Mathematical Aspects of Quantum Information Theory:

Lecture 4

Dario Trevisan

Università di Pisa
dario.trevisan@unipi.it

Plan

1. Inequalities (conclusion)
   - Lieb's concavity theorem
2. Distances
   - Trace distance
   - Fidelity
   - Quantum optimal transport

Plan

Theorem (Lieb's concavity theorem, monotonicity version)

Let H, $\tilde{H}$ be finite dimensional quantum systems,

- *let $\Phi : \mathcal{L}(\tilde{H}) \to \mathcal{L}(H))$ be CP and unital*
- *so that $\Phi^\dagger$ is a quantum channel from H to $\tilde{H}$,*
- *let $X, Y \in \mathcal{O}_{\geq}(H)$ be positive, $K \in \mathcal{L}(\tilde{H})$.*

Then, for every $t \in [0, 1]$,

$$\mathrm{tr}[\Phi(K)^* X^{1-t} \Phi(K) Y^t] \leq \mathrm{tr}[K^* \Phi^\dagger(X)^{1-t} K \Phi^\dagger(Y)^t].$$

In the case $K = \mathbb{1}_H$, we have $\Phi(\mathbb{1}_H) = \mathbb{1}_{\tilde{H}}$, hence the inequality becomes

Theorem (Lieb's concavity theorem, monotonicity version)

Let H, $\tilde{H}$ be finite dimensional quantum systems,

- *let $\Phi : \mathcal{L}(\tilde{H}) \to \mathcal{L}(H))$ be CP and unital*
- *so that $\Phi^\dagger$ is a quantum channel from H to $\tilde{H}$,*
- *let $X, Y \in \mathcal{O}_{\geq}(H)$ be positive, $K \in \mathcal{L}(\tilde{H})$.*

Then, for every $t \in [0, 1]$,

$$\mathrm{tr}[\Phi(K)^* X^{1-t} \Phi(K) Y^t] \leq \mathrm{tr}[K^* \Phi^\dagger(X)^{1-t} K \Phi^\dagger(Y)^t].$$

In the case $K = \mathbb{1}_H$, we have $\Phi(\mathbb{1}_H) = \mathbb{1}_{\tilde{H}}$, hence the inequality becomes

Remarks

- The inequality
$$\operatorname{tr}[\Phi^\dagger(X^{1-t}Y^t)] \leq \operatorname{tr}[\Phi^\dagger(X)^{1-t}\Phi^\dagger(Y)^t]$$
seems a Hölder inequality $p = 1/(1-t)$ and $p' = 1/t$.
- By monotonicity of operator means, we already have
$$\Phi^\dagger(X \sharp_t Y) \leq \Phi^\dagger(X) \sharp_t \Phi^\dagger(Y),$$
- But X and Y do not necessarily commute. The main idea is to move to a "higher" level to partially restore commutativity.
- The concavity version of Lieb's theorem states that
$$(A, B) \mapsto \operatorname{tr}[K^* A^{1-t} K B^t]$$
is concave on $\mathcal{O}_{\geq 0}(H) \times \mathcal{O}_{\geq 0}(H)$. To deduce it from the monotonicity version, use $\Phi^\dagger = \operatorname{tr}_{\mathbb{C}^2}$ on $H \otimes \mathbb{C}^2$ with
$$X = \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix}, \quad Y = \begin{pmatrix} B_0 & 0 \\ 0 & B_1 \end{pmatrix}, \quad K = \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix}.$$

Proof

- For simplicity, consider $H = \tilde{H}$. Recall that $\mathcal{L}(H)$ is a Hilbert space endowed with the Hilbert-Schmidt scalar product ($\langle A|B\rangle = \mathrm{tr}[A^* B]$) (Dirac notation $|X\rangle, |Y\rangle, |K\rangle \in \mathcal{L}(H)$).
- The map Φ is linear from $\mathcal{L}(H)$ into itself, $\Phi|K\rangle = |\Phi(K)\rangle$.
- Its adjoint is $\Phi^* = \Phi^\dagger$.
- We use Φ as a single Kraus operator to define a CP map

$$\tilde{\Phi} : \mathcal{L}(\mathcal{L}(H)) \to \mathcal{L}(\mathcal{L}(H)), \quad A \mapsto \tilde{\Phi}(A) = \Phi^* A \Phi.$$

- We introduce two further operators:
  - the left-multiplication by $X \in \mathcal{L}(H)$

$$L_X \in \mathcal{L}(\mathcal{L}(H)), \quad L_X : \mathcal{L}(H) \to \mathcal{L}(H), \quad L_X |A\rangle \mapsto |XA\rangle \,,$$

  - the right multiplication by $Y \in \mathcal{L}(H)$,

$$R_Y \in \mathcal{L}(\mathcal{L}(H)), \quad R_Y : \mathcal{L}(H) \to \mathcal{L}(H), \quad R_Y |A\rangle \mapsto |AY\rangle \,.$$

- Basic properties:

$$L_X L_{X'} = L_{XX'}, \quad R_Y R_{Y'} = R_{Y'Y}, \quad L_X^* = L_{X^*}, \quad R_Y^* = R_{Y^*}.$$

- Any operator L_X commutes with any R_Y, since

- Therefore,

$$(L_X)\sharp_t(R_Y) = L_X^{1-t} R_Y^t = L_{X^{1-t}} R_{Y^t}.$$

- It holds $L_{\Phi^\dagger(X)} \geq \tilde{\Phi}(L_X) = \Phi^* L_X \Phi$, since for every $K \in \mathcal{L}(H)$,

- By the monotonicity of the operator mean,

- To conclude, take the scalar product both sides with $K \in \mathcal{L}(H)$.

Distances between classical probability distributions

How to compare two states of a quantum system? The answer of course depends on the application and justifies a large variety of "distances".

We discuss the quantum analogues of

1. the total variation distance $\Rightarrow$ trace distance
2. the Hellinger distance, or Bhattacharyya coefficient $\Rightarrow$ Fidelity
3. the Kantorovich-Wasserstein distance, $\Rightarrow$ several analogues.

Plan

1. Inequalities (conclusion)
   - Lieb's concavity theorem
2. Distances
   - Trace distance
   - Fidelity
   - Quantum optimal transport

- For probability distributions p, q over Ω, the total variation distance is

$$\|p-q\|_{TV} = \frac{1}{2}\sum_{\omega\in\Omega} |p(\omega)-q(\omega)| \in [0,1],$$

- Given a Markov kernel $N = (N(\omega, x))_{\omega\in\Omega, x\in\mathcal{X}}$,

$$\|(N^\dagger p) - (N^\dagger q)\|_{TV} \leq \|p-q\|_{TV}.$$

- Dual representation:

$$\|p-q\|_{TV} = \sup_{V\subseteq\Omega} \sum_{\omega\in V} p(\omega)-q(\omega).$$

(optimizer: $V = \{\omega\in\Omega : p(\omega) > q(\omega)\}$)

- We can further relax from sets to functions:

$$\|p-q\|_{TV} = \sup_{a:\Omega\to[0,1]} \sum_{\omega\in V} a(\omega)\,(p(\omega)-q(\omega))\,.$$

- On a finite-dimensional quantum system H, given states $\rho, \sigma \in \mathcal{S}(H)$, one defines their trace distance

$$D_{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2}\mathrm{tr}|\rho - \sigma|$$

where $|\rho - \sigma|$ is defined via spectral calculus on $\rho - \sigma \in \mathcal{O}(H)$.

- It is a special case of p-Schatten norm

$$\|A\|_p := \mathrm{tr}[|A|^p]^{1/p} = \mathrm{tr}[(A^* A)^{p/2}]^{1/p}.$$

- If both ρ and σ are diagonal with respect to the same basis $(|i\rangle)_{i \in I}$, i.e.,

$$\rho = \sum_{i \in I} p_i |i\rangle \langle i| , \quad \sigma = \sum_{i \in I} q_i |i\rangle \langle i| ,$$

for classical probability densities $(p_i)_{i \in I}$, $(q_i)_{i \in I}$, then

$$D_{\mathrm{tr}}(\rho, \sigma) = \|p - q\|_{TV}.$$

- $\rho = |\psi\rangle \langle\psi|$, $\sigma = |\varphi\rangle \langle\varphi| \in \mathcal{S}(H)$ are pure states then,

$$D_{\mathrm{tr}}(\rho, \sigma) = \sqrt{1 - |\langle\psi|\varphi\rangle|^2}.$$

- Dual representation

$$D_{\mathrm{tr}}(\rho, \sigma) = \sup_{V < H} \mathrm{tr}[\mathbb{1}_V(\rho - \sigma)] = \sup_{V < H} \mathbb{P}_\rho(V) - \mathbb{P}_\sigma(V),$$

- and “relaxed” version

$$D_{\mathrm{tr}}(\rho, \sigma) = \sup_{A \in \mathcal{O}(H), \sigma(A) \subseteq [0,1]} (A)_\rho - (A)_\sigma.$$

- Sketch of proof:

- $\Rightarrow$ the triangle inequality and, for any quantum channel $\Phi^\dagger$,

$$D_{\mathrm{tr}}(\Phi^\dagger(\rho), \Phi^\dagger(\sigma)) \leq D_{\mathrm{tr}}(\rho, \sigma).$$

- Examples: $D_{\mathrm{tr}}(\mathrm{tr}_H[\rho], \mathrm{tr}_H[\sigma]) \leq D_{\mathrm{tr}}(\rho, \sigma). \quad \forall \rho, \sigma \in \mathcal{S}(H \otimes K)$,

$$\|\mathbb{P}_\rho(X = \cdot) - \mathbb{P}_\sigma(X = \cdot)\|_{TV} \leq D_{\mathrm{tr}}(\rho, \sigma), \qquad \text{for any } X = (\mathbb{1}_{V_x})_{x \in \mathcal{X}}.$$

Plan

1. Inequalities (conclusion)
   - Lieb's concavity theorem
2. Distances
   - Trace distance
   - Fidelity
   - Quantum optimal transport

- The (squared) Hellinger distance between probability distributions p, q is

$$H^2(p,q) = \frac{1}{2}\sum_{\omega\in\Omega} |\sqrt{p(\omega)} - \sqrt{q(\omega)}|^2.$$

- The Bhattacharyya coefficient is

$$BC(p,q) = \sum_{\omega\in\Omega} \sqrt{p(\omega)q(\omega)} \in [0,1],$$

so that

$$H(p,q) = 1 - BC(p,q).$$

- For every Markov kernel N,

$$H(N^\dagger p, N^\dagger q) \leq H(p,q)$$

or equivalently

$$BC(N^\dagger p, N^\dagger q) \geq BC(p,q).$$

- In the quantum setting the analogues of H and BC have a natural interpretation, in particular for pure states $\rho = |\psi\rangle\langle\psi|$.
- The fidelity between ρ, $\sigma \in \mathcal{S}(H)$ is defined as

$$F(\rho,\sigma) = \left(\mathrm{tr}[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}]\right)^2.$$

- The analogue of the Hellinger distance is the Bures metric

$$D_B(\rho,\sigma)^2 = 2\left(1 - \sqrt{F(\rho,\sigma)}\right),$$

- For pure states $\rho = |\psi\rangle\langle\psi|$, $\sigma = |\varphi\rangle\langle\varphi|$, since $\sqrt{\rho} = \rho$,

$$F(\rho,\sigma) = |\langle\psi|\varphi\rangle|^2$$

- Several equivalent ways to write the fidelity:

$$F(\rho,\sigma) = \mathrm{tr}[\left|\sqrt{\rho}\sqrt{\sigma}\right|^2] = \frac{1}{2}\left(\mathrm{tr}[(\rho\sharp_{1/2}\sigma^{-1})\sigma] + \mathrm{tr}[(\sigma\sharp_{1/2}\rho^{-1})\rho]\right),$$

- Fon any quantum channel $\Phi^\dagger$,

$$F(\Phi^\dagger(\rho),\Phi^\dagger(\sigma)) \geq F(\rho,\sigma),$$

- The proof uses the variational representation

$$F(\rho,\sigma) = \sup\left\{\left|\mathrm{tr}[X]\right| : X \in \mathcal{L}(H), \text{ such that } M = \begin{pmatrix} \rho & X \\ X^* & \sigma \end{pmatrix} \geq 0\right\}.$$

Plan

- The classical optimal transport problem (Monge-Kantorovich) searches for the cheapest way to move two probability distributions p, q, with respect to a displacement cost c.

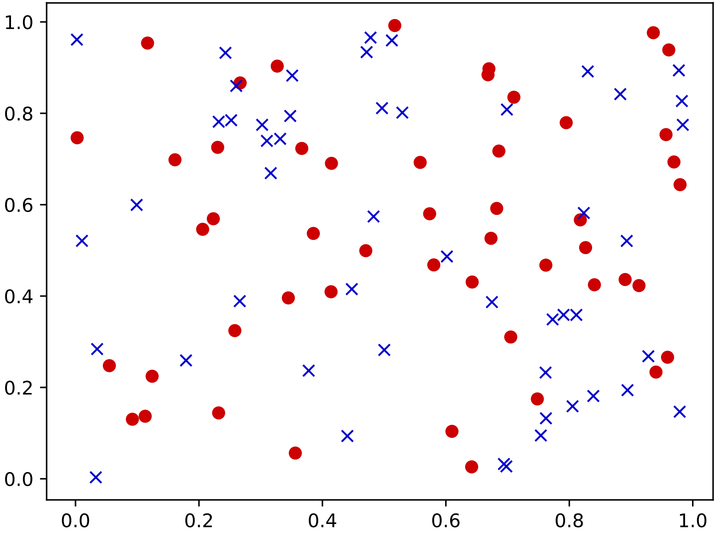
- The precise definition is given by the variational problem

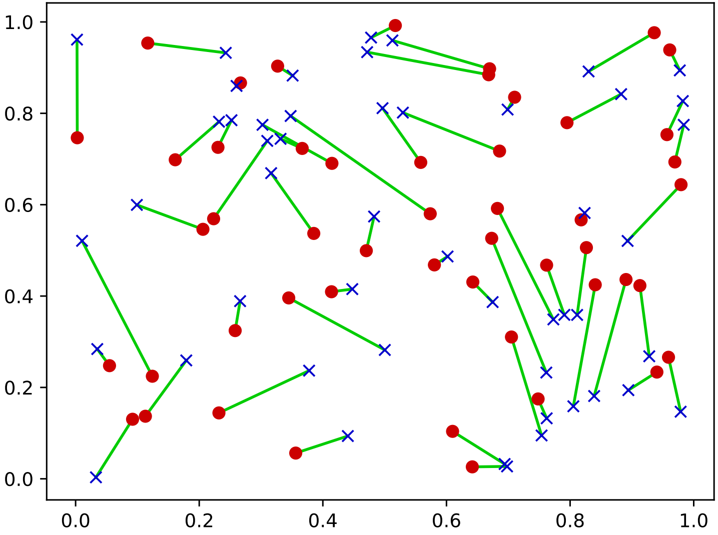
$$W^c(p,q) := \inf_{\pi \in \mathcal{C}(p,q)} \sum_{x,y \in \mathcal{X}} c(x,y)\pi(x,y),$$

where $\mathcal{C}(p,q)$ are the couplings between p, q, i.e., $\pi = (\pi(x,y))_{x,y\in\mathcal{X}}$,

$$\forall x, y \in \mathcal{X}, \quad 0 \le \pi(x,y) \le 1,$$

$$\sum_{x\in\mathcal{X}} \pi(x,y) = q(y), \sum_{y\in\mathcal{X}} \pi(x,y) = p(x).$$


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

- We give a "dynamical" description via the conditional probabilities

$$N(x, y) = \pi(y|x) = \frac{\pi(x, y)}{p(x)},$$

which define a Markov kernel such that $N^{\dagger} p = q$ (a transport plan)

- A further point of view is the dual formulation,

$$W^c(p, q)$$
$$= \sup \left\{ \sum_{x \in \mathcal{X}} f(x)p(x) + \sum_{y \in \mathcal{X}} g(y)q(y) \ : \ f(x) + g(y) \leq c(x, y) \ \forall x, y \in \mathcal{X} \right\}.$$

- When $c(x, y) = d(x, y)$ is a distance the dual problem can be restricted to 1-Lipschitz functions, i.e., the Kantorovich problem

$$W^d(p, q) = \sup_{f \text{ is 1-Lip}} \left\{ \sum_{x \in \mathcal{X}} f(x) \left(p(x) - q(x) \right) \right\}.$$

- Optimal transport is a fundamental problem in operation research and combinatorial optimization, which in recent decades found applications in
  - PDE's
  - Riemannian geometry
  - computer science
  - statistics and machine learning...
- The first proposals for quantum optimal transport date back to the 1990's (Connes, Zyczkowski).
- Recently, more formulations have been proposed: Agredo, Carlen-Maas, Golse-Mouhot-Paul-Caglioti, ...).

OT via quantum couplings

- This approach is considered by Golse, Mouhot, Paul and Caglioti.

- Given a quantum system H, consider two copies $H = H_1 = H_2$, introduce a cost operator as $C \in \mathcal{O}(H_1 \otimes H_2)$, e.g. a sum of squares

$$C = \sum_{i \in I} (A_i \otimes \mathbb{1}_{H_2} - \mathbb{1}_{H_1} \otimes A_i)^2.$$

- Quantum couplings $\mathcal{C}(\rho, \sigma)$ are density operators $\Pi \in \mathcal{S}(H_1 \otimes H_2)$ with

$$\rho = \mathrm{tr}_{H_2}[\Pi], \quad \sigma = \mathrm{tr}_{H_1}[\Pi].$$

- The optimal transport cost is

$$W_C(\rho, \sigma) = \inf_{\Pi \in \mathcal{C}(p,q)} \mathrm{tr}[C\Pi].$$

- Many open questions: is $\sqrt{W_C(\rho,\sigma)}$ is an actual distance?

- Given a quantum channel $\Phi^\dagger$, determine how much it expands the cost:

$$W_C(\Phi^\dagger(\rho), \Phi^\dagger(\sigma)) \leq \|\Phi^\dagger\|_{W_C} W_C(\rho,\sigma), \quad \text{for every } \rho,\sigma \in \mathcal{S}(H)?$$

- This OT is used in the infinite-dimensional (CCR) setting to investigate quantitatively semiclassical limits.

OT via quantum channels

- With DePalma we proposed to use channels such that $\Phi^\dagger(\rho) = \sigma$ as quantum plans. How to define a cost functional?

- Consider the "sum of squares" case. In the classical case,

$$\sum_{x,y\in\mathcal{X}} c(x,y)\pi(x,y) = \sum_{i\in I}\sum_{x,y\in\mathcal{X}} (f_i(x) - f_i(y))^2 \pi(x,y)$$

$$= \sum_{i\in I}\sum_{x\in\mathcal{X}} f_i^2(x)p(x) - \sum_{y\in\mathcal{X}} f_i^2(y) - 2\sum_{x,y\in\mathcal{X}} f_i(x)f_i(y)\pi(x,y).$$

- We rewrite using $N(x,y) = \pi(y|x)$ instead of π,

$$\sum_{x,y\in\mathcal{X}} f_i(x)f_i(y)\pi(x,y) = \sum_{x\in\mathcal{X}} f_i(x)p(x)(Nf_i)(x).$$

- As a quantum analogue of $\sum_{x\in\mathcal{X}} f_i(x)p(x)(Nf_i)(x)$, we propose

$$\mathrm{tr}[A_i\sqrt{\rho}\Phi(A_i)\sqrt{\rho}] = \mathrm{tr}[(\sqrt{\rho}A_i)^*\Phi(A_i)\sqrt{\rho}] = \langle\sqrt{\rho}A_i|\Phi(A_i)\sqrt{\rho}\rangle\,.$$

- The full expression of the cost becomes

$$\mathrm{Cost}(\Phi,\rho,\sigma) = \sum_{i\in I}\mathrm{tr}[A_i^2\rho] + \mathrm{tr}[A_i^2\sigma] - 2\mathrm{tr}[A_i\sqrt{\rho}\Phi(A_i)\sqrt{\rho}].$$

- We minimize with respect to quantum plans,

$$W_P(\rho,\sigma) = \inf_{\Phi^\dagger(\rho)=\sigma}\mathrm{Cost}(\Phi,\rho,\sigma).$$

- An optimal plan between ρ and itself in the identity channel $\Phi = \mathbb{1}_{\mathcal{L}(H)}$ and

$$W_P(\rho,\sigma) \geq \frac{1}{2}\left(W_P(\rho,\rho) + W_P(\sigma,\sigma)\right).$$

- A modified triangle inequality holds:

$$\sqrt{W_P(\rho,\tau)} \leq \sqrt{W_P(\rho,\sigma)} + \sqrt{W_P(\sigma,\sigma)} + \sqrt{W_P(\sigma,\tau)}.$$

- Recalling the definition of the cost, inequality

$$W_P(\rho,\sigma) \geq \frac{1}{2}\left(W_P(\rho,\rho) + W_P(\sigma,\sigma)\right),$$

  turns out to be equivalent to

$$\sum_{i\in I} \mathrm{tr}[A_i\sqrt{\rho}\Phi(A_i)\sqrt{\rho}] \leq \frac{1}{2}\sum_{i\in I} \mathrm{tr}[A_i\sqrt{\rho}A_i\sqrt{\rho}] + \mathrm{tr}[A_i\sqrt{\sigma}A_i\sqrt{\sigma}]$$

  for every plan $\Phi^\dagger(\rho) = \sigma$.
- Argue for each $A_i = A$. By Cauchy-Schwarz inequality,

$$\begin{aligned}\mathrm{tr}[A\sqrt{\rho}\Phi(A)\sqrt{\rho}] &= \mathrm{tr}[(\rho^{1/4}A\rho^{1/4})(\rho^{1/4}\Phi(A)\rho^{1/4})] \\ &\leq \frac{1}{2}\mathrm{tr}[A\sqrt{\rho}A\sqrt{\rho}] + \frac{1}{2}\mathrm{tr}[\Phi(A)\sqrt{\rho}\Phi(A)\sqrt{\rho}]\end{aligned}$$

- By Lieb's concavity theorem with $t = 1/2$,

$$\mathrm{tr}[\Phi(A)\sqrt{\rho}\Phi(A)\sqrt{\rho}] \leq \mathrm{tr}[A\sqrt{\Phi^\dagger(\rho)}A\sqrt{\Phi^\dagger(\rho)}].$$

OT via Lipschitz operators

- In the classical case, we can use Kantorovich duality to define W^d:

$$W^d(p,q) = \sup_{f \text{ is 1-Lip}} \left\{ \sum_{x \in \mathcal{X}} f(x)\,(p(x) - q(x)) \right\}.$$

- A similar strategy in the quantum setting dates back to Connes: define first what are Lipschitz observables and obtain the cost via duality.
- Recently, we proposed to consider the case of product systems

$$H = \bigotimes_{i \in I} H_i,$$

providing a quantum analogue of OT with respect to the Hamming distance

- Recall that on sets $\Pi_{i \in I} \mathcal{X}_i$,

$$d_{\mathrm{Ham}}((x_i)_{i \in I}, (y_i)_{i \in I}) = \sum_{i \in I} 1_{\{x_i \neq y_i\}}.$$

- $f : \Pi_{i \in I} \mathcal{X}_i \to \mathbb{R}$ is (Hamming) 1-Lipschitz if and only if, for every $i \in I$,

$$|f(x) - f(y)| \leq 1$$

whenever x, y differ only at the coordinate i (write $x \sim_i y$).

- Equivalently, define the oscillation at $i \in I$ as

$$\partial_i f = \sup_{x \sim_i y} |f(x) - f(y)| = 2 \inf_{g_i} \sup_{x} |f(x) - g_i(x)|$$

where g_i does not depend upon the coordinate i. Then,

$$\|f\|_{\mathrm{Lip}} = \max_{i \in I} \partial_i f.$$

- On a product system $H = \bigotimes_{i \in I} H_i$, for every $i \in I$ and observable $A \in \mathcal{O}(H)$, define

$$\partial_i A = \sup \left\{ 2 \left\| A - G_i \otimes \mathbb{1}_{H_i} \right\|_\infty \; : \; G_i \in \mathcal{O}(\bigotimes_{j \neq i} H_j) \right\},$$

- The quantum Lipschitz constant of $A \in \mathcal{O}(H)$ is

$$\|A\|_L := \max_{i \in I} \partial_i A.$$

- The quantum Wasserstein distance of order 1 between $\rho, \sigma \in \mathcal{S}(H)$ is

$$\begin{aligned} \|\rho - \sigma\|_{W_1} &= \sup \left\{ \mathrm{tr}[A(\rho - \sigma)] \; : \; \|A\|_L \leq 1 \right\} \\ &= \sup \left\{ (A)_\rho - (A)_\sigma \; : \; \|A\|_L \leq 1 \right\} \end{aligned}$$

- Back to the classical case, forget about the product structure (i.e., considers the set $\mathcal{X} = \bigotimes_{i=1}^{1} \mathcal{X}$ a single factor): then the Hamming distance is the trivial distance and

$$W^{d_{trivial}}(p,q) = \|p-q\|_{TV}.$$

- Since

$$1_{\{x \neq y\}} \leq \sum_{i \in I} 1_{\{x_i \neq y_i\}} \leq |I| 1_{\{x \neq y\}},$$

  this leads to a comparison between OT distances.
- Also in the quantum case, we can compare

$$D_{\mathrm{tr}}(\rho,\sigma) \leq \|\rho-\sigma\|_{W_1} \leq |I| D_{\mathrm{tr}}(\rho,\sigma).$$

- For product states $\rho = \otimes_{i\in I}\rho_i$, $\sigma = \otimes_{i\in I}\sigma_i$, then

$$\|\rho - \sigma\|_{W_1} = \sum_{i\in I} D_{\mathrm{tr}}(\rho_i, \sigma_i).$$

- Exercise: Compute the Wasserstein distance of order 1 between any two Bell states on the composite system $H = \mathbb{C}^2 \otimes \mathbb{C}^2$, e.g.

$$\rho = \frac{1}{2}\left(|00\rangle + |11\rangle\right)\left(\langle 00| + \langle 11|\right),$$

$$\sigma = \frac{1}{2}\left(|01\rangle + |10\rangle\right)\left(\langle 01| + \langle 10|\right).$$