On quantum Wasserstein distance

József Pitrik
Wigner Research Centre for Physics, Budapest

Common work with Géza Tóth, Gergely Bunth, Dánel Virosztek and Tamás Titkos

Dedicated to the memory of Dénes Petz (1953 – 2018) on the occasion of his 70th birthday.

Prof. Dénes Petz and Prof. Fumio Hiai at the end of the '90s

1. The classical (Monge-Kantorovich) optimal transport problem
    - Monge Formulation
    - Kantorovich Formulation

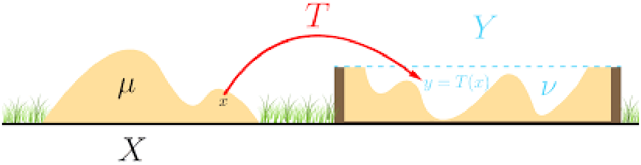
2. Wasserstein spaces
    - p-Wasserstein distance
    - Wasserstein barycenters

3. Quantum optimal transport
    - Basics
    - Transport by quantum couplings
    - Transport by quantum channels
    - Our contribution

What is Optimal Transport (OT)?

- The optimal transport problem seeks the most efficient way of transporting one distribution of mass into another.
- The problem was originally studied by Gaspard Monge in 1781: "Given a pile of sand and a pit of equal volume, how can one optimally transport the sand into the pit?"
  In: Mémoire sur la théorie des déblais et les remblais (Note on the theory of land excavation and infill)

Gaspard Monge
1746-1818

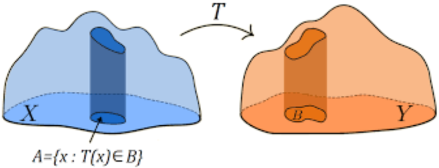


The classical optimal transport problem - Monge Formulation

- X – sand space : complete separable metric space with its Borel σ-algebra
- Y – pit space : complete separable metric space with its Borel σ-algebra
- $\mu \in \mathcal{P}(X)$ the sand distribution - probability measure over X
- $\nu \in \mathcal{P}(Y)$ the shape of the pit - probability measure over Y
- $c : X \times Y \to [0, \infty]$ Borel measurable **cost function**: $c(x, y)$ represents the cost of moving a unit of mass from $x \in X$ to $y \in Y$
- $T : X \to Y$ **transport map**

The map $T : X \to Y$ must be mass-preserving:

$$\mu(T^{-1}(B)) = \nu(B), \text{ for all } B \subset Y \text{ Borel}$$



$\nu \in \mathcal{P}(Y)$ is **push-forward measure** of $\mu \in \mathcal{P}(X)$ under the map T if

$$(T_{\#}\mu)(B) := \mu(T^{-1}(B)) = \nu(B),$$

for all $B \subset Y$ Borel measurable set. In other words if X is a random variable such that $Law(X) = \mu$, then

$$Law(T(X)) = T_{\#}\mu.$$

The total transport cost of the map $T : X \to Y$:

$$C(T) := \int_X c(x, T(x)) \mathrm{d}\mu(x)$$

The Monge problem

For given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to [0, \infty]$ to find the optimal transport map $T : X \to Y$, i.e. to solve

$$\inf\{C(T) = \int_X c(x, T(x)) \mathrm{d}\mu(x) : T_{\#}\mu = \nu\}$$

What can we say about the solution of the Monge problem?

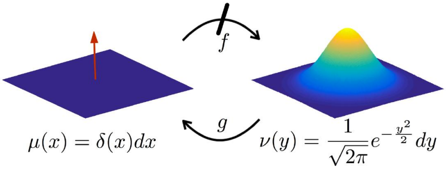
A transport map may not exist!

For example if $\mu = \delta_{x_0}$ is the Dirac measure at some $x_0 \in X$ but ν is not, then the set $B = \{T(x_0)\}$ satisfies

$$\mu(T^{-1}(B)) = 1 > \nu(B),$$

so no such T can exist! Why?

Because the mass at x_0 must be sent to a unique point $T(x_0)$, i.e. splitting the grains of sand is not allowed!



Remarks:

- The existence and the uniqueness of the solution depend heavily on the structure of the space, and on the cost function.
- Monge originally considered the case $X = Y = \mathbb{R}^3$, and the cost was the Euclidean distance $c(x, y) = \|x - y\|$. This original problem was extremely difficult, and the Academy of Paris offered a prize for its solution.
- The existence thory for the Monge problem was not fully understood until 1995. (Brenier '87, Gangbo & McCann '95.)

In the case

$$X = Y = \mathbb{R}^n, \quad c(x,y) = \|x-y\|^p, \quad 0 < p < \infty,$$

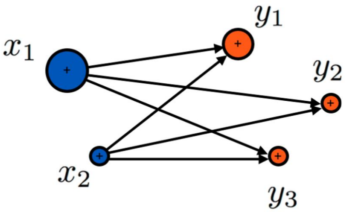
μ, ν are compactly supported:

- For $p > 1$, if μ, ν are absolutely continous with respect to Lebesgue measure, then there is a unique solution to the Monge problem.
- For $p = 2$ and $n \geq 2$ the unique optimal transport map is $T = \nabla\varphi$ for some convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$.
- For $p = 1$, if μ, ν are absolutely continous with respect to Lebesgue measure, then there are solutions of the Monge problem, but there is no uniqueness.
- For $p < 1$, there is in general no solution of the Monge problem, except if μ and ν are concentrated on disjoint sets.

The classical optimal transport problem - Kantorovich Formulation

Working on optimal allocation of scarce resources during World War II, Kantorovich revisited the optimal transport problem in 1942.

Leonid Kantorovich
1912-1986



Resource allocation

The classical optimal transport problem - Kantorovich Formulation

- X – sand space : complete separable metric space with its Borel σ-algebra
- Y – pit space : complete separable metric space with its Borel σ-algebra
- $\mu \in \mathcal{P}(X)$ the sand distribution - probability measure over X
- $\nu \in \mathcal{P}(Y)$ the shape of the pit - probability measure over Y
- $c : X \times Y \to [0, \infty]$ Borel measurable **cost function**: $c(x, y)$ represents the cost of moving a unit of mass from $x \in X$ to $y \in Y$

Instead of transport maps, we consider probability measures on the product space $X \times Y$. If $\pi \in \mathcal{P}(X \times Y)$, then $\pi(A \times B)$ is the amount of sand transported from the subset $A \subseteq X$ into the part of the pit represented by $B \subseteq Y$.

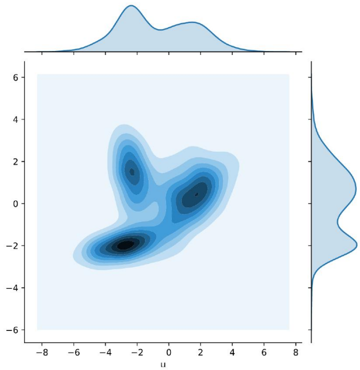
- The total mass sent from A is $\pi(A \times Y)$, and the total mass sent to B is $\pi(X \times B)$.
- π is mass-preserving iff

$$\pi(A \times Y) = \mu(A) \quad \text{for all } A \subset X \text{ Borel}$$

$$\pi(X \times B) = \nu(B) \quad \text{for all } B \subset Y \text{ Borel}$$

A probability measure π satisfying these conditions will be called **coupling** or **transport plan** of μ and ν.
The set of such couplings is denoted by $\Pi(\mu, \nu)$.

- If $\pi \in \Pi(\mu, \nu)$, then $\pi|_X = \mu$ and $\pi|_Y = \nu$ are the marginals.
- $\Pi(\mu, \nu)$ is never empty: it always contains the product measure $\mu \otimes \nu$ defined by $[\mu \otimes \nu](A \times B) = \mu(A)\nu(B)$



[1]

[1] Source: Wikipedia

Transport map vs. coupling

Let $T : X \to Y$ satisfy $T_{\#}\mu = \nu$. Consider the map

$$Id \times T : X \to X \times Y, \quad x \mapsto (x, T(x)),$$

and define

$$\pi_T := (Id \times T)_{\#}\mu \in \mathcal{P}(X \times Y).$$

Then $\pi_T \in \Pi(\mu, \nu)$, i.e.

$$\pi_T|_1 = \mu \quad \text{and} \quad \pi_T|_2 = \nu.$$

The total cost associated with $\pi \in \Pi(\mu, \nu)$ is

$$C(\pi) = \int_{X \times Y} c(x, y) \mathrm{d}\pi(x, y).$$

The Kantorovich problem

For given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to [0, \infty]$ to find the optimal transport plan $\pi \in \Pi(\mu, \nu)$, i.e. to solve

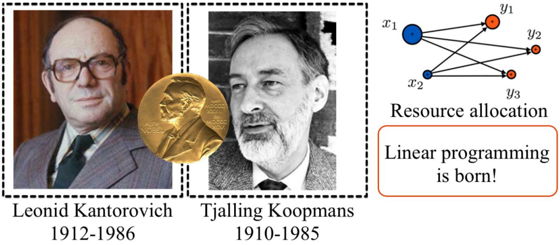
$$\inf\{C(\pi) = \int_{X \times Y} c(x, y) \mathrm{d}\pi(x, y) : \pi \in \Pi(\mu, \nu)\}$$

Probabilistic view:

$$\inf_{(X,Y)} \{\mathbb{E}[c(X, Y)] : X \sim \mu \text{ and } Y \sim \nu\}$$

Both the objective function $C(\pi)$ and the constraints for the coupling are linear in π, so the problem can be seen as infinite-dimensional linear programming.

In 1975, Kantorovich shared the Nobel Memorial Prize in Economic Sciences with Tjalling Koopmans "for their contributions to the theory of optimum allocation of resources."



In the case of discrete probability densities, with the transport plan

$$\pi(x_i, y_j) \geq 0$$

such that

$$\sum_j \pi(x_i, y_j) = p_i, \quad \sum_i \pi(x_i, y_j) = q_j,$$

the problem becomes **linear optimization with linear constaints**:

$$\min_\pi \sum_i \sum_j c(x_i, y_j)\pi(x_i, y_j)$$

that can be solved via simplex algorithm.

Kantorovich vs. Monge

- The Kantorovich problem admits a solution when the cost is continous.
- The Kantorovich problem is a relaxation of the Monge problem, because to each transport map T one can associate a coupling π_T, by

$$\pi_T(A \times B) := \mu(A \cap T^{-1}(B)), \quad \text{for all Borel } A \subseteq X,\ B \subseteq Y$$

  with the same cost, i.e. $C(T) = C(\pi_T)$.
- It follows that

$$\inf_{T : T_{\#}\mu = \nu} C(T) = \inf_{\pi_T : T_{\#}\mu = \nu} C(\pi) \geq \inf_{\pi \in \Pi(\mu,\nu)} C(\pi) = C(\pi^*),$$

  for some optimal π^*.

What is a Wasserstein space?

- Let $\mathcal{W}_p(X)$ be the set of Borel probability measures with finite p'th moment defined on a given complete separable metric space (X, d):

$$\mathcal{W}_p(X) = \left\{ \mu \in \mathcal{P}(X) \,\middle|\, \int_X d(x, \hat{x})^p \, \mathrm{d}\mu(x) < \infty \text{ for some } \hat{x} \in X \right\}.$$

- The **p-Wasserstein metric** W_p, for $p \geq 1$ on $\mathcal{W}_p(X)$ is then defined as the optimal transport problem with the cost function $c(x, y) = d^p(x, y)$. For $\mu, \nu \in \mathcal{W}_p(X)$

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{X^2} d(x, y)^p \, \mathrm{d}\pi(x, y) \right)^{\frac{1}{p}}.$$

where $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}\left(X^2\right) \,\middle|\, \pi|_1 = \mu, \pi|_2 = \nu \right\}$ is the collection of all *transport plans* between μ and ν.

The space of sufficiently concentrated probability measures $\mathcal{W}_p(X)$ endowed with the metric W_p is a separable and complete metric space, called **p–Wasserstein space**.

Example: quadratic Wasserstein distance of two Gaussians
$P = \mathcal{N}(m, C)$ is a normal distribution on $\mathbb{R}^n$ if its probability density function is

$$p(x) = \frac{\exp\left(-\frac{1}{2}(x-m)^T C^{-1}(x-m)\right)}{\sqrt{(2\pi)^n \det C}},$$

where $m \in \mathbb{R}^n$ is its expected value and C is a symmetric postive-definite $n \times n$ matrix, the covariance matrix.

If $P_1 = \mathcal{N}(m_1, C_1)$ and $P_2 = \mathcal{N}(m_2, C_2)$, then their 2-Wasserstein distance, wrt. the usual Euclidean norm on $\mathbb{R}^n$ is

$$W_2(P_1, P_2)^2 = \|m_1 - m_2\|_2^2 + \mathrm{Tr}\left(C_1 + C_2 - 2(C_2^{1/2} C_1 C_2^{1/2})^{1/2}\right).$$

Fun fact: if ρ_1 and ρ_2 are density matrices, then their Bures distance D_B is given by

$$D_B^2(\rho_1, \rho_2) = \mathrm{Tr}\left(\rho_1 + \rho_2 - 2(\rho_2^{1/2} \rho_1 \rho_2^{1/2})^{1/2}\right),$$

and their *fidelity* is

$$F(\rho_1, \rho_2) = \mathrm{Tr}\,(\rho_2^{1/2} \rho_1 \rho_2^{1/2})^{1/2}.$$

In general if (X, Σ) is a measurable space and $\mathcal{P}(X)$ is the space of probability measures on X, there is a lot of possibility to define distances and divergences between two diributions $P, Q \in \mathcal{P}(X)$ to measure their dissimilarity:

- The **Total Variation (TV)** distance

$$TV(P, Q) = \sup_{A \in \Sigma} |P(A) - Q(A)|.$$

- The **Kullback-Leibler divergence (KL)**

$$KL(P||Q) = \begin{cases} \int_X \log\left(\frac{p(x)}{q(x)}\right) p(x) \mathrm{d}\mu(x), & \text{if } \mathrm{supp}\,(P) \cap \ker Q = \{0\} \\ +\infty, & \text{if } \mathrm{supp}\,(P) \cap \ker Q \neq \{0\}, \end{cases}$$

where $P(A) = \int_A p(x) \mathrm{d}\mu(x)$ and $Q(A) = \int_A q(x) \mathrm{d}\mu(x)$ for all $A \in \Sigma$.

- The **Jensen-Shannon divergence (JS)**

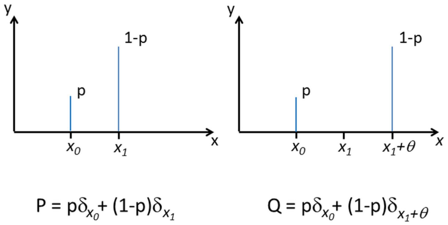
$$JS(P,Q) = KL(P||M) + KL(Q||M),$$

where $M = \frac{P+Q}{2}$ is the mixture.

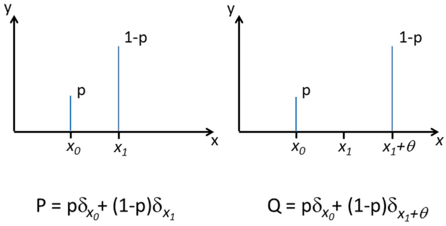
These distances are useful, but they have some drawbacks:

1. We cannot use them to compare P and Q when one is discrete and the other is continous.
2. These distances ignore the underlying geometry of the space.

Example



- $TV(P,Q) = \begin{cases} 1-p & \text{if } \Theta \neq 0 \\ 0 & \text{if } \Theta = 0 \end{cases}$
- $KL(P||Q) = \begin{cases} +\infty & \text{if } \Theta \neq 0 \\ 0 & \text{if } \Theta = 0 \end{cases}$



- $JS(P, Q) = (1 - p)\log 2$
- The 1-Wasserstein (Earth-Mover) distance depends on Θ !

$$W_1(P, Q) = \Theta(1 - p)$$

- Thus, the Wasserstein metric on probability spaces is sensitive to the "underlying" metric!

Wasserstein barycenters

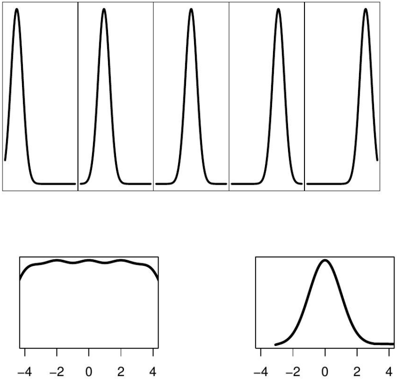
When we average different objects – such as distributions, data sets or images – we would like to make sure that we get back a similar objects. Suppose we have a set of distributions $P_1, P_2, \ldots, P_n$. How do we summarize these distributions with one "typical" distribution? We could take the average or Euclidean barycenter:

$$\frac{1}{n}\sum_{i=1}^{n} P_i.$$

A generalization of the average is the following. Let (X, d) be a metric space. The **barycenter** of the points $x_1, x_2, \ldots, x_n \in X$ is defined by

$$BC_d(x_1, x_2, \ldots, x_n) = \arg\min_{x} \frac{1}{n}\sum_{i=1}^{n} d^2(x, x_i).$$

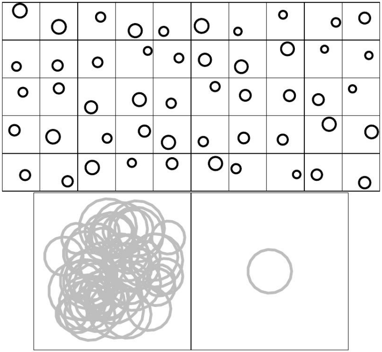
Example 1[2]



Top: Five distibutions. Bottom left: Euclidean average of the distributions. Bottom right: 1-Wasserstein barycenter.

[2] Kolouri et al. Optimal Mass Transport: Signal processing and machine-learning applications. IEEE Signal Processing Magazine 34(4) (2017):43–59.

Example 2[3]



Top: We take some random cirles and take a uniform distibution on each circle. Bottom left: Euclidean average of the distributions. Bottom right: 1-Wasserstein barycenter.

[3] Kolouri et al. Optimal Mass Transport: Signal processing and machine-learning applications. IEEE Signal Processing Magazine 34(4) (2017):43–59.

Basics of quantum optimal transport

- several different approaches:
  - Biane and Voiculescu (free probability)
  - Carlen and Maas (dynamical interpretation)
  - Golse, Mouhot, and Paul (static interpretation)
  - De Palma and Trevisan (quantum channels)
  - Życzkowski and Słomczyński (semi-classical approach)
- most relevant approaches for us are that of Golse-Mouhot-Paul[4] and De Palma-Trevisan[5]

[4] F. Golse, C. Mouhot and T. Paul, *On the mean-field and classical limits of quantum mechanics,* Commun. Math. Phys., **343** (2016), 165–205.

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

Classical vs Quantum: a dictionary

X, Y spaces (sand and pit)	$\leftrightarrow$	$\mathcal{H}, \mathcal{K}$ Hilbert spaces
$x \in X$	$\leftrightarrow$	$\|\psi\rangle \in \mathcal{H}$ ket vectors
$X \times Y$ product spaces	$\leftrightarrow$	$\mathcal{H} \otimes \mathcal{K}$ tensor product
$\mathcal{P}(X)$ prob. measures on X	$\leftrightarrow$	$\mathcal{S}(\mathcal{H})$ quantum state space
		(psd operators, with trace 1)
$\mu, \nu \in \mathcal{P}(X)$	$\leftrightarrow$	$\rho, \sigma \in \mathcal{S}(\mathcal{H})$ quantum states
δ_x Dirac measures	$\leftrightarrow$	$\|\psi\rangle\langle\psi\| \in \mathcal{S}(\mathcal{H})$ pure states
		(1-rank projections)

$\pi \in \mathcal{P}(X \times Y)$ joint distributions	$\leftrightarrow$	$\Pi \in \mathcal{S}(\mathcal{H} \otimes \mathcal{K})$ bipartite states
$p_i = \sum_j \pi(x_i, y_j), q_j = \sum_i \pi(x_i, y_j)$	$\leftrightarrow$	$\rho = \mathrm{Tr}_{\mathcal{K}} \Pi, \sigma = \mathrm{Tr}_{\mathcal{H}} \Pi$
marginal distributions		marginal states
$T : X \to Y$ transport map	$\leftrightarrow$	$\Phi : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{K})$ CPTP maps
		(quantum channels)

Basics of non-commutative optimal transport

- when measuring an observable quantity A on a quantum system being in the state $\rho \in$, the probability of the outcome lying in an interval $[a, b] \subset \mathbb{R}$ is $\mathrm{tr}_{\mathcal{H}} \left(\rho \, E_A \left([a, b] \right) \right)$, where E_A is the spectral measure of A
- a quantum state *encapsulates several classical probability distributions*, each corresponding to a physical quantity we are interested in
- let $A^{(1)}, \ldots, A^{(k)}$ be observable quantities, let us fix the initial state ρ_1 and the final state ρ_2
- let $X_i^{(j)}$ denote the random variable obtained by measuring $A^{(j)}$ in ρ_i, that is, $\mathbb{P} \left(X_i^{(j)} \in [a, b] \right) = \mathrm{tr}_{\mathcal{H}} \left(\rho_i \, E^{(j)} \left([a, b] \right) \right)$
- so the squared OT distance of the quantum states $\rho_1, \rho_2 \in$ should read as

$$D^2 \left(\rho_1, \rho_2 \right) = \inf_{\left(X_i^{(1)}, \ldots, X_i^{(k)} \right) \text{ is given by } \rho_i \, (i \in \{1,2\})} \left\{ \sum_{j=1}^{k} \mathrm{E} \left(X_1^{(j)} - X_2^{(j)} \right)^2 \right\}.$$

QOT via quantum couplings

The approach of Golse, Mouhot and Paul[6]

- quantum couplings are defined as

$$\mathcal{C}(\rho, \omega) = \{\pi \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H}) \,|\, \mathrm{tr}_2 \pi = \rho, \; \mathrm{tr}_1 \pi = \omega\},$$

- cost operators

$$C = \sum_{j=1}^{M} (A_j \otimes I - I \otimes A_j)^2$$

where $A_j \in \mathcal{L}^{sa}(\mathcal{H})$.

- optimal transport cost:

$$D_C^2(\rho, \omega) = \inf_{\pi \in \mathcal{C}(\rho, \omega)} \mathrm{tr}\, \pi C$$

[6] F. Golse, C. Mouhot and T. Paul, *On the mean-field and classical limits of quantum mechanics,* Commun. Math. Phys., **343** (2016), 165–205.

QOT via quantum channels

Recall: in the classical case, for $T : X \to Y$ satisfying $T_{\#}\mu = \nu$,

$$\pi_T := (Id \times T)_{\#}\mu \in \mathcal{P}(X \times Y) \in \Pi(\mu, \nu).$$

Purification

Given a state $\rho \in \mathcal{S}(\mathcal{H})$, a **purification** $\gamma \in \mathcal{S}(\mathcal{H} \otimes \mathcal{K})$ pure such that

$$\mathrm{Tr}_{\mathcal{K}}\gamma = \rho.$$

Canonical choice: $\mathcal{K} = \mathcal{H}^*$ and $\mathcal{H} \otimes \mathcal{H}^* \approx \mathcal{T}_2(\mathcal{H})$ by

$$\sum_{i,j} x_{ij}|i\rangle \otimes \langle j| \in \mathcal{H} \otimes \mathcal{H}^* \quad \longleftrightarrow \quad \sum_{i,j} x_{ij}|i\rangle\langle j| \in \mathcal{T}_2(\mathcal{H}).$$

$$\rho \in \mathcal{S}(\mathcal{H}) \mapsto ||\sqrt{\rho}\rangle\rangle \in \mathcal{H} \otimes \mathcal{H}^*$$

- Use spectral theorem to diagonalize

$$\rho = \sum_i p_i |i\rangle\langle i|$$

with ortonormal basis $(|i\rangle)_i$.

- Then $\sqrt{\rho} = \sum_i \sqrt{p_i}|i\rangle\langle i|$, hence

$$||\sqrt{\rho}\rangle\rangle = \sum_i \sqrt{p_i}|i\rangle \otimes \langle i|.$$

- Taking the partial traces we get

$$\begin{aligned} Tr_{\mathcal{H}^*}\left(||\sqrt{\rho}\rangle\rangle\,\langle\langle\sqrt{\rho}||\right) = \sum_i p_i|i\rangle\langle i| &= \rho \\ Tr_{\mathcal{H}}\left(||\sqrt{\rho}\rangle\rangle\,\langle\langle\sqrt{\rho}||\right) = \sum_i p_i\langle i|\otimes|i\rangle &= \rho^T. \end{aligned}$$

The approach of De Palma and Trevisan[7]

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

[7] 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 very strange thing

- $D_C(\rho,\sigma) = D_C(\sigma,\rho)$ $\surd$
- 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

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

which is the famous **the Wigner – Yanase information**!

- For any $\rho, \tau, \sigma \in \mathcal{S}(\mathcal{H})$ the modified triangle inequality holds:

$$D_C(\rho, \sigma) \leq D_C(\rho, \tau) + D_C(\tau, \tau) + D_C(\tau, \sigma)$$

  It is not known whether the term $D(\tau, \tau)$ can in fact be removed.

- conjecture (DPT): a modified version of the quantum optimal transport distance defined by

$$d_C(\rho, \omega) := \sqrt{D_C^2(\rho, \omega) - \frac{1}{2}\left(D_C^2(\rho, \rho) + D_C^2(\omega, \omega)\right)}$$

  is a true metric for all quadratic cost operator C up to some non-degeneracy assumptions on the A_j's generating C to ensure the definiteness of d_C, that is, that $d_C(\rho, \omega) = 0$ only if $\rho = \omega$

Our contribution

Triangle inequality for quantum Wasserstein divergences

Theorem (Bunth-Titkos-Virosztek-P. (2023))

The triangle inequality

$$d_C(\tau, \rho) + d_C(\rho, \omega) \geq d_C(\tau, \omega)$$

holds for any $\tau, \omega \in \mathcal{S}(\mathcal{H})$*, any* $\rho \in \mathcal{P}_1(\mathcal{H})$*, and any quadratic cost* C*.*

Our contribution[8]

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.

[8] 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

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

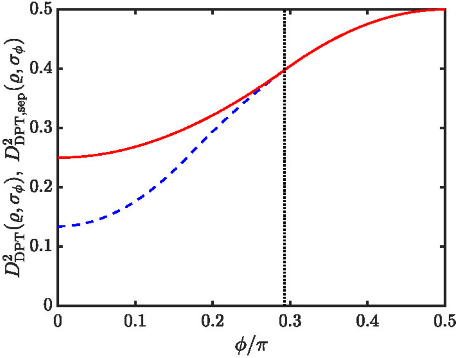
holds, then all optimal Π for $D(\rho, \sigma)$ is entangled.

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

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

and

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



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

The modified sef-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!