langle v|B|v\rangle$ implies $A|v\rangle = B|v\rangle$.
- Hence, $V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2}) = (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})$.
- It follows by an easy calculation that

$$\|V^\dagger(\Delta + t)^{-1}(X^{1/2})\|^2 = \|V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2})\|^2 = \|(\Delta + t)^{-1}(X^{1/2})\|^2.$$

Proof of the main theorem

- Equality in DPI if and only if for all $t > 0$

$$\langle \Phi(X)^{1/2}, V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2}) \rangle = \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1} (\Phi(X)^{1/2}) \rangle .$$

- Let $A \geq B$, then $\langle v|A|v\rangle = \langle v|B|v\rangle$ implies $A|v\rangle = B|v\rangle$.
- Hence, $V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2}) = (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})$.
- It follows by an easy calculation that

$$\|V^\dagger(\Delta + t)^{-1}(X^{1/2})\|^2 = \|V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2})\|^2 = \|(\Delta + t)^{-1}(X^{1/2})\|^2 .$$

Proof of the main theorem

- Equality in DPI if and only if for all $t > 0$

$$\langle \Phi(X)^{1/2}, V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2})\rangle = \langle \Phi(X)^{1/2}, (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})\rangle.$$

- Let $A \geq B$, then $\langle v|A|v\rangle = \langle v|B|v\rangle$ implies $A|v\rangle = B|v\rangle$.
- Hence, $V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2}) = (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})$.
- It follows by an easy calculation that

$$\|V^\dagger(\Delta + t)^{-1}(X^{1/2})\|^2 = \|V^\dagger(\Delta + t)^{-1} V(\Phi(X)^{1/2})\|^2 = \|(\Delta + t)^{-1}(X^{1/2})\|^2.$$

Proof of the main theorem

- $\|V^\dagger(\Delta + t)^{-1}(X^{1/2})\| = \|(\Delta + t)^{-1}(X^{1/2})\|^2$.
- If $\|W^\dagger|\xi\rangle\|^2 = \||\xi\rangle\|^2$ for an arbitrary contraction W, then $WW^\dagger|\xi\rangle = |\xi\rangle$
- Recall: $V^\dagger(\Delta + t)^{-1}(X^{1/2}) = (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})$
- Then: $V(\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2}) = (\Delta + t)^{-1}(X^{1/2})$.
- That is, the resolvents of Δ_Φ and Δ coincide on the vectors $\Phi(X)^{1/2}$ and $X^{1/2}$, respectively (modulo V).

Proof of the main theorem

- $\|V^\dagger(\Delta + t)^{-1}(X^{1/2})\| = \|(\Delta + t)^{-1}(X^{1/2})\|^2$.
- If $\|W^\dagger|\xi\rangle\|^2 = \||\xi\rangle\|^2$ for an arbitrary contraction W, then $WW^\dagger|\xi\rangle = |\xi\rangle$
- Recall: $V^\dagger(\Delta + t)^{-1}(X^{1/2}) = (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})$
- Then: $V(\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2}) = (\Delta + t)^{-1}(X^{1/2})$.
- That is, the resolvents of Δ_Φ and Δ coincide on the vectors $\Phi(X)^{1/2}$ and $X^{1/2}$, respectively (modulo V).

Proof of the main theorem

- $\|V^\dagger(\Delta+t)^{-1}(X^{1/2})\| = \|(\Delta+t)^{-1}(X^{1/2})\|^2$.
- If $\|W^\dagger|\xi\rangle\|^2 = \||\xi\rangle\|^2$ for an arbitrary contraction W, then $WW^\dagger|\xi\rangle = |\xi\rangle$
- Recall: $V^\dagger(\Delta+t)^{-1}(X^{1/2}) = (\Delta_\Phi+t)^{-1}(\Phi(X)^{1/2})$
- Then: $V(\Delta_\Phi+t)^{-1}(\Phi(X)^{1/2}) = (\Delta+t)^{-1}(X^{1/2})$.
- That is, the resolvents of Δ_Φ and Δ coincide on the vectors $\Phi(X)^{1/2}$ and $X^{1/2}$, respectively (modulo V).

Proof of the main theorem

- $\|V^\dagger(\Delta + t)^{-1}(X^{1/2})\| = \|(\Delta + t)^{-1}(X^{1/2})\|^2$.
- If $\|W^\dagger|\xi\rangle\|^2 = \||\xi\rangle\|^2$ for an arbitrary contraction W, then $WW^\dagger|\xi\rangle = |\xi\rangle$
- Recall: $V^\dagger(\Delta + t)^{-1}(X^{1/2}) = (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})$
- Then: $V(\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2}) = (\Delta + t)^{-1}(X^{1/2})$.
- That is, the resolvents of Δ_Φ and Δ coincide on the vectors $\Phi(X)^{1/2}$ and $X^{1/2}$, respectively (modulo V).

Proof of the main theorem

- $\|V^\dagger(\Delta + t)^{-1}(X^{1/2})\| = \|(\Delta + t)^{-1}(X^{1/2})\|^2$.

- If $\|W^\dagger|\xi\rangle\|^2 = \||\xi\rangle\|^2$ for an arbitrary contraction W, then $WW^\dagger|\xi\rangle = |\xi\rangle$

- Recall: $V^\dagger(\Delta + t)^{-1}(X^{1/2}) = (\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2})$

- Then: $V(\Delta_\Phi + t)^{-1}(\Phi(X)^{1/2}) = (\Delta + t)^{-1}(X^{1/2})$.

- That is, the resolvents of Δ_Φ and Δ coincide on the vectors $\Phi(X)^{1/2}$ and $X^{1/2}$, respectively (modulo V).

Proof of the main theorem

- The resolvent of an operator O determines the projections onto the eigenspaces of O.
- Hence, for every polynomial p we have

$$Vp(\Delta_\Phi)(\Phi(X)^{1/2}) = p(\Delta)(X^{1/2}).$$

- **Stone-Weierstrass approximation theorem:** polynomials are dense in the space of continuous functions
- Hence, for every continuous f we have

$$Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2}).$$

Proof of the main theorem

- The resolvent of an operator O determines the projections onto the eigenspaces of O.
- Hence, for every polynomial p we have

$$Vp(\Delta_\Phi)(\Phi(X)^{1/2}) = p(\Delta)(X^{1/2}).$$

- **Stone-Weierstrass approximation theorem:** polynomials are dense in the space of continuous functions
- Hence, for every continuous f we have

$$Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2}).$$

Proof of the main theorem

- The resolvent of an operator O determines the projections onto the eigenspaces of O.
- Hence, for every polynomial p we have

$$Vp(\Delta_\Phi)(\Phi(X)^{1/2}) = p(\Delta)(X^{1/2}).$$

- **Stone-Weierstrass approximation theorem:** polynomials are dense in the space of continuous functions
- Hence, for every continuous f we have

$$Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2}).$$

Proof of the main theorem

- The resolvent of an operator O determines the projections onto the eigenspaces of O.
- Hence, for every polynomial p we have
$$Vp(\Delta_\Phi)(\Phi(X)^{1/2}) = p(\Delta)(X^{1/2}).$$
- **Stone-Weierstrass approximation theorem:** polynomials are dense in the space of continuous functions
- Hence, for every continuous f we have
$$Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2}).$$

Proof of the main theorem

- $Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2})$ for all continuous f.
- Choose $f(x) = x^{it}$, then

$$
\begin{aligned}
V\Delta_\Phi^{it}(\Phi(X)^{1/2}) &= \Delta^{it}(X^{1/2}) \\
\Leftrightarrow V(\Phi(Y)^{it}\Phi(X)^{-it}\Phi(X)^{1/2}) &= Y^{it}X^{-it}X^{1/2} \\
\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it})X^{1/2} &= Y^{it}X^{-it}X^{1/2} \\
\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it}) &= Y^{it}X^{-it} \\
\Leftrightarrow \Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) &= X^{it}Y^{-it}
\end{aligned}
$$

The last line follows from taking the adjoint and using the fact that $\Lambda(A)^\dagger = \Lambda(A^\dagger)$ for a positive map Λ.

Proof of the main theorem

- $Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2})$ for all continuous f.
- Choose $f(x) = x^{it}$, then

$$V\Delta_\Phi^{it}(\Phi(X)^{1/2}) = \Delta^{it}(X^{1/2})$$

$$\Leftrightarrow V(\Phi(Y)^{it}\Phi(X)^{-it}\Phi(X)^{1/2}) = Y^{it}X^{-it}X^{1/2}$$

$$\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it})X^{1/2} = Y^{it}X^{-it}X^{1/2}$$

$$\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it}) = Y^{it}X^{-it}$$

$$\Leftrightarrow \Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) = X^{it}Y^{-it}$$

The last line follows from taking the adjoint and using the fact that $\Lambda(A)^\dagger = \Lambda(A^\dagger)$ for a positive map Λ.

Proof of the main theorem

- $Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2})$ for all continuous f.
- Choose $f(x) = x^{it}$, then

$$V\Delta_\Phi^{it}(\Phi(X)^{1/2}) = \Delta^{it}(X^{1/2})$$

$$\Leftrightarrow V(\Phi(Y)^{it}\Phi(X)^{-it}\Phi(X)^{1/2}) = Y^{it}X^{-it}X^{1/2}$$

$$\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it})X^{1/2} = Y^{it}X^{-it}X^{1/2}$$

$$\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it}) = Y^{it}X^{-it}$$

$$\Leftrightarrow \Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) = X^{it}Y^{-it}$$

The last line follows from taking the adjoint and using the fact that $\Lambda(A)^\dagger = \Lambda(A^\dagger)$ for a positive map Λ.

Proof of the main theorem

- $Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2})$ for all continuous f.
- Choose $f(x) = x^{it}$, then

$$
\begin{aligned}
V\Delta_\Phi^{it}(\Phi(X)^{1/2}) &= \Delta^{it}(X^{1/2}) \\
\Leftrightarrow V(\Phi(Y)^{it}\Phi(X)^{-it}\Phi(X)^{1/2}) &= Y^{it}X^{-it}X^{1/2} \\
\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it})X^{1/2} &= Y^{it}X^{-it}X^{1/2} \\
\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it}) &= Y^{it}X^{-it} \\
\Leftrightarrow \Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) &= X^{it}Y^{-it}
\end{aligned}
$$

The last line follows from taking the adjoint and using the fact that $\Lambda(A)^\dagger = \Lambda(A^\dagger)$ for a positive map Λ.

Proof of the main theorem

- $Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2})$ for all continuous f.
- Choose $f(x) = x^{it}$, then

$$
\begin{aligned}
V\Delta_\Phi^{it}(\Phi(X)^{1/2}) &= \Delta^{it}(X^{1/2}) \\
\Leftrightarrow V(\Phi(Y)^{it}\Phi(X)^{-it}\Phi(X)^{1/2}) &= Y^{it}X^{-it}X^{1/2} \\
\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it})X^{1/2} &= Y^{it}X^{-it}X^{1/2} \\
\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it}) &= Y^{it}X^{-it} \\
\Leftrightarrow \Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) &= X^{it}Y^{-it}
\end{aligned}
$$

The last line follows from taking the adjoint and using the fact that $\Lambda(A)^\dagger = \Lambda(A^\dagger)$ for a positive map Λ.

Proof of the main theorem

- $Vf(\Delta_\Phi)(\Phi(X)^{1/2}) = f(\Delta)(X^{1/2})$ for all continuous f.
- Choose $f(x) = x^{it}$, then

$$
\begin{aligned}
V\Delta_\Phi^{it}(\Phi(X)^{1/2}) &= \Delta^{it}(X^{1/2}) \\
\Leftrightarrow V(\Phi(Y)^{it}\Phi(X)^{-it}\Phi(X)^{1/2}) &= Y^{it}X^{-it}X^{1/2} \\
\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it})X^{1/2} &= Y^{it}X^{-it}X^{1/2} \\
\Leftrightarrow \Phi^\dagger(\Phi(Y)^{it}\Phi(X)^{-it}) &= Y^{it}X^{-it} \\
\Leftrightarrow \Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) &= X^{it}Y^{-it}
\end{aligned}
$$

The last line follows from taking the adjoint and using the fact that $\Lambda(A)^\dagger = \Lambda(A^\dagger)$ for a positive map Λ.

Proof of the main theorem

- Proven so far:

$$D(X\|Y) = D(\Phi(X)\|\Phi(Y))$$
$$\Longrightarrow \Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) = X^{it}Y^{-it} \quad \text{for all } t > 0.$$

- To prove sufficiency of $\Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) = X^{it}Y^{-it}$, differentiate this at $t = 0$:

$$\Phi^\dagger(\log\Phi(X) - \log\Phi(Y)) = \log X - \log Y$$

- Using the definition of $\Phi^\dagger$, this implies equality in the data processing inequality.

Proof of the main theorem

- Proven so far:

$$D(X\|Y) = D(\Phi(X)\|\Phi(Y))$$
$$\implies \Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) = X^{it}Y^{-it} \quad \text{for all } t > 0.$$

- To prove sufficiency of $\Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) = X^{it}Y^{-it}$, differentiate this at $t = 0$:

$$\Phi^\dagger(\log\Phi(X) - \log\Phi(Y)) = \log X - \log Y$$

- Using the definition of $\Phi^\dagger$, this implies equality in the data processing inequality.

Proof of the main theorem

- Proven so far:

$$D(X\|Y) = D(\Phi(X)\|\Phi(Y))$$
$$\implies \Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) = X^{it}Y^{-it} \quad \text{for all } t > 0.$$

- To prove sufficiency of $\Phi^\dagger(\Phi(X)^{it}\Phi(Y)^{-it}) = X^{it}Y^{-it}$, differentiate this at $t = 0$:

$$\Phi^\dagger(\log\Phi(X) - \log\Phi(Y)) = \log X - \log Y$$

- Using the definition of $\Phi^\dagger$, this implies equality in the data processing inequality.

Table of Contents

von Neumann entropy

For a state ρ define the **von Neumann entropy**

$$S(\rho) = -\operatorname{Tr}\rho\log\rho = -D(\rho\|\mathbb{1}).$$

- $S(\rho) = H(\{\lambda_i\}_i)$ where λ_i are the eigenvalues of ρ and $H(\{p_i\}_i) = -\sum_i p_i \log p_i$ is the **Shannon entropy**.
- $0 \leq S(\rho) \leq \log\dim\mathcal{H}$ for all states ρ on $\mathcal{H}$.
- **Additivity:** $S(\rho\otimes\sigma) = S(\rho) + S(\sigma)$
- **Subadditivity:** Let ρ_{AB} be a state on a bipartite system $\mathcal{H}_A\otimes\mathcal{H}_B$ and set $\rho_A = \operatorname{Tr}_B\rho_{AB}$ and $\rho_B = \operatorname{Tr}_A\rho_{AB}$. Then,

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B).$$

von Neumann entropy

For a state ρ define the **von Neumann entropy**

$$S(\rho) = -\operatorname{Tr}\rho\log\rho = -D(\rho\|\mathbb{1}).$$

- $S(\rho) = H(\{\lambda_i\}_i)$ where λ_i are the eigenvalues of ρ and $H(\{p_i\}_i) = -\sum_i p_i \log p_i$ is the **Shannon entropy**.
- $0 \leq S(\rho) \leq \log\dim\mathcal{H}$ for all states ρ on $\mathcal{H}$.
- **Additivity:** $S(\rho\otimes\sigma) = S(\rho) + S(\sigma)$
- **Subadditivity:** Let ρ_{AB} be a state on a bipartite system $\mathcal{H}_A\otimes\mathcal{H}_B$ and set $\rho_A = \operatorname{Tr}_B\rho_{AB}$ and $\rho_B = \operatorname{Tr}_A\rho_{AB}$. Then,

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B).$$

von Neumann entropy

For a state ρ define the **von Neumann entropy**

$$S(\rho) = -\operatorname{Tr}\rho\log\rho = -D(\rho\|\mathbb{1}).$$

- $S(\rho) = H(\{\lambda_i\}_i)$ where λ_i are the eigenvalues of ρ and $H(\{p_i\}_i) = -\sum_i p_i \log p_i$ is the **Shannon entropy**.
- $0 \leq S(\rho) \leq \log\dim\mathcal{H}$ for all states ρ on $\mathcal{H}$.
- **Additivity:** $S(\rho\otimes\sigma) = S(\rho) + S(\sigma)$
- **Subadditivity:** Let ρ_{AB} be a state on a bipartite system $\mathcal{H}_A\otimes\mathcal{H}_B$ and set $\rho_A = \operatorname{Tr}_B\rho_{AB}$ and $\rho_B = \operatorname{Tr}_A\rho_{AB}$. Then,

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B).$$

von Neumann entropy

For a state ρ define the **von Neumann entropy**

$$S(\rho) = -\operatorname{Tr} \rho \log \rho = -D(\rho \| \mathbb{1}).$$

- $S(\rho) = H(\{\lambda_i\}_i)$ where λ_i are the eigenvalues of ρ and $H(\{p_i\}_i) = -\sum_i p_i \log p_i$ is the **Shannon entropy**.
- $0 \leq S(\rho) \leq \log \dim \mathcal{H}$ for all states ρ on $\mathcal{H}$.
- **Additivity:** $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$
- **Subadditivity:** Let ρ_{AB} be a state on a bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$ and set $\rho_A = \operatorname{Tr}_B \rho_{AB}$ and $\rho_B = \operatorname{Tr}_A \rho_{AB}$. Then,

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B).$$

Strong subadditivity of von Neumann entropy

Strong subadditivity [Lieb and Ruskai 1973]:

- Let ρ_{ABC} be a tripartite state, and denote by $\rho_{AB}, \rho_{BC}, \rho_B$ the corresponding marginals. Then:

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}).$$

- **Quantum conditional mutual information**

$$I(A;C|B)_\rho = S(AB) + S(BC) - S(B) - S(ABC)$$

where $S(AB) \equiv S(\rho_{AB})$ etc.

- Definition of qCMI analogous to classical quantity!
- SSA $\Longrightarrow I(A;C|B)_\rho \geq 0$ for all ρ_{ABC}.

Strong subadditivity of von Neumann entropy

Strong subadditivity [Lieb and Ruskai 1973]:

- Let ρ_{ABC} be a tripartite state, and denote by $\rho_{AB}, \rho_{BC}, \rho_B$ the corresponding marginals. Then:

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}).$$

- **Quantum conditional mutual information**

$$I(A;C|B)_\rho = S(AB) + S(BC) - S(B) - S(ABC)$$

where $S(AB) \equiv S(\rho_{AB})$ etc.

- Definition of qCMI analogous to classical quantity!
- SSA $\Longrightarrow I(A;C|B)_\rho \geq 0$ for all ρ_{ABC}.

Strong subadditivity of von Neumann entropy

Strong subadditivity [Lieb and Ruskai 1973]:

- Let ρ_{ABC} be a tripartite state, and denote by $\rho_{AB}, \rho_{BC}, \rho_B$ the corresponding marginals. Then:

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}).$$

- **Quantum conditional mutual information**

$$I(A;C|B)_\rho = S(AB) + S(BC) - S(B) - S(ABC)$$

where $S(AB) \equiv S(\rho_{AB})$ etc.

- Definition of qCMI analogous to classical quantity!
- SSA $\Longrightarrow I(A;C|B)_\rho \geq 0$ for all ρ_{ABC}.

Strong subadditivity of von Neumann entropy

Strong subadditivity [Lieb and Ruskai 1973]:

- Let ρ_{ABC} be a tripartite state, and denote by $\rho_{AB}, \rho_{BC}, \rho_B$ the corresponding marginals. Then:

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}).$$

- **Quantum conditional mutual information**

$$I(A; C|B)_\rho = S(AB) + S(BC) - S(B) - S(ABC)$$

where $S(AB) \equiv S(\rho_{AB})$ etc.

- Definition of qCMI analogous to classical quantity!
- SSA $\implies I(A; C|B)_\rho \geq 0$ for all ρ_{ABC}.

Strong subadditivity of von Neumann entropy

Strong subadditivity:

$$S(\rho_{ABC}) + S(\rho_B) \le S(\rho_{AB}) + S(\rho_{BC})$$

- Highly non-trivial statement!
- However, easy proof using data processing inequality:

$$\begin{aligned} S(\rho_A) + S(\rho_{BC}) - S(\rho_{ABC}) &= D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) \\ &\ge D(\rho_{AB} \| \rho_A \otimes \rho_B) \\ &= S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \end{aligned}$$

- What happens if we have equality in SSA?
- By the above: equality in SSA $\Leftrightarrow I(A;C|B) = 0$

Strong subadditivity of von Neumann entropy

Strong subadditivity:

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})$$

- Highly non-trivial statement!
- However, easy proof using data processing inequality:

$$\begin{aligned} S(\rho_A) + S(\rho_{BC}) - S(\rho_{ABC}) &= D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) \\ &\geq D(\rho_{AB} \| \rho_A \otimes \rho_B) \\ &= S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \end{aligned}$$

- What happens if we have equality in SSA?
- By the above: equality in SSA $\Leftrightarrow I(A;C|B) = 0$

Strong subadditivity of von Neumann entropy

Strong subadditivity:

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})$$

- Highly non-trivial statement!
- However, easy proof using data processing inequality:

$$\begin{aligned} S(\rho_A) + S(\rho_{BC}) - S(\rho_{ABC}) &= D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) \\ &\geq D(\rho_{AB} \| \rho_A \otimes \rho_B) \\ &= S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \end{aligned}$$

- What happens if we have equality in SSA?
- By the above: equality in SSA $\Leftrightarrow I(A;C|B) = 0$

Strong subadditivity of von Neumann entropy

Strong subadditivity:

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})$$

- Highly non-trivial statement!
- However, easy proof using data processing inequality:

$$\begin{aligned} S(\rho_A) + S(\rho_{BC}) - S(\rho_{ABC}) &= D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) \\ &\geq D(\rho_{AB} \| \rho_A \otimes \rho_B) \\ &= S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \end{aligned}$$

- What happens if we have equality in SSA?
- By the above: equality in SSA $\Leftrightarrow I(A;C|B) = 0$

Strong subadditivity of von Neumann entropy

Equality in SSA [Hayden et al. 2004]:

- We applied DPI with respect to partial trace over C.
- Equality condition (recovery map formulation):

  There is a recovery map $\mathcal{R}_{B\to BC} : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_{BC})$ s.t.

  $$\mathcal{R}_{B\to BC}(\rho_{AB}) = \rho_{ABC} \qquad \mathcal{R}_{B\to BC}(\sigma_{AB}) = \sigma_{ABC}$$

- Hence, we obtain:

  $$I(A;C|B)_\rho = 0 \quad \Longleftrightarrow \quad \exists \mathcal{R}_{B\to BC} \text{ with } \rho_{BC} = \mathcal{R}_{B\to BC}(\rho_B)$$

- That is, $A \leftrightarrow B \leftrightarrow C$ forms a **(short) quantum Markov chain** iff we can recover C from B alone.

Strong subadditivity of von Neumann entropy

Equality in SSA [Hayden et al. 2004]:

- We applied DPI with respect to partial trace over C.
- Equality condition (recovery map formulation):

  There is a recovery map $\mathcal{R}_{B\to BC}\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_{BC})$ s.t.

$$\mathcal{R}_{B\to BC}(\rho_{AB}) = \rho_{ABC} \qquad \mathcal{R}_{B\to BC}(\sigma_{AB}) = \sigma_{ABC}$$

- Hence, we obtain:

$$I(A;C|B)_\rho = 0 \quad \iff \quad \exists \mathcal{R}_{B\to BC} \text{ with } \rho_{BC} = \mathcal{R}_{B\to BC}(\rho_B)$$

- That is, $A \leftrightarrow B \leftrightarrow C$ forms a **(short) quantum Markov chain** iff we can recover C from B alone.

Strong subadditivity of von Neumann entropy

Equality in SSA [Hayden et al. 2004]:

- We applied DPI with respect to partial trace over C.
- Equality condition (recovery map formulation):

  There is a recovery map $\mathcal{R}_{B\to BC}\colon \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_{BC})$ s.t.

$$\mathcal{R}_{B\to BC}(\rho_{AB}) = \rho_{ABC} \qquad \mathcal{R}_{B\to BC}(\sigma_{AB}) = \sigma_{ABC}$$

- Hence, we obtain:

$$I(A;C|B)_\rho = 0 \quad \Longleftrightarrow \quad \exists \mathcal{R}_{B\to BC} \text{ with } \rho_{BC} = \mathcal{R}_{B\to BC}(\rho_B)$$

- That is, $A \leftrightarrow B \leftrightarrow C$ forms a **(short) quantum Markov chain** iff we can recover C from B alone.

Strong subadditivity of von Neumann entropy

Equality in SSA [Hayden et al. 2004]:

- We applied DPI with respect to partial trace over C.
- Equality condition (recovery map formulation):

  There is a recovery map $\mathcal{R}_{B\to BC} : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_{BC})$ s.t.

  $$\mathcal{R}_{B\to BC}(\rho_{AB}) = \rho_{ABC} \qquad \mathcal{R}_{B\to BC}(\sigma_{AB}) = \sigma_{ABC}$$

- Hence, we obtain:

  $$I(A;C|B)_\rho = 0 \quad \iff \quad \exists \mathcal{R}_{B\to BC} \text{ with } \rho_{BC} = \mathcal{R}_{B\to BC}(\rho_B)$$

- That is, $A \leftrightarrow B \leftrightarrow C$ forms a **(short) quantum Markov chain** iff we can recover C from B alone.

Table of Contents

Conclusion and open problems

Summary:

- Divergences (or relative entropies) play an important role in Classical and Quantum Information Theory.
- Their crucial property is the **data processing inequality**.
- The **quantum relative entropy** is an important example in Quantum Information Theory.
- We derived an **equality condition** in the DPI for the quantum relative entropy.
- We saw how this gives rise to the notion of **quantum Markov chains**.

Conclusion and open problems

Summary:

- Divergences (or relative entropies) play an important role in Classical and Quantum Information Theory.
- Their crucial property is the **data processing inequality**.
- The **quantum relative entropy** is an important example in Quantum Information Theory.
- We derived an **equality condition** in the DPI for the quantum relative entropy.
- We saw how this gives rise to the notion of **quantum Markov chains**.

Conclusion and open problems

Summary:

- Divergences (or relative entropies) play an important role in Classical and Quantum Information Theory.
- Their crucial property is the **data processing inequality**.
- The **quantum relative entropy** is an important example in Quantum Information Theory.
- We derived an **equality condition** in the DPI for the quantum relative entropy.
- We saw how this gives rise to the notion of **quantum Markov chains**.

Conclusion and open problems

Summary:

- Divergences (or relative entropies) play an important role in Classical and Quantum Information Theory.
- Their crucial property is the **data processing inequality**.
- The **quantum relative entropy** is an important example in Quantum Information Theory.
- We derived an **equality condition** in the DPI for the quantum relative entropy.
- We saw how this gives rise to the notion of **quantum Markov chains**.

Conclusion and open problems

Summary:

- Divergences (or relative entropies) play an important role in Classical and Quantum Information Theory.
- Their crucial property is the **data processing inequality**.
- The **quantum relative entropy** is an important example in Quantum Information Theory.
- We derived an **equality condition** in the DPI for the quantum relative entropy.
- We saw how this gives rise to the notion of **quantum Markov chains**.

Conclusion and open problems

Generalized divergence:

- Crucial property of a divergence: DPI
- Popular other choices: **Rényi divergences**

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}(\rho^\alpha \sigma^{1-\alpha}) \qquad \alpha \in [0, 2]$$

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}\left(\sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha}\right)^\alpha \qquad \alpha \in [1/2, \infty)$$

- $\lim_{\alpha\to 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma) = \lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma)$.
- Both satisfy DPI in the given range.

[Petz 1986; Frank and Lieb 2013; Beigi 2013]

Conclusion and open problems

Generalized divergence:

- Crucial property of a divergence: DPI
- Popular other choices: **Rényi divergences**

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}(\rho^\alpha \sigma^{1-\alpha}) \qquad \alpha \in [0, 2]$$

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}\left(\sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha}\right)^\alpha \qquad \alpha \in [1/2, \infty)$$

- $\lim_{\alpha \to 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma) = \lim_{\alpha \to 1} \widetilde{D}_\alpha(\rho\|\sigma)$.
- Both satisfy DPI in the given range.

[Petz 1986; Frank and Lieb 2013; Beigi 2013]

Conclusion and open problems

Generalized divergence:

- Crucial property of a divergence: DPI
- Popular other choices: **Rényi divergences**

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log\mathrm{Tr}(\rho^\alpha\sigma^{1-\alpha}) \qquad \alpha\in[0,2]$$

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log\mathrm{Tr}\left(\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha}\right)^\alpha \qquad \alpha\in[1/2,\infty)$$

- $\lim_{\alpha\to1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma) = \lim_{\alpha\to1}\widetilde{D}_\alpha(\rho\|\sigma)$.
- Both satisfy DPI in the given range.

[Petz 1986; Frank and Lieb 2013; Beigi 2013]

Conclusion and open problems

Generalized divergence:

- Crucial property of a divergence: DPI

- Popular other choices: **Rényi divergences**

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}(\rho^\alpha \sigma^{1-\alpha}) \qquad \alpha \in [0, 2]$$

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}\left(\sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha}\right)^\alpha \quad \alpha \in [1/2, \infty)$$

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

- Both satisfy DPI in the given range.

  [Petz 1986; Frank and Lieb 2013; Beigi 2013]

Conclusion and open problems

- α**-Rényi divergence**:

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1}\log \operatorname{Tr}(\rho^\alpha \sigma^{1-\alpha})$$

- Equality conditions: Same as for $D(\rho\|\sigma)$! [Hiai et al. 2011]

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it} \quad \text{for all } t \in \mathbb{R}.$$

- α**-sandwiched Rényi divergence**:

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

[Müller-Lennert et al. 2013; Wilde et al. 2014]

- Equality conditions? [FL, Rouzé, Datta]

$\longrightarrow$ *CQIF seminar, Fri, February 26, 12.00 (MR12)!*

Conclusion and open problems

- **α-Rényi divergence**:

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}(\rho^\alpha \sigma^{1-\alpha})$$

- Equality conditions: Same as for $D(\rho\|\sigma)$! [Hiai et al. 2011]

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it} \quad \text{for all } t \in \mathbb{R}.$$

- **α-sandwiched Rényi divergence**:

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}\left(\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha}\right)^\alpha$$

[Müller-Lennert et al. 2013; Wilde et al. 2014]

- Equality conditions? [FL, Rouzé, Datta]

$\longrightarrow$ *CQIF seminar, Fri, February 26, 12.00 (MR12)!*

Conclusion and open problems

- α**-Rényi divergence**:

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}(\rho^\alpha \sigma^{1-\alpha})$$

- Equality conditions: Same as for $D(\rho\|\sigma)$! [Hiai et al. 2011]

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it} \quad \text{for all } t \in \mathbb{R}.$$

- α**-sandwiched Rényi divergence**:

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}\left(\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha}\right)^\alpha$$

[Müller-Lennert et al. 2013; Wilde et al. 2014]

- Equality conditions? [FL, Rouzé, Datta]

$\longrightarrow$ *CQIF seminar, Fri, February 26, 12.00 (MR12)!*

Conclusion and open problems

- **α-Rényi divergence**:

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \operatorname{Tr}(\rho^\alpha \sigma^{1-\alpha})$$

- Equality conditions: Same as for $D(\rho\|\sigma)$! [Hiai et al. 2011]

$$\Phi^\dagger\left(\Phi(X)^{it}\Phi(Y)^{-it}\right) = X^{it}Y^{-it} \quad \text{for all } t \in \mathbb{R}.$$

- **α-sandwiched Rényi divergence**:

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

[Müller-Lennert et al. 2013; Wilde et al. 2014]

- Equality conditions? [FL, Rouzé, Datta]

$\longrightarrow$ *CQIF seminar, Fri, February 26, 12.00 (MR12)!*

References

Beigi, S. (2013). *Journal of Mathematical Physics* 54.12, p. 122202. arXiv: 1306.5920 [quant-ph].

Frank, R. L. and E. H. Lieb (2013). *Journal of Mathematical Physics* 54.12, p. 122201. arXiv: 1306.5358 [math-ph].

Hayden, P. et al. (2004). *Communications in Mathematical Physics* 246.2, pp. 359–374. arXiv: quant-ph/0304007.

Hiai, F. et al. (2011). *Reviews in Mathematical Physics* 23.07, pp. 691–747.

Lieb, E. H. and M. B. Ruskai (1973). *Physical Review Letters* 30 (10), pp. 434–436.

Müller-Lennert, M. et al. (2013). *Journal of Mathematical Physics* 54.12, p. 122203. arXiv: 1306.3142 [quant-ph].

Müller-Hermes, A. and D. Reeb (2015). *arXiv preprint*. arXiv: 1512.06117 [quant-ph].

Petz, D. (1986). *Reports on Mathematical Physics* 23.1, pp. 57–65.

— (1988). *The Quarterly Journal of Mathematics* 39.1, pp. 97–108.

Stinespring, W. F. (1955). *Proceedings of the American Mathematical Society* 6.2, pp. 211–216.

Wilde, M. M. et al. (2014). *Communications in Mathematical Physics* 331.2, pp. 593–622. arXiv: 1306.1586 [quant-ph].

Thank you very much for your attention!