On Geometry of Quantum State Space

József Pitrik
HUN-REN Wigner Research Centre for Physics, Budapest
HUN-REN Alfréd Rényi Institute of Mathematics, Budapest
Department of Analysis and Operations Research, BUTE

Common work with Géza Tóth (UPV/EHU Bilbao)

1. Dissimilarity measures for probabilities
   - Measures of Dissimilarity
   - f-Divergences
   - The Riemannian structure of statistical manifolds
2. Quantum Counterparts
   - The quantum f-divergences
   - Monotone metrics
   - Monotone metrics as induced metric by f-divergences
   - Metric adjusted skew informations
3. Our contribution

Measures of Dissimilarity

Denote $\mathcal{P}(X)$ the space of probability measures on $X = \mathbb{R}^n$. A function D is called a **divergence** if

$$D : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R}, \quad D(p, q) \geq 0 \text{ with } D(p, q) = 0 \text{ iff } p = q.$$

A divergence D is

1. *symmetric* if $D(p, q) = D(q, p)$ for all $p, q \in \mathcal{P}(X)$.
2. fulfills the *triangle inequality* if $D(p, q) + D(q, r) \geq D(p, r)$ for all $p, q, r \in \mathcal{P}(X)$.
3. *monotone* or fulfills the *Data Processing Inequalities (DPI)* if

$$D(S(p), S(q)) \leq D(p, q),$$

for any *stochastic map (Markov Kernel)* $S : \mathcal{P}(X) \to \mathcal{P}(X)$.

Remarks:

- If we consider discrete probability distributions i.e. $p = (p_1, p_2, \ldots, p_n)^T$, with $(\forall i : p_i \geq 0)$, $\sum_{i=1}^{n} p_i = 1$, then any stochastic map S is given by a left-stochastic *transition matrix* (S_{ij}), i.e. $(\forall i, j : S_{ij} \geq 0)$ and $(\forall j : \sum_i S_{ij} = 1)$.
- We can interpret columns of a transition matrix as vectors of a conditional probability: $S_{.j} = p(.|j)$, which effectly randomizes the input probability vectors.
- If for a divergence D the conditions (1) and (2) hold then it is called a **distance**.

For simplicity, we primarily consider discrete probability distributions.

f-Divergences (Imre Csiszár, 1967)

For any given convex function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that $f(1) = 0$, an f-**divergence** is defined as

$$D_f(p, q) = \sum_i p_i f\left(\frac{q_i}{p_i}\right).$$

Some properties

- For $f^*(x) = xf(x^{-1})$ we have $D_f(p, q) = D_{f^*}(q, p)$ so D_f is symmetric if $f^* = f$.
- The convexity of f ensures that D_f is monotone (fulfills the DPI).
- Of course, not every monotone divergence constitutes an f-divergence: for any non-decreasing $g : \mathbb{R}^+ \to \mathbb{R}^+$ with $g(0) = 0$, $g(D_f(.,.))$ gives a monotone divergence, which is not an f-divergence.

Csiszár's characterisation

Theorem Assume that a number $C(p, q) \in \mathbb{R}$ is associated to probability distributions on the same set X for all finite set X. If

1. $C(p, q)$ is invariant under the permutations of the basic set X, and
2. if $\mathcal{A}$ is a partition of X and $p_{\mathcal{A}}(A) := \sum_{x \in A} p(x)$, then $C(p_{\mathcal{A}}, q_{\mathcal{A}}) \leq C(p, q)$ with equality iff $p_{\mathcal{A}}(A)q(x) = q_{\mathcal{A}}(A)p(x)$ whenever $x \in A \in \mathcal{A}$,

then there exists a convex function $f : \mathbb{R}^+ \to \mathbb{R}$ which is continous at 0 and $C(p, q) = D_f(p, q)$ for every p, q.

Some examples for f-divergences

- For $f(x) = \frac{1}{2}|1-x|$ the f-divergence is

$$T(p,q) = \frac{1}{2}\sum_i |p_i - q_i|$$ **the Total Variation (TV) distance**.

- For $f(x) = \frac{1-x^\alpha}{1-\alpha}$, $0 \leq \alpha < 1$ the f-divergences are

$$H_\alpha(p,q) = \frac{1}{1-\alpha}\left(1 - \sum_i p_i^\alpha q_i^{1-\alpha}\right) = \frac{1}{1-\alpha}(1 - \xi_\alpha(p,q))$$

the **Hellinger divergences** parametrized by α. Here $\xi_\alpha(p,q) = \sum_i p_i^\alpha q_i^{1-\alpha}$ are called **Chernoff coefficients**.

The $\alpha = 1/2$ case is crucial:

$$H_{1/2}(p,q) = 2(1 - \sum_i \sqrt{p_i q_i}) = 2(1 - F(p,q))$$

squared Hellinger distance, where $F(p,q) = \xi_{1/2}(p,q) = \sum_i \sqrt{p_i q_i}$ is the **Bhattacharyya coefficient**.

- For $f(x) = -\log x$ we get for the f-divergence

$$KL(p,q) = \sum_i p_i \log \frac{p_i}{q_i}$$

the **Kullback-Leibler divergence or relative entropy**.

Example: Hypothesis Testing Tasks

Objective: correctly identify, based on the outcomes $\mathrm{x}^\mathrm{n} = (x_1, \ldots, x_n)$ of n independent rounds, which of given PDs p and q is the one governing the experiment (the r. v. $X \sim p$ or q)
Decision function: $\mathcal{D} : X^n \to \{0, 1\}$, s.t
if $\mathcal{D}(\mathrm{x_n}) = 0 \Rightarrow$ one concludes p to be correct PD
if $\mathcal{D}(\mathrm{x_n}) = 1 \Rightarrow$ one concludes q to be correct PD
Hypothesis Testing:
H_0: "p is the true PD"
H_1: "q is the true PD"
Types of errors:
type-I errors ("false positive"): H_0 is rejected based on the data despite actually holding true $\Rightarrow P_n(q|p)$
type-II errors ("false negative"): H_0 is maintained although the data has actually been generated in accordance with $H_1 \Rightarrow P_n(p|q)$

I. Symmetric Hypothesis Testing

The goal is minimising the *average error probability*

$$p_n^{err} := \pi_p P_n(q|p) + \pi_q P_n(p|q),$$

where π_p and π_q are the *a priori* probabilities of p and q resp. $(\pi_p + \pi_q = 1)$

1. **Single-shot scenario** (n = 1)

$$p_{min}^{err} = \frac{1}{2}(1 - T(p, q))$$

2. **Asymptotic scenario** $(n \to \infty)$

$$p_{n,min}^{err} \leq \xi(p, q)^n,$$

where $\xi(p, q) := \min_{0 \leq \alpha \leq 1} \xi_\alpha(p, q)$ is the **Chernoff bound**.

$$\lim_{n \to \infty} p_{n,min}^{err} = \exp[n \ln \xi(p, q)]$$

II. Asymmetric Hypothesis Testing
Finding optimal inference strategy for which $P_n(p|q)$ (and the type-II error) is minimal, while simultaneously assuring that $P_n(q|p) \leq \epsilon$ for some $0 < \epsilon < 1$. For the probability of II-type error we get

$$\lim_{n\to\infty} P_{n,min}(p|q) = \exp[-nKL(p,q) + o(n)],$$

according to the *Stein's lemma*.

The statistical manifold

Consider a familiy $\mathcal{M}$ of probability distributions on X and suppose each element of $\mathcal{M}$, a PD, may be parametrized using n real-valued variables: $\theta = (\theta_1, \dots, \theta_n)$, i.e.

$$\mathcal{M} = \{p_\theta = p(x;\theta) : \theta = (\theta_1, \dots, \theta_n) \in \Theta\},$$

where $\Theta \subset \mathbb{R}^n$ is open and the mapping $\theta \mapsto p_\theta$ is injective. S is called an n-dimensional **statistical (parametrical) model or manifold**. θ_i's are called **coordinates** and **the tangent space** associated with a given point $p \in \mathcal{M}$ is denoted by $\mathcal{T}_p\mathcal{M}$.

Examples

1. **Normal distributions**

$$p(x;\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-m)^2}{2\sigma^2}\}$$

$X = \mathbb{R}$, $n = 2$, $\theta = (m, \sigma)$,
$\Theta = \{(m, \sigma) : -\infty < m < \infty, 0 < \sigma < \infty\}$

2. **$\mathcal{P}(X)$ for finite X**

$$p(x_i;\theta) = \begin{cases} \theta_i & \text{if } 1 \leq i \leq n \\ 1 - \sum_{i=1}^{n} \theta_i & \text{if } i = 0 \end{cases}$$

$X = (x_0, x_1, \ldots, x_n)$, $\Theta = \{(\theta_1, \ldots, \theta_n) : (\forall i : \theta_i > 0), \sum_{i=1}^{n} \theta_i < 1\}$

A **Riemannian metric** on a statistical manifold $\mathcal{M}$ is defined as a smooth mapping

$$\mathfrak{g}_p : \mathcal{T}_p\mathcal{M} \times \mathcal{T}_p\mathcal{M} \to \mathbb{R}^+$$

where $p \in \mathcal{M}$ is a PD and it defines an inner products $\langle ., . \rangle_{\mathfrak{g}_p} \geq 0$ for vectors contained in the *tangent space* $\mathcal{T}_p\mathcal{M}$. If $(e_i)_i$ is an ON basis in $\mathcal{T}_p\mathcal{M}$ then with the matrix $g_{ij}(p) = \langle e_i, e_j \rangle_{\mathfrak{g}_p}$, for all $u, v \in \mathcal{T}_p\mathcal{M}$ we have

$$\mathfrak{g}_p(u, v) \equiv \langle u, v \rangle_{\mathfrak{g}_p} = \sum_{i,j} g_{ij}(p) u_i v_j = u^T g(p) v.$$

For an arbitrary curve $\gamma^{(p,q)} : [a, b] \ni t \mapsto u \in \mathcal{M}$ connectig points $p = \gamma^{(p,q)}(a)$ and $q = \gamma^{(p,q)}(b)$ in $\mathcal{M}$, the tangent vectors along the curve read

$$\dot{\gamma}(t) := \frac{d\gamma^{(p,q)}}{dt}|_u = \sum_i \dot{\gamma}_i e_i.$$

The lenght of the curve is defined as

$$|\gamma^{(p,q)}| = \int_a^b \sqrt{\sum_{i,j} g_{ij}\dot{\gamma}_i\dot{\gamma}_j}\, dt,$$

and the squared infinitesimal segment of length along the curve is

$$d\ell^2 = \sum_{i,j} g_{ij}\dot{\gamma}_i\dot{\gamma}_j dt = \langle\dot{\gamma}(t), \dot{\gamma}(t)\rangle_{\mathfrak{g}} dt.$$

Divergence-Induced metrics

Consider a divergence $D(p, q)$ which is smooth in its both arguments $p, q \in \mathcal{M}$. For some orthonormal basis $(e_i)_i$ in $\mathcal{T}_p\mathcal{M}$ we define a matrix

$$g_D(p)_{ij} := \langle e_i, e_j \rangle_{\mathfrak{g}_D(p)} = -\frac{\partial^2}{\partial t \partial s} D(p + te_i, p + se_j)|_{t=s=0},$$

then D defines an inner product, and hence a metric $\mathfrak{g}_D(p)$ for any $u, v \in \mathcal{T}_p\mathcal{M}$ at $p \in \mathcal{M}$ as

$$\mathfrak{g}_D(p)(u, v) = \langle u, v \rangle_{\mathfrak{g}_D(p)} = v^T g_D(p) u.$$

We can give the Taylor expansion for D, when perturbing by $\delta p \in \mathcal{T}_p\mathcal{M}$ a given $p \in \mathcal{M}$ onto $p + \delta p \in \mathcal{M}$ s.t. $\mathrm{supp}\,(p+\delta p) \subseteq \mathrm{supp}\,(p)$ as

$$D(p, p+\delta p) = \frac{1}{2}\mathfrak{g}_D(p)(\delta p, \delta p) + \mathrm{O}(\delta p^3) = \frac{1}{2}\delta p^T g_D(p)\delta p + \mathrm{O}(\delta p^3),$$

since the less degree terms of δp absent as $D(p,p) = 0$ to be a global minimum.

A numerical function f defined on pairs of probability distributions is **monotone** if

$$f(S(p), S(q)) \leq f(p,q),$$

for all stochastic map S and for all PDs p, q.

Monotone divergence induce monotone metric

A metric is **monotone metric** if its geodesic distance is a monotone function. For a stochastic map S let $p' = S(p)$ and $\delta p' = S(\delta p)$. The monotonicity of the divergence D implies

$$D(p\ , p' + \delta p') \leq D(p, p + \delta p),$$

which is equivalent with

$$g_D(p) \geq S^T g_D(S(p)) S \quad \text{up to } \mathrm{O}(\delta p^3).$$

The Fisher Information Matrix

Motivation

We seek a quantitative measure of the extent that two distributions $p(x;\theta)$ and $p(x;\theta+d\theta)$ can be distinguished. Consider the relative difference

$$\Delta = \frac{p(x;\theta+d\theta)-p(x;\theta)}{p(x;\theta)} = \sum_i \frac{\frac{\partial p(x;\theta)}{\partial\theta_i}d\theta_i}{p(x;\theta)} = \sum_i \frac{\partial \log p(x;\theta)}{\partial\theta_i}d\theta_i$$

The expected value

$$\langle\Delta\rangle = \mathbb{E}_p\Delta = \int dxp(x;\theta)\sum_i \frac{\partial \log p(x;\theta)}{\partial\theta_i}d\theta_i =$$

$$\sum_i d\theta_i \frac{\partial}{\partial\theta_i}\int dxp(x;\theta) = 0.$$

We consider the variance

$$\langle \Delta^2 \rangle = \mathbb{E}_p \Delta^2 = \int dx p(x;\theta) \sum_{i,j} \frac{\partial \log p(x;\theta)}{\partial \theta_i} \frac{\partial \log p(x;\theta)}{\partial \theta_j} d\theta_i d\theta_j$$

as a measure of the infinitesimal difference between the two distributions, so we could say $d\ell^2 = \mathbb{E}_p \Delta^2$ between the points θ and $\theta + d\theta$ on the manifold. It suggests introducing the matrix (g_{ij})

$$g_{ij}(\theta) = \int dx p(x;\theta) \frac{\partial \log p(x;\theta)}{\partial \theta_i} \frac{\partial \log p(x;\theta)}{\partial \theta_j} =$$

$$\int dx \frac{1}{p(x;\theta)} \frac{\partial p(x;\theta)}{\partial \theta_i} \frac{\partial p(x;\theta)}{\partial \theta_j}$$

the **Fisher information matrix** and for the infinitesimal distance $d\ell$

$$d\ell^2 = \sum_{ij} g_{ij}(\theta) d\theta_i d\theta_j.$$

Rao was the first to realize in 1945 that a statistical manifold (a family of densities) can be considered as a Riemannian manifold where Fisher information matrix plays the role of metrics.

Theorem (Chentsov)

Up to a constant factor the Fisher information matrix yields the only monotone family of Riemann metrics on the class of finite probability simplexes.

The Fisher-Rao metric

We restrict ourself to the open probability simplex

$$\mathcal{P}_n = \{p = (p_1, \ldots, p_n) : p_i > 0, \sum_i p_i = 1\}.$$

The tangent space is given by

$$\mathcal{T}\mathcal{P}_n = \{u \in \mathbb{R}^n : \sum_i u_i = 0\}.$$

Then the **Fisher-Rao metric** is given by

$$\mathfrak{g}_p(u, v) = \langle u, v \rangle_{\mathfrak{g}(p)} = \sum_i \frac{u_i v_i}{p_i},$$

for all tangent vector $u, v \in \mathcal{T}\mathcal{P}_n$.

Fisher-Rao metric as induced metrics by f-divergences

For a sufficiently smooth convex function $f : \mathbb{R}^+ \to \mathbb{R}^+$, with $f(1) = 0$, for the monotone f-divergence $D_f(p, q) = \sum_i p_i f\left(\frac{p_i}{q_i}\right)$ the induced metric $\mathfrak{g}_{D_f}(p)$ at $p \in \mathcal{P}_n$ is given by

$$\mathfrak{g}_{D_f}(p)(u, v) = -\frac{\partial^2}{\partial t \partial s} D_f(p + tu, p + sv)|_{t=s=0} =$$

$$= f''(1) \sum_i \frac{u_i v_i}{p_i},$$

for all $u, v \in \mathcal{T}\mathcal{P}_n$, according to the Chentsov Theorem.

Notations

- $\mathcal{H}$ is a finite dimensional Hilbert space
- $\mathcal{B}(\mathcal{H})$ is the algebra of linear operators
- $\mathcal{B}(\mathcal{H})^+ = \{A \in \mathcal{B}(\mathcal{H}) : A \geq 0\}$
- $\mathcal{B}(\mathcal{H})^{++} = \{A \in \mathcal{B}(\mathcal{H}) : A > 0\}$
- $\mathcal{S}(\mathcal{H}) = \{\rho \in \mathcal{B}(\mathcal{H})^+, \mathrm{Tr}\,\rho = 1\}$ is the **quantum state space** and its element are called **density matrices**.
- The **Hilbert-Schmidt inner product** is

$$\langle A, B \rangle = \mathrm{Tr}\, A^* B, \quad A, B \in \mathcal{B}(\mathcal{H})$$

- The **left** and **right multiplications** of $A \in \mathcal{B}(\mathcal{H})$ are

$$L_A X = AX, \quad R_A X = XA, \quad X \in \mathcal{B}(\mathcal{H})$$

If $A, B \in \mathcal{B}(\mathcal{H})^+$, then L_A and L_B are positive operators on $\mathcal{B}(\mathcal{H})$ with $L_A R_B = R_B L_A$.

The quantum f-divergence of Petz

Definition

Assume that $f:(0,\infty)\to\mathbb{R}$ is an **operator convex function**. For $A,B\in\mathcal{B}(\mathcal{H})^{++}$ with spectral decomposition $A=\sum_{a\in\mathrm{Sp}(A)} aP_a$ and $B=\sum_{b\in\mathrm{Sp}(B)} bQ_b$, the **(standard)** f**-divergence** is

$$S_f(A,B):=\langle B^{1/2}, f(L_A R_{B^{-1}})B^{1/2}\rangle=\mathrm{Tr}\, B^{1/2} f(L_A R_{B^{-1}})(B^{1/2}),$$

with

$$f(L_A R_{B^{-1}})=\sum_{a\in\mathrm{Sp}(A)}\sum_{b\in\mathrm{Sp}(B)} f(ab^{-1})L_{P_a}R_{Q_b},$$

which can be extended to general $A,B\in\mathcal{B}(\mathcal{H})^{+}$ as

$$S_f(A,B):=\lim_{\varepsilon\searrow 0} S_f(A+\varepsilon I, B+\varepsilon I).$$

Properties of S_f

- **Joint convexity**

  For every $\rho_i, \sigma_i \in \mathcal{S}(\mathcal{H})$ and $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$ we have

$$S_f(\sum_i \lambda_i \rho_i, \sum_i \lambda_i \sigma_i) \leq \sum_i \lambda_i S_f(\rho_i, \sigma_i).$$

- **Monotonicity**

  Assume that $\Phi : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$ is a completely positive trace-preserving (CPTP) map. Then for any density matrices ρ, σ the DPI holds, i.e.

$$S_f(\Phi(\rho), \Phi(\sigma)) \leq S_f(\rho, \sigma).$$

The maximal f-divergence of Matsumoto

It is natural to ensure the generalization of a divergence to be independent of the measurement choice. This, in particular, can be achieved by performing maximization over all POVM available, i.e.

$$D(\rho, \sigma) := \max_{\{\Pi_i\}_i} D(p, q),$$

such that $p_i = \mathrm{Tr}\,(\rho \Pi_i)$, $q_i = \mathrm{Tr}\,(\sigma \Pi_i)$, where $\Pi_i \geq 0$ and $\sum_i \Pi_i = I$.

Definition

Assume that $f:(0,\infty)\to\mathbb{R}$ is an **operator convex function**. For $A,B\in\mathcal{B}(\mathcal{H})^{++}$ the **maximal f-divergence** is

$$S_f^{max}(A,B):=\mathrm{Tr}\,B^{1/2}f(B^{-1/2}AB^{-1/2})B^{1/2},$$

which can be extended to general $A,B\in\mathcal{B}(\mathcal{H})^{+}$ as

$$S_f^{max}(A,B):=\lim_{\varepsilon\searrow 0}S_f(A+\varepsilon I,B+\varepsilon I).$$

Properties of S_f^{max}

- S_f^{max} is jointly convex
- S_f^{max} is monotone under CPTP maps (DPI)
- S_f^{max} is maximal among monotone quantum f-divergences

$$S_f(A,B) \leq S_f^{max}(A,B), \quad A,B \in \mathcal{B}(\mathcal{H})^+.$$

Examples

1. $S_{x^2}(A,B) = \mathrm{Tr}\, A^2 B^{-1} = S_{x^2}^{max}(A,B)$
2. $S_{x\log x}(A,B) = \mathrm{Tr}\, A(\log A - \log B) = S(A,B)$
3. $S_{x\log x}^{max} = \mathrm{Tr}\, A \log(A^{1/2} B^{-1} A^{1/2}) = S_{BS}(A,B)$

$$S(A,B) \leq S_{BS}(A,B)$$

Petz's monotone metrics

- Denote $\mathcal{D}_n$ the invertible $n \times n$ density matrices. This is a differentiable manifold with tangent space at the footpoint ρ

$$\mathcal{T}_\rho \mathcal{D}_n \equiv \{A \in M_n^{sa} : \mathrm{Tr}\, A = 0\}$$

- Recall that a Riemannian metric $\mathfrak{g}_\rho$ with footpoint ρ on $\mathcal{D}_n$ is called **monotone metric** if

$$\mathfrak{g}_{T(\rho)}(T(A), T(A)) \leq \mathfrak{g}_\rho(A, A)$$

for all TPCP map T and $A \in \mathcal{T}_\rho \mathcal{D}_n$.

An operator monotone function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is called **symmetric** if $f(x) = xf(x^{-1})$ and **normalized** if $f(1) = 1$. With each symmetric and normalized operator monotone function f we associate its **Morozova-Chentsov function** via

$$c_f(x,y) := \frac{1}{yf(\frac{x}{y})}, \quad \text{and conversely} \quad f(x) = \frac{1}{c_f(x,1)}.$$

There is a plethora of suitable f functions, for example

$$\frac{2x^{\alpha+1/2}}{1+x^{2\alpha}}, \quad \frac{x-1}{\log x}, \quad \frac{x-1}{\log x}\frac{2\sqrt{x}}{1+x}, \quad \frac{1+x}{2}, \ldots$$

where $\alpha \in [0, 1/2]$.

Petz's Theorem on characterisation of monotone metrics

Theorem

There exists a bijective correspondence between monoton metrics on $\mathcal{D}_n$ and symmetric, normalized operator monotone functions f on $(0, \infty)$, given by

$$\mathfrak{g}_\rho^f(A, B) = \langle A, (f(L_\rho R_\rho^{-1})R_\rho)^{-1} B\rangle = \mathrm{Tr}\,(A c_f(L_\rho, R_\rho)(B)),$$

where $A, B \in \mathcal{T}_\rho \mathcal{D}_n$.

At a point where ρ is diagonal, $\rho = \mathrm{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, the lenght square of any tangent vector A is

$$\mathfrak{g}_\rho^f(A, A) = \|A\|^2 = \frac{1}{4}\left(f''(1) \sum_i \frac{A_{ii}^2}{\lambda_i} + 2 \sum_{i<j} c_f(\lambda_i, \lambda_j)|A_{ij}|^2 \right).$$

Monotone metrics as induced metric by f-divergences

Theorem (Lesniewski-Ruskai)

Any monotone metric is obtained from standard f-divergence by derivation

$$\mathfrak{g}_\rho^{\mathfrak{f}}(A,B) = -\frac{\partial^2}{\partial t \partial s} S_F(\rho + tA, \rho + sB)|_{t=s=0},$$

where the relation of the function F to the function f is

$$\frac{1}{f(x)} = \frac{F(x) + xF(x^{-1})}{(x-1)^2}.$$

Metric adjusted skew informations

It is useful to decompose the tangent space

$$\mathcal{T}_\rho \mathcal{D}_n = \{A \in \mathcal{M}_n : A = A^*, \operatorname{Tr} A = 0\}$$

as the direct sum of a "commuting" and a "non-commuting" part w.r.t. ρ. We set

$$(\mathcal{T}_\rho \mathcal{D}_n)^c = \{A \in \mathcal{T}_\rho \mathcal{D}_n : [A, \rho] = 0\}$$

the **commuting part** and define $(\mathcal{T}_\rho \mathcal{D}_n)^o$ as the **orthogonal complement** of $(\mathcal{T}_\rho \mathcal{D}_n)^c$ w.r.t. the Hilbert-Schmidt inner product. Then

$$\mathcal{T}_\rho \mathcal{D}_n = (\mathcal{T}_\rho \mathcal{D}_n)^c \oplus (\mathcal{T}_\rho \mathcal{D}_n)^o.$$

Whenever $A \in (\mathcal{T}_\rho \mathcal{D}_n)^c$, we have $\mathfrak{g}^f_\rho(A, A) = \operatorname{Tr} \rho^{-1} A^2$.

A tipical element of $(\mathcal{T}_\rho \mathcal{D}_n)^o$ is $i[\rho, K]$, $(K \in M_n^{sa})$ and we can define the **metric adjusted skew information** by

$$I_\rho^f(K) = \frac{f(0)}{2} \mathfrak{g}_\rho^f(i[\rho, K], i[\rho, K]).$$

- With the choice $f(x) = \frac{(\sqrt{x}+1)^2}{4}$ we get the **Wigner-Yanase information**:

$$I_\rho^f(K) = -\frac{1}{2} \mathrm{Tr}\,([K, \sqrt{\rho}])^2 = \mathrm{Tr}\,\rho K^2 - \mathrm{Tr}\,\rho^{1/2} K \rho^{1/2} K.$$

- With the choice $f(x) = \frac{1+x}{2}$ we get the **quantum Fisher information**, with $\rho = \sum_k \lambda_k |k\rangle\langle k|$:

$$I_\rho^f(K) = F_Q[\rho, K] = 2 \sum_{k,l} \frac{(\lambda_k - \lambda_l)^2}{\lambda_k + \lambda_l} |\langle k|K|l\rangle|^2.$$

- For a general f we can write explicitly:

$$I_\rho^f(K) = \frac{f(0)}{2} \sum_{k,l} \frac{(\lambda_k - \lambda_l)^2}{\lambda_l f(\lambda_k/\lambda_l)} |\langle k|K|l\rangle|^2.$$

QOT via quantum channels

The approach of De Palma and Trevisan[1]

- For any $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, the set $\mathcal{M}(\rho, \sigma)$ of *quantum transport maps* from ρ to σ is the set of the quantum channels (CPTP maps) such that

$$\Phi : \mathcal{T}_1(\mathrm{supp}\,(\rho)) \to \mathcal{T}_1(\mathcal{H}), \quad \Phi(\rho) = \sigma.$$

- We can associate with any $\Phi \in \mathcal{M}(\rho, \sigma)$ the quantum state $\Pi_\Phi \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H}^*)$ by

$$\Pi_\Phi = \left(\Phi \otimes I_{\mathcal{T}_1(\mathcal{H}^*)}\right) \left(||\sqrt{\rho}\rangle\rangle \langle\langle\sqrt{\rho}||\right).$$

[1] G. De Palma and D. Trevisan, *Quantum optimal transport with quantum channels*, Ann. Henri Poincaré **22** (2021), 3199–3234.

- Since
$$\mathrm{Tr}_{\mathcal{H}}\Pi_{\Phi} = \rho^T \quad \text{ad} \quad \mathrm{Tr}_{\mathcal{H}^*}\Pi_{\Phi} = \sigma,$$
where X^T is the transpose map, i.e. $X^T \langle\phi| = \langle\phi| X$, it induce the following definition:
- The set of **quantum couplings** assosiated with $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ is
$$\mathcal{C}(\rho, \sigma) = \{\Pi \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H}^*) : \mathrm{Tr}_{\mathcal{H}}\Pi = \rho^T, \mathrm{Tr}_{\mathcal{H}^*}\Pi = \sigma\}.$$
- De Palma and Trevisan showed that for any $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, the map $\Phi \mapsto \Pi_{\Phi}$ is a bijection between $\mathcal{M}(\rho, \sigma)$ and $\mathcal{C}(\rho, \sigma)$, that is in striking contrast to the classical case, the quantum couplings are in one-to-one correspondance with the quantum transport maps.
- Why? The primary reason: quantum channels **can "split mass"**, i.e. they can send pure states to mixed states.

- The **cost operator** for fixed self-adjoint operators $\{A_i\}_{i=1}^N$:

$$C = \sum_{j=1}^{N} \left(A_j \otimes I_{\mathcal{H}^*} - I_{\mathcal{H}} \otimes A_j^T \right)^2$$

- The transport cost for a coupling Π is

$$C(\Pi) = \mathrm{Tr}_{\mathcal{H}\otimes\mathcal{H}^*} \Pi C$$

- The **quantum Wasserstein (pseudo-)distance** $D_C(\rho, \sigma)$ is defined by

$$D_C^2(\rho, \sigma) = \inf_{\Pi \in \mathcal{C}(\rho,\sigma)} C(\Pi)$$

Some strange properties

- $D_C(\rho,\sigma) = D_C(\sigma,\rho)$ $\surd$
- If $\rho = \sigma$ then the optimal transport map corresponds to the identity map $\Phi = I$, so $D_C(\rho,\rho)^2 = C\left(\left|\left|\sqrt{\rho}\right\rangle\right\rangle\left\langle\left\langle\sqrt{\rho}\right|\right|\right)$ and

$$
\begin{aligned}
D_C(\rho,\rho)^2 &= -\sum_{i=1}^{N} \mathrm{Tr}\left([A_i, \sqrt{\rho}]^2\right) \\
&= 2\sum_{i=1}^{M} \left(\mathrm{Tr}\left(\rho A_i^2\right) - \mathrm{Tr}\left(\sqrt{\rho} A_i \sqrt{\rho} A_i\right)\right),
\end{aligned}
$$

  which is the famous **the Wigner – Yanase information**!
- For any $\rho, \tau, \sigma \in \mathcal{S}(\mathcal{H})$ the modified triangle inequality holds:

$$
D_C(\rho,\sigma) \le D_C(\rho,\tau) + D_C(\tau,\tau) + D_C(\tau,\sigma).
$$

Our contribution[2]

A bipartite quantum state is **separable** if it can be given as

$$\sum_k p_k |\Psi_k\rangle\langle\Psi_k| \otimes |\Phi_k\rangle\langle\Phi_k|,$$

with $\sum_k p_k = 1$. If a state cannot be written in this form, then it is called **entangled**. We denote the convex set of separable states by $\mathcal{S}_{sep}$. We define the modified **quantum Wasserstein (pseudo-)distance** by

$$D^2_{sep}(\rho, \sigma) = \inf_{\Pi} C(\Pi) = \inf_{\Pi} \sum_{j=1}^{N} \mathrm{Tr}\left(A_j \otimes I_{\mathcal{H}^*} - I_{\mathcal{H}} \otimes A_j^T\right)^2 \Pi,$$

where $\Pi \in \mathcal{C}(\rho, \sigma) \cap \mathcal{S}_{sep}$ are the separable couplings of the marginals ρ and σ.

[2] Géza Tóth, J.P. *Quantum Wasserstein distance based on an optimization over separable states*, Quantum 7 (2023), 1143

- For two qubits, **it is computable numerically** with semidefinite programming.
- In general,

$$D_{sep}(\rho,\sigma) \geq D(\rho,\sigma).$$

- If the relation

$$D_{sep}(\rho,\sigma) > D(\rho,\sigma)$$

holds, then all optimal Π for $D(\rho,\sigma)$ is entangled.

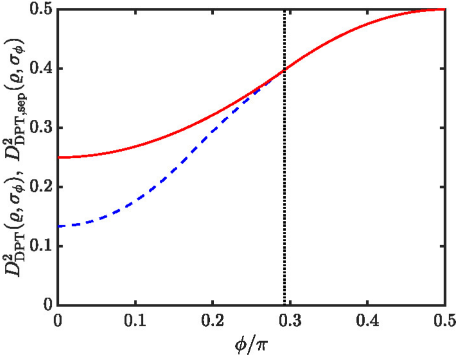
Let us consider the distance between two single-qubit mixed states

$$\rho = \frac{1}{2}|1\rangle\langle 1| + \frac{1}{4}I,$$

and

$$\sigma_\phi = e^{-i\frac{\sigma_y}{2}\phi}\rho^{+i\frac{\sigma_y}{2}\phi},$$

for $N = 1$ and $A_1 = \sigma_z$.



Thus, an entangled Π can be cheaper than a separable one.

The modified self-distance

- For the self-distance in the modified case for $N = 1$ we get

$$D_{sep}(\rho, \rho)^2 = \frac{1}{4} F_Q[\rho, A],$$

where

$$F_Q[\rho, A] = 2 \sum_{k,l} \frac{(\lambda_k - \lambda_l)^2}{\lambda_k + \lambda_l} |\langle k|A|l\rangle|^2,$$

the **quantum Fisher information** of the state $\rho = \sum_k \lambda_k |k\rangle\langle k|$ w.r.t the selfadjoint operator A.

- Note that

$$I_\rho(A) \leq \frac{1}{4} F_Q[\rho, A] \leq (\Delta A)^2_\rho,$$

where $I_\rho(A)$ is the Wigner-Yanase information and $(\Delta A)^2_\rho$ is the variance.

Summary

- For the quantum Wasserstein distance, we restrict the optimization to separable states.
- Then, the self-distance is the quarter of the quantum Fisher information.
- We found a fundamental connection from quantum optimal transport to quantum entanglement theory and quantum metrology.

Thank you for your kind attention!