Optimal transport: classical and quantum

József Pitrik
Research Group on Quantum Information and Quantum Metrology

Work in progress with with Dániel Virosztek and Tamás Titkos, Rényi Institute, 'Momentum' Optimal Transport and Quantum Information Geometry Research Group

ELKH Eötvös Loránd Research Network

Why Optimal Transport (OT) theory?

OT has seen an increasing amount of attention from the applications:

- Signal and data analysis
- Machine learning
- Neural architecture search
- Image processing
- Modeling population dynamics in biology or social sciences
- Economics
- Weather and climate models
- **Quantum information theory!**
- etc.

The methods generated from OT theory are competitive with the current state-of-the-art methods!

OT pave the way towards a beautiful interplay between:

- partial differential equations
- fluid mechanics
- geometry
- probability theory
- functional analysis
- geometric measure theory, etc.

Very recently OT gained extreme popularity, because many researchers in different areas of mathematics understood that this topic was strongly linked to their subject.

Cédric Villani
(Fields Medal in 2010)

Alessio Figalli
(Fields Medal in 2018)

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

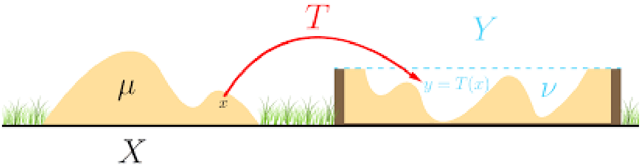
2. Wasserstein spaces
   - p-Wasserstein distance
   - Wasserstein barycenters
   - Wasserstein geodesics
   - Some modern approaches to the OT problem

3. Quantum optimal transport
   - Basics
   - "Quantum optimal transport is cheaper"

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

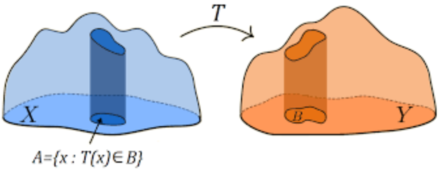


The classical optimal transport problem - Monge Formulation

- 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:

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



$\nu \in \mathcal{P}(Y)$ 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

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

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

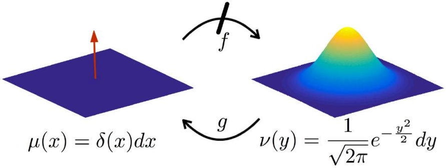
A transport map may not exist!

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 uniquness 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

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

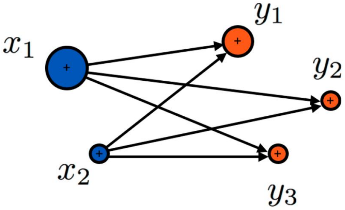
μ, ν are compactly supported:

- For $p > 1$, if μ, ν are absolutely continous with respec 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 respec 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 $B \subseteq Y$.

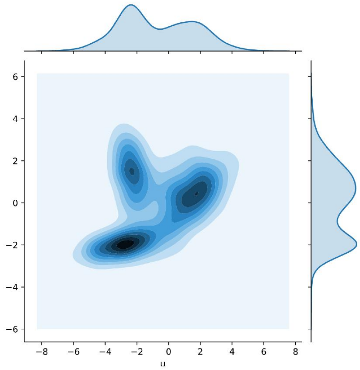
- The total mass sent from A is $\pi(A \times Y)$, and the total mass sent to B is $\pi(X \times B)$.
- π is mass-preserving iff

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

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

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

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

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

Probabilistic view:

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

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.”

Leonid Kantorovich 1912-1986 | Tjalling Koopmans 1910-1985

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

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

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)**

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

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)**

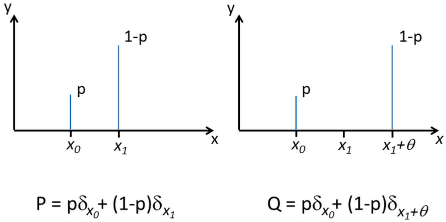
$$JS(P,Q) = KL(P||M) + KL(Q||M),$$

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

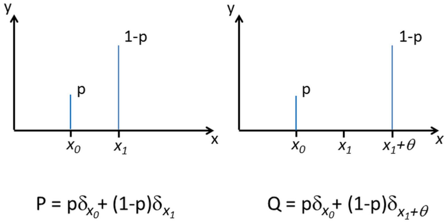
These distances are useful, but they have some drawbacks:

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 !

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

p-Wasserstein for 1D probability measures

- For absolutely continous probability measures μ and ν on $\mathbb{R}$ we can define their cumulative distribution functions

$$F_\mu(x) = \mu((-\infty, x)) \quad \text{and} \quad F_\nu(x) = \nu((-\infty, x))$$

- *p*-Wasserstein distance expressed by cumulative distribution functions:
  - Vallender:[2]

$$W_1(\mu, \nu) = \int_0^1 \left| F_\mu^{-1}(x) - F_\nu^{-1}(x) \right| \, \mathrm{d}x$$

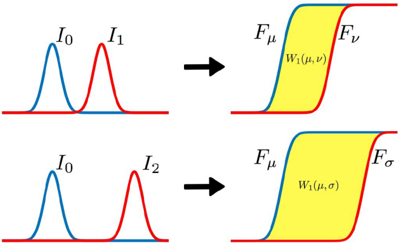
  - this can be generalized:[3]

$$W_p(\mu, \nu) = \left(\int_0^1 \left| F_\mu^{-1}(x) - F_\nu^{-1}(x) \right|^p \mathrm{d}x \right)^{\frac{1}{p}} \qquad (p > 1, \ \mu, \nu \in \mathcal{W}_p(\mathbb{R}))$$

[2] S. S. Vallender, *Calculation of the Wasserstein distance between probability distributions on the line,* Theory Probab. Appl. 18 (1973), 784–786.

[3] C. Villani, *Topics in optimal transportation,* Graduate studies in Mathematics vol. 58, American Mathemtical Society, Providence, RI, 2003.

$$W_p(\mu,\nu) = \left(\int_0^1 \left|F_\mu^{-1}(x) - F_\nu^{-1}(x)\right|^p \mathrm{d}x\right)^{\frac{1}{p}}$$



Note that the distances and divergences above do not provide a sensible distance between I_0, I_1 and I_2 while the p-Wasserstein distance does!

Wasserstein barycenters

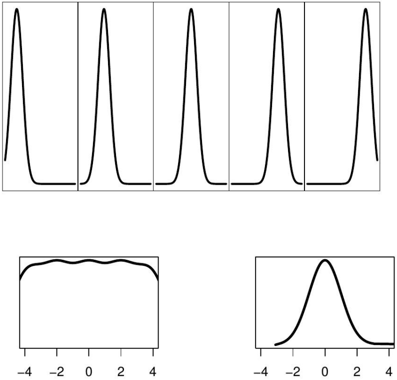
When we average different objects – such as distributions, data sets or images – we would like to make sure that we get back a similar objects. Suppose we have a set of distributions $P_1, P_2, \ldots, P_n$. How do we summarize these distributions with one "typical" distribution? We could take the average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

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

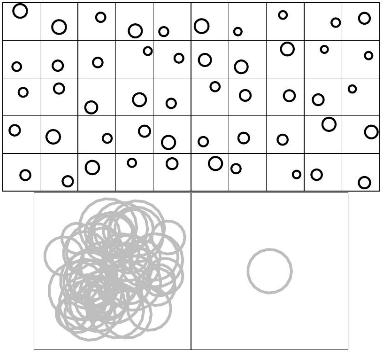
Example 1[4]



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

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

Example 2[5]



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.

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

2-Wasserstein geodesics

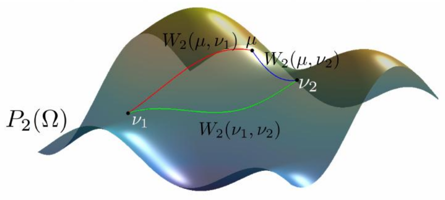
- The set of continous measures on Ω together with the 2-Wasserstein metric forms a Riemannian manifold, denoted by $\mathcal{P}_2(\Omega)$. (F. Otto, 2001.)
- Given the 2-Wasserstein space, the **geodesic** between μ and ν is the shortest curve on $\mathcal{P}_2(\Omega)$ that connects these measures.
- Let ρ_t for $t \in [0,1]$ parametrizes the geodesic curve on $\mathcal{W}_2(X)$ with $\rho_0 = \mu$ and $\rho_1 = \nu$.
- If T is the optimal transport map (it exist in this case!) we define

$$T_t(x) := (1-t)x + tT(x) \quad \text{(McCann interpolation)}$$

- Then the geodesic ρ_t is given by

$$\rho_t = (T_t)_{\#}\mu.$$

Recall that the push-forward measure is defined by $(T_{\#}\mu)(B) := \mu(T^{-1}(B)) = \nu(B)$ for all Borel B.



It is straightforward to show that the geodesic ρ_t is a constant speed geodesic, ie.

$$W_2(\mu, \rho_t) = tW_2(\mu, \nu).$$

Example [6]

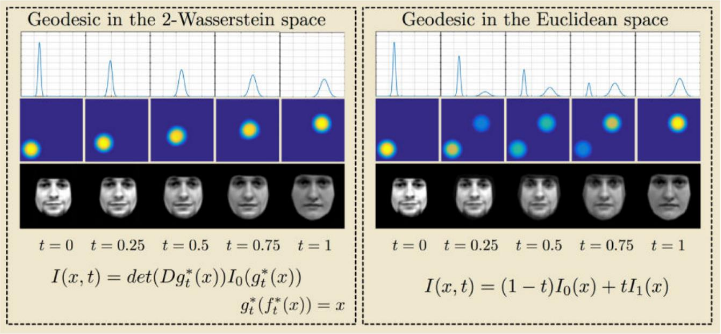


Fig. 2.
Geodesics in the 2-Wasserstein space (left panel), and in the Euclidean space (right panel) between various one and two-dimensional PDFs. Note that the geodesic in the 2-Wasserstein space captures the nonlinear structure of the signals and images and provides a natural morphing.

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

Dynamical interpretation – Benamou-Brenier formula

The problem goes back to the fluid mechanics: we like to model an incompressible, inviscid fluid in a bounded, smooth open set $\Omega \subset \mathbb{R}^n$ $(n = 2, 3)$. $p = p(t, x) \in \mathbb{R}$ is the pressure of the fluid at time t and position x, the unknown is the velocity field of the fluid:

$$v(t, x) : \mathbb{R}^+ \times \Omega \to \mathbb{R}^n \quad \text{(=tangent space to } \Omega\text{)}$$

The incompressible **Euler equation**:

$$\begin{cases} \frac{\partial v}{\partial t} + v \cdot \nabla v = -\nabla p & \text{(Euler equation)} \\ \nabla \cdot v = 0 & \text{(incompressibility condition)} \\ v \cdot \nu = 0 \text{ on } \partial\Omega & \text{(no flux condition).} \end{cases}$$

V.I. Arnold (1966) showed that the Euler eqation can be represented as geodesic equations on the (infinite-dimensional) group of diffeomorphisms equipped with a certain Riemannian metric.

Benamou and Brenier (2000)

Let $\mu_0, \mu_1 \in \mathcal{W}_2(\mathbb{R}^d)$, and let T be the optimal transport map between μ_0 and μ_1.

$$T_t(x) = (1-t)x + tT(x), \quad x \in \mathbb{R}^d$$

is the McCann interpolation, for which $T_0(x) = x$ and $T_1(x) = T(x)$. Then the geodesic curve in $\mathcal{W}_2(\mathbb{R}^d)$ between μ_0 and μ_1 is given by:

$$\mu_t := \mu_0 \circ T^{-1} = \rho_t \mathrm{d}x, \quad t \in [0,1],$$

where $\mu_0 = \rho_0 \mathrm{d}x$. If we define the velocity field by

$$v_t : \mathbb{R}^d \to \mathbb{R}^d, \quad \frac{\mathrm{d}T_t}{\mathrm{d}t} = v_t(T_t),$$

then μ_t satisfies the classical continuity equation:

$$\frac{\partial \rho_t}{\partial t} + \nabla(\rho_t \cdot v_t) = 0.$$

Benamou and Brenier defined for a given pair (ρ_t, v_t) solving the continuity equation with the tangentiality of v_t on the boundary the total kinetic energy (or action) by

$$A[\rho_t, v_t] := \int_0^1 \int_\Omega \|v_t(x)\|^2 \rho_t(x) \mathrm{d}x \mathrm{d}t,$$

and showed that

$$W_2^2(\mu_0, \mu_1) = \inf\{A[\rho_t, v_t] : \rho_0 = \mu_0, \rho_1 = \mu_1, \partial_t \rho_t + \nabla(v_t \rho_t) = 0\}.$$

In the fluid dynamical interpretation $\rho_t(x)$ stands for the density of particles.

Gradient flows – Otto calculus

Gradient flows are evolutionary systems driven by a potential (energy), in the sense that the energy decreases along solutions, *as fast as possible*. The two ingredients of the problem are:

- the driving energy
- *"as fast as possible"* $\Longrightarrow$ the dissipation mechanism

Example

A curve $x : [0, T] \to \mathbb{R}^n$ is the gradient flow of a potential $E : \mathbb{R}^n \to \mathbb{R}$ starting at $x_0 \in \mathbb{R}^n$ if

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} x(t) = -\nabla E(x(t)), \\ x(0) = x_0. \end{cases}$$

Note that, for a solution $x(t)$ of the gradient flow

$$\frac{\mathrm{d}}{\mathrm{d}t}E(x(t)) = \nabla E(x(t))\dot{x}(t) = -\|\nabla E(x(t))\|^2 \leq 0,$$

thus

- E decreases along the curve $x(t)$
- $\frac{\mathrm{d}}{\mathrm{d}t}E(x(t)) = 0$ iff $\|\nabla E(x(t))\| = 0$ i.e. $x(t)$ is a critical point a E
- convexity of the energy E determines *stability* and *long time behavior*

If we like to generalize this concept from $\mathbb{R}^n$ to more exciting spaces X, to define a gradient flow we need gradients (tangent plane) and scalar product (that is we have to define the dissipation mechanism).

Example

If $X = L^2(\mathbb{R}^n)$, we define the operator ∇ by

$$\langle \nabla E(f), g \rangle = \lim_{t \to 0} \frac{E(f + tg) - E(f)}{t}, \quad f, g \in L^2(\mathbb{R}^n).$$

With the choice of

$$E(f) = \frac{1}{2} \int_{\mathbb{R}^n} \|\nabla f\|^2 \mathrm{d}x$$

called **Dirichlet energy functional**, the gradient flow is the **heat equation**

$$\frac{\partial}{\partial t} f(x, t) = \triangle f(x, t).$$

F. Otto at al. discovered that by replacing the Dirichlet energy functional with the **entropy functional** $\int f \log f$, and the L^2 norm with the 2-Wasserstein distance, the 2-Wasserstein gradient flow is again the heat equation.

Gradient flows in the 2-Wasserstein space

It turns out that many PDE's of mathematical physics admit desciptions in the form of gradient flows of some energy functional E on $\mathcal{W}_2$. Moreover, all Wasserstein gradient flows are of the form

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \quad \text{with} \quad v = -\nabla \frac{\partial E}{\partial \rho}.$$

energy functional	gradient flow	
$E(\rho) = \int \rho \log \rho$	$\frac{d}{dt}\rho = \Delta \rho$	$v = -\frac{\nabla \rho}{\rho}$
$E(\rho) = \frac{1}{m-1} \int \rho^m$	$\frac{d}{dt}\rho = \Delta \rho^m$	$v = -m\rho^{m-2}\nabla \rho$
$E(\rho) = \int V\rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla V \rho)$	$v = -\nabla V$
$E(\rho) = \int (K * \rho)\rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla (K * \rho)\rho)$	$v = -\nabla (K * \rho)$

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[7] and De Palma-Trevisan[8]

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

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

Basics of quantum optimal transport

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

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

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

Basics of quantum optimal transport

The approach of De Palma and Trevisan[9]

- 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

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

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

Basics of quantum optimal transport

- 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.

Basics of quantum optimal transport

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

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

- The transport cost for a coupling Π is

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

- The **quantum Wasserstein 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

- $D_C(\rho, \sigma) = D_C(\sigma, \rho)$ $\checkmark$
- If $\rho = \sigma$ then the optimal transport map corresponds to the identity, 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

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

  the Wigner – Yanase information, i.e. there is some deep connection with the **quantum Fisher information!**
- For any $\rho, \tau\sigma \in \mathcal{S}(\mathcal{H})$ the modified triangle inequality holds:

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

"Quantum optimal transport is cheaper"

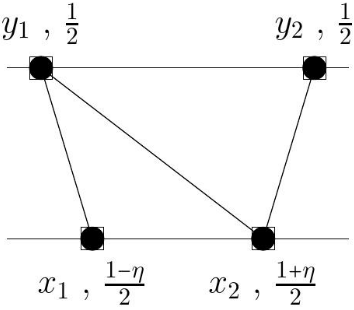
- the following example is taken from Caglioti-Golse-Paul[10]
- the setting: $\mathcal{H} = L^2\left(\mathbb{R}^d\right)$, the cost is defined by position and momentum:

$$C = (\hat{p} \otimes I - I \otimes \hat{p})^2 + (\hat{q} \otimes I - I \otimes \hat{q})^2 - 2d\hbar$$

$$= (x-y)^2 - \hbar^2 (\nabla_x - \nabla_y)^2 - 2d\hbar = -4\hbar^2 \nabla^2_{x-y} + (x-y)^2 - 2d\hbar$$

- $\frac{1}{2}(C + 2d\hbar)$ is the Hamiltonian of the quantum harmonic oscillator in the variable $(x-y)/\sqrt{2}$ and hence $C \geq 0$

[10] E. Caglioti, F. Golse, T. Paul, *Quantum optimal transport is cheaper*, J. Stat. Phys. **181** (2020), 149–162.



- let $d = 1$ and consider the classical OT problem with $\mu = \frac{1+\eta}{2}\delta_a + \frac{1-\eta}{2}\delta_{-a}$ and $\nu = \frac{1}{2}\delta_a + \frac{1}{2}\delta_{-a}$ where $\eta > 0$ so $\eta/2$ is transported from a to $-a$, and the quadratic cost is $2\eta a^2$

"Quantum optimal transport is cheaper"

- the "quantized classical" coupling is $Q_c =$

$$= \frac{1}{2} |a\rangle \langle a| \otimes |a\rangle \langle a| + \frac{1-\eta}{2} |-a\rangle \langle -a| \otimes |-a\rangle \langle -a| + \frac{\eta}{2} |a\rangle \langle a| \otimes |-a\rangle \langle -a|$$

where $|a\rangle$ is a is a coherent state of null momentum localized at a, i.e., $\langle x| \, |a\rangle = (\pi\hbar)^{-\frac{1}{4}} e^{-\frac{(x-a)^2}{2\hbar}}$, with marginals

$$\mathrm{tr}_2 Q_c =: R = \frac{1+\eta}{2} |a\rangle \langle a| + \frac{1-\eta}{2} |-a\rangle \langle -a|$$

and

$$\mathrm{tr}_1 Q_c =: S = \frac{1+\lambda}{2} |\phi_+\rangle \langle \phi_+| + \frac{1-\lambda}{2} |\phi_-\rangle \langle \phi_-|$$

where $\phi_\pm = \frac{|a\rangle \pm |-a\rangle}{\sqrt{2(1\pm\lambda)}}$ and $\lambda = \langle a| \, |-a\rangle = e^{-a^2/\hbar}$

"Quantum optimal transport is cheaper"

- now let
$$Q_q := \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$
in the basis $\{\phi_\pm \otimes \phi_\pm\}$
- clearly, $\mathrm{tr}_2 Q_q = \mathrm{tr}_1 Q_q = \mathrm{tr} Q_q = 0$
- therefore, $Q_\varepsilon := Q_c + \varepsilon Q_q$ is a **coupling** of R and S (checking the positivity is tricky) for $0 < \varepsilon << 1$
- and $\mathrm{tr}\, CQ_q = -\frac{8a^2\lambda^2}{1-\lambda^2} < 0$, hence

$$\mathrm{tr}\, CQ_\varepsilon = \mathrm{tr}\, CQ_0 - \varepsilon \frac{8a^2\lambda^2}{1-\lambda^2} < \mathrm{tr}\, CQ_0 = d^2_{W_2}(\mu, \nu)$$

Future plan

- Understand the intimate connections between quantum Wasserstein distances and Fisher information metrics.
- Quantum Wasserstein geodesics
- Quantum Wasserstein barycenters
- Describe the isometric structure of p-Wasserstein spaces in some important cases.

Thank you for your kind attention!