

# Lectures on Optimal Transport. May, 2022. KAIST

Young-Heon Kim (Department of Mathematics, University of British Columbia)

Lecture 1 The Monge-Kantorovich problem and duality.

Lecture 2 Wasserstein geometry of the space of probability measures.

Lecture 3 Entropic regularization of optimal transport.

Lecture 4 Application of optimal transport to developmental processes.

**Lecture 5** Multimarginal optimal transport. Wasserstein barycentre. **Today!**

Seminar Optimal Brownian stopping with free target and the supercooled Stefan problem.

Lecture 6 Optimal martingale transport

Lecture 7 Optimal Brownian martingale transport

# Some references for the lectures

- ▶ Lecture 1, 2, and 3:
  - ▶ Villani: Topics in Optimal Transport. Book
  - ▶ Villani: Optimal Transport. Old and New. Book
  - ▶ Cuturi & Payré: Computational Optimal Transport. Book
- ▶ Lecture 4
  - ▶ Schiebinger: <https://broadinstitute.github.io/wot/tutorial/>
  - ▶ Lavanant, Zhang, Kim, Schiebinger,: Towards a mathematical theory of trajectory inference. <https://arxiv.org/abs/2102.09204>
- ▶ **Lecture 5**
  - ▶ Cuturi & Payré: Computational Optimal Transport. Book
  - ▶ Kim & Pass: Wasserstein Barycenters over Riemannian manifolds. Adv. in Math. 2017.
- ▶ Lecture 6
  - ▶ Ghoussoub, Kim, & Lim: Structure of optimal martingale transport in general dimensions. Ann. Prob. 2019.
- ▶ Lecture 7
  - ▶ Ghoussoub, Kim, & Palmer: PDE Methods For Optimal Skorokhod Embeddings. Calc. Var. 2019.
  - ▶ Ghoussoub, Kim, & Palmer: A solution to the Monge transport problem for Brownian martingales. Ann. Prob. 2021.
  - ▶ I. Kim & Y. Kim. The Stefan problem and free targets of optimal Brownian martingale transport. Preprint. 2021

## **Geometric average of shapes**

# Probability measures

- ▶ Randomness

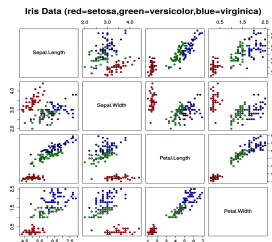


- ▶ Distributions

- ▶ distribution of mass, temperature, etc.
- ▶ Images



- ▶ Datasets



# Barycentre

- ▶  $X$  = a metric space with distance *dist*.  
e.g.  $X = \mathbb{R}^n$ ,  $\text{dist}(x, y) = |x - y|$ .
- ▶  $\mu$  = a probability measure on  $X$ .



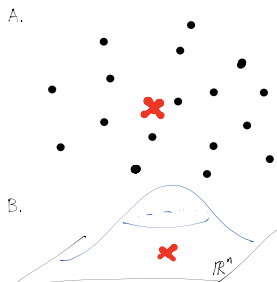
Barycenter:

$$\text{Bar}(\mu) \in \underset{x \in X}{\operatorname{argmin}} \int \text{dist}^2(x, y) d\mu(y)$$

- ▶  $\operatorname{argmin}$  = the set of minimizers
- ▶  $\int \dots d\mu$  is the same as  $\sum_i \dots \mu_i$  in the finite (discrete) set case.

# Barycentre

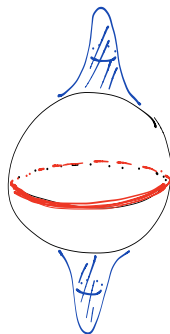
- ▶ Centre of mass
- ▶ **Geometric average** of many points in the distribution
- ▶ A central point of a distribution
- ▶ **Example:** Where would they put the central postal office in Daejeon?



- ▶ Unique on the Euclidean space  $X = \mathbb{R}^n$   
(since  $\text{dist}^2(x, y) = |x - y|^2$  is uniformly convex)

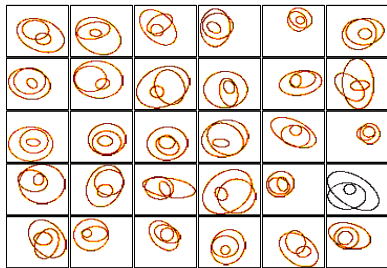
# Barycentre

May not be unique in general:  
e.g. Consider distributions on the round sphere.

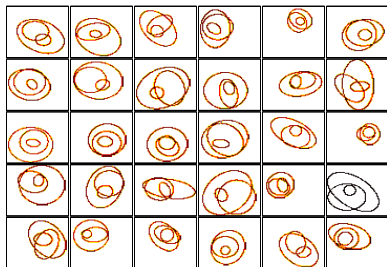


## An infinite dimensional problem: Average of many shapes

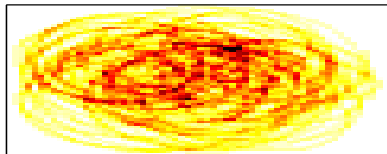
Consider a set of “shapes”.  
What is the **average** of those shapes?







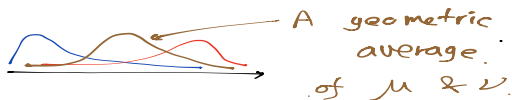
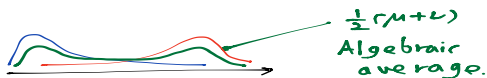
Algebraic average of the shapes



(Image provided by Marco Cuturi)

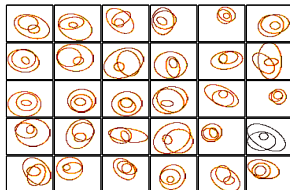
Question: **Geometric average** of shapes?

# Comparison between algebraic average and geometric average



## Question: **Geometric average** of shapes?

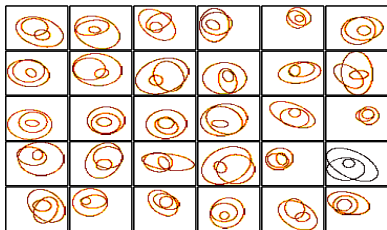
- ▶  $X$  = the space of shapes.
  - ▶ Very large or infinite dimensional space.
- ▶ Consider **a notion of distance** on  $X$ :
  - ▶ which shape is close to which, and how much.
- ▶ Population (distribution) of shapes:



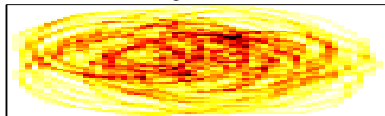
- ▶ probability distribution over  $X$ .
- ▶ Consider a barycenter of such **population** over  $X$ .
- ▶ The barycenter itself is a shape, which we can regard as the geometric average.

# GEOMETRIC average of shapes.

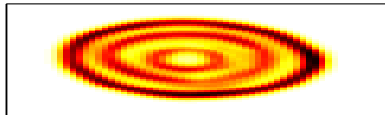
- ▶ Data/image interpolation:  
e.g. Recover the original image from many distorted samples.
- ▶ E.g. What is the 'average' shape of healthy heart, that one can compare a defective one with it?
- ▶ ETC.



algebraic



geometric



(Image provided by Marco Cuturi)

**Geometric averages via optimal transport: Wasserstein barycenters.**

# Optimal transport with cost function $c(x, y) = \text{dist}^2(x, y)$ and Wasserstein barycenters

Optimal transport  $\pi$  between  $\mu$  and  $\nu$  is a transport plan ( $\pi \in \Pi(\mu, \nu)$ ) that minimizes

$$\int_X \int_X \text{dist}^2(x, y) d\pi(x, y).$$

- ▶ In discrete setting:

$$\text{Minimize } \sum_i \sum_j \text{dist}(x_i, y_j)^2 \pi_{ij}$$

subject to  $\pi \in \Pi(\mu, \nu)$ .

- ▶ A **distance** between probability measures

$$d_W(\mu, \nu) = \sqrt{\min_{\pi \in \Pi(\mu, \nu)} \int_X \int_X \text{dist}^2(x, y) d\pi(x, y)}$$

called **the Wasserstein distance**.

- ▶  $P(X)$  = "the space of probability measures on  $X$ ", becomes a natural **metric space** with  $d_W$ :
  - ▶ Isometric imbedding  $X \ni x \mapsto \delta_x \in P(X)$ .  $d_W(\delta_x, \delta_y) = \text{dist}(x, y)$ .

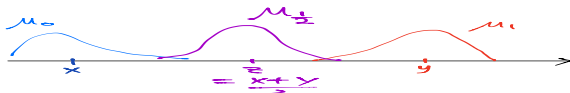
# An optimal transport way of averaging: displacement interpolation. [McCann]

$$I_{1/2} : (x, y) \mapsto \frac{x+y}{2}$$

For optimal plan  $\pi = \operatorname{argmin}_{\pi \in \Pi(\mu, \nu)} \int \operatorname{dist}^2(x, y) d\pi(x, y)$ ,

displacement interpolation

$$\mu_{1/2} = (I_{1/2})_{\#} \pi$$



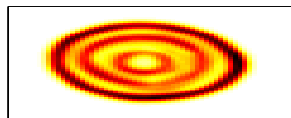
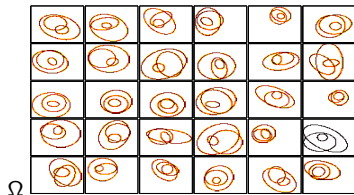
# An optimal transport way of averaging: Wasserstein barycentre

$$(X, \text{dist}) \implies (P(X), d_W).$$

$\Omega$  a probability measure on  $P(X)$  (i.e.  $\Omega \in P(P(X))$ .)

$$\text{Bar}_W(\Omega) \in \operatorname{argmin}_{\mu \in P(X)} \int_{P(X)} d_W^2(\mu, \nu) d\Omega(\nu).$$

- Wasserstein barycenter is a probability measure on  $X$ .



$\text{Bar}_W(\Omega)$

(Image provided by Marco Cuturi)



## Examples of Wasserstein barycenter

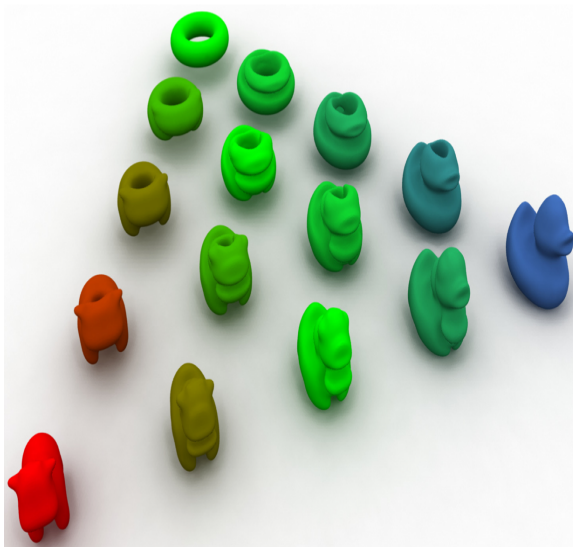


Image provided by Marco Cuturi

# Wasserstein barycentre

$\Omega$  is a distribution on the space of probability measures.

- ▶ [McCann].  $\Omega = (1 - \lambda)\delta_{\nu_0} + \lambda\delta_{\nu_1}$ .  
 $Bar_W(\Omega)$  is **displacement interpolation**.
- ▶ [Agueh & Carlier]. **Wasserstein barycenter**.  
 $\Omega = \sum_i \lambda_i \delta_{\nu_i}$ , *finitely many measures  $\nu_i$* ,  
 $X = \mathbb{R}^n$  Euclidean space.
- ▶ [K. & Pass]. General  $\Omega \in P(P(X))$ .  
 $X =$  **Riemannian manifold**.
  - ▶ Riemannian manifold: a curved space that generalizes the Euclidean space.

**It is related to multimarginal optimal transport**

# Multimarginal Optimal Transport

- ▶  $\mu_1, \dots, \mu_N$  probability measures on  $X_1, \dots, X_N$ , respectively.
- ▶  $\pi \in \Pi(\mu_1, \dots, \mu_N)$ .  
Transport plan with multiple marginals  $\mu_1, \dots, \mu_N$ .  
 $\pi$  is a probability measure on  $X_1 \times \dots \times X_N$ .
- ▶ Cost function  $c(x_1, \dots, x_N)$ , on  $X_1 \times \dots \times X_N$ .
- ▶ Multimarginal Optimal Transport:

$$\min_{\pi \in \Pi(\mu_1, \dots, \mu_N)} \int_{X_1 \times \dots \times X_N} c(x_1, \dots, x_N) d\pi(x_1, \dots, x_N).$$

- ▶ Motivation:
  - ▶ **[Economics]** find a most cost-effective matching between resources/agents:
    - ▶ To make a good faculty of science, we need to hire (match) mathematicians, statisticians, physicists, chemists, biologists, computer scientists, etc, etc.
  - ▶ From the previous lecture,  $\pi$  is like a matching between cell distributions  $\mu_i$ 's at different time points  $t_1, t_2, \dots, t_N$ .
  - ▶ **[Physics] density functional theory**: describe the state of many particles that are correlated with each other.
  - ▶ **[Mathematics]** What is the **distance** between measures? What is the **barycenter** between several measures?

## Theorem (Gangbo & Swiech, Agueh & Carlier, K. & Pass)

### Assume

- ▶  $X = a$  Riemannian manifold
- ▶  $c(x_1, \dots, x_N) = \inf_{y \in X} \sum_i \lambda_i \text{dist}^2(y, x_i)$ .
- ▶  $\mu_1 \ll \text{vol}$ .

### Then

- ▶ the optimal solution  $\pi^*$  of the multimarginal OT uniquely exists,
- ▶  $\pi^*$  is supported on a graph of a function  $F : x_1 \mapsto (F_2(x_1), \dots, F_N(x_1))$ ,
- ▶ and  $C_{\#}^{\bar{\lambda}} \pi^*$  is the unique Wasserstein barycenter of  $\frac{1}{N} \sum_i \lambda_i \delta_{\mu_i} \in P(P(X))$ 
  - ▶ for  $C^{\bar{\lambda}}(x_1, \dots, x_N)$  is the Riemannian center of mass of  $\sum_i \lambda \delta_{x_i}$ , defined for  $\pi^*$ -a.e.  $(x_1, \dots, x_N)$ .

**Remark:** Similar results by [K. & Pass] under more general but structurally similar conditions on  $c$ .

## Other types of costs for multi-marginal problems:

- ▶  $c(x_1, x_2, \dots, x_N) = \sum_{i \neq j}^N \frac{1}{|x_i - x_j|}$ ,
  - ▶ arises in density functional theory for Coulombic electronic interactions in quantum physics
  - ▶ (Cotar-Friesecke-Kluppelberg '11 and Buttazzo-De Pascale-Gori-Giorgi '12).
- ▶  $c(x_1, x_2, \dots, x_N) = - \sum_{i \neq j}^N |x_i - x_j|^2$ ,
  - ▶ arises when Coulombic interactions are replaced by repulsive, harmonic oscillator interactions.

These are costs for which solutions can be **non-unique** and have **high dimensional support**.

## **Properties of Wasserstein barycentres**

## Theoretical results: Properties of $\text{Bar}_W$ .

$X$  = Riemannian manifold.

$\text{vol}$  = Riemannian volume; it is a version of Lebesgue measure in the Riemannian space.

- ▶ **Unique** if  $\Omega[\{\nu \in P(X) \mid \nu \ll \text{vol}\}] > 0$ .
- ▶ **[Absolute continuity]**  
 $\text{Bar}_W(\Omega) \ll \text{vol}$  if  $\Omega[\{\nu \in P(X) \mid \nu \ll \text{vol}\}] > 0$ .  
So,  $\text{Bar}_W(\Omega)$  has a density function.
- ▶ These are proven in the following cases respectively.
  - ▶ [McCann].  $\Omega = (1 - \lambda)\delta_{\nu_0} + \lambda\delta_{\nu_1}$ .  
 $\text{Bar}_W(\Omega)$  is **displacement interpolation**.  
when  $\nu_0$  or  $\nu_1$  is  $\ll \text{Leb}$ .
  - ▶ [Agueh & Carlier]. **Wasserstein barycenter**.  
 $\Omega = \sum_i \lambda_i \delta_{\nu_i}$ , **finitely many measures  $\nu_i$** ,  
 $X = \mathbb{R}^n$  Euclidean space  
and  $\nu_1 \ll \text{Leb}$  (and  $\nu_j$ 's are compactly supported)
  - ▶ [K. & Pass]. **General**  $\Omega \in P(P(X))$ .  
 $X$  = **Riemannian manifold**  
and  $\Omega[\{\nu \in P(X) \mid \nu \ll \text{vol}\}] > 0$ ,  
(and  $\nu_j$ 's are compactly supported)



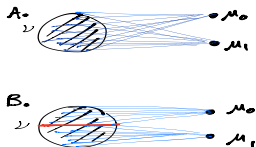
# Uniqueness from (linear) strict convexity

- If  $\nu \ll \text{vol}$ , then  $\mu \mapsto d_W^2(\mu, \nu)$  is strictly convex:

$$d_W^2(\underbrace{(1-t)\mu_0 + t\mu_1}_{\text{linear interpolation}}, \nu) < (1-t)d_W^2(\mu_0, \nu) + td_W^2(\mu_1, \nu),$$

unless  $\mu_0 = \mu_1$ .

Example:



- If  $\Omega[\{\nu \in P(X) \mid \nu \ll \text{vol}\}] > 0$

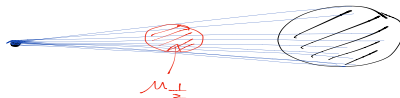
then  $\mu \mapsto \int_{P(X)} d_W^2(\mu, \nu) d\Omega(\nu)$  is strictly convex.

# Absolute continuity from non-concentration phenomena

## Example

$$\mu = \delta_x, \quad d\nu(y) = f(y)d\text{vol}(y)$$

$$d\mu_{1/2}\left(\frac{x+y}{2}\right) = 2^n f(y) d\text{vol}$$



## Absolute continuity. When $X = \mathbb{R}^n$ .

- ▶ Let  $\bar{\mu}$  be the Wasserstein barycenter of  $\nu_1, \dots, \nu_N$  over  $\mathbb{R}^n$ .
- ▶ For  $C^{\vec{\lambda}}(x_1, \dots, x_N) = \operatorname{argmin}_x \sum_i \lambda_i |x - x_i|^2$ ,  
 $C^{\vec{\lambda}}(x_1, \dots, x_N) = \sum \lambda_i x_i$ .
- ▶ For its inverse  $(C^{\vec{\lambda}})^{-1} : \operatorname{supp} \bar{\mu} \rightarrow \operatorname{supp} \pi^*$   
(where  $\pi^*$  the optimal solution of the multimarginal problem),  
we have

$$\|(C^{\vec{\lambda}})^{-1}\|_{Lip} \leq \frac{1}{\lambda_1} \quad \text{on the support of } \bar{\mu}.$$

- ▶ This implies

$$\left\| \frac{d\bar{\mu}}{dx} \right\|_{\infty} \leq \frac{1}{\lambda^n} \left\| \frac{d\nu_1}{dx} \right\|_{\infty}.$$

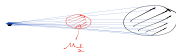
- ▶ For general  $\Omega$  over  $P(\mathbb{R}^d)$ , take approximation of  $\Omega$  by finitely supported measures  $\Omega^M = \frac{1}{N} \sum_i \delta_{\nu_i^N}$  and let  $N \rightarrow \infty$ .

# Absolute continuity: $X = \text{general (compact) Riemannian case.}$

**Assume for simplicity**  $\lambda_1 = \dots = \lambda_N = 1/N$ .

- ▶ When  $N = 2$ :

$$\|Bar_W(\nu_1, \delta_x)\|_\infty \leq C \|\nu_1\|_\infty$$



where

$C = C(\text{Ricci curvature of } X, \text{diam}, \text{dim})$ .

- ▶ For  $\nu_1, \nu_2 \in P(X)$ , approximate  $\nu_2$  by  $\frac{1}{k} \sum_{i=1}^k \delta_{x_i}$ ,

$$\left\| Bar_W(\nu_1, \frac{1}{k} \sum_{i=1}^k \delta_{x_i}) \right\|_\infty \leq C \|\nu_1\|_\infty$$

- ▶ Take limit.

- ▶ When  $N \geq 3$ ,

- ▶ Approximate  $\nu_i, i = 2, \dots, N$ , by  $\hat{\nu}_i = \frac{1}{k_i} \sum_j^{k_j} \delta_{y_j^i}$ .

- ▶ Get

$$\|Bar_W(\nu_1, \hat{\nu}_2, \dots, \hat{\nu}_N)\|_\infty \leq C \|\nu_1\|_\infty$$

- ▶ This step is not so trivial.
- ▶ Take limit.

- ▶ For general  $\Omega$  over  $P(X)$  (so  $\Omega \in P(P(X))$ ), take approximation of  $\Omega$  by finitely supported measures  $\Omega^N = \frac{1}{N} \sum_i \delta_{\nu_i^N}$  and let  $N \rightarrow \infty$ .

## Further properties of $\text{Bar}_W$ ?

- ▶ It is a big challenge to draw geometric/analytic information of the Wasserstein barycenter.
  - ▶ e.g. Smoothness properties of the density?

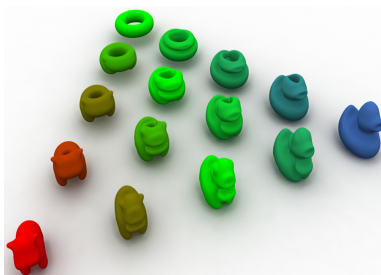


Image provided by Marco Cuturi

## Tools from PDE: Moge-Ampère equation:

For two measures:

- ▶  $\Omega = (1 - \lambda)\delta_{\nu_0} + \lambda\delta_{\nu_1}$ ,  $\nu_i = \rho_i \text{Leb}$ .
- ▶ Understanding the **displacement interpolation**  $Bar_W(\Omega)$  can be reduced to studying the solution  $u$  to the **Monge-Ampère equation**:
- ▶  $u$  is a convex function with

$$\rho_1(\nabla u(x)) \det(\nabla^2 u(x)) = \rho_0(x).$$

▶

Let  $\bar{\rho} =$  density of  $Bar_W(\Omega)$ .

Then,

$$\bar{\rho}((1 - \lambda)x + \lambda \nabla u(x)) \det \left( (1 - \lambda)I + \lambda \nabla^2 u(x) \right) = \rho_0(x).$$

# Regularity of optimal transport

Smooth  $u$  if things are good (convexity, positivity, good curvature, etc).

- ▶ Delanoë, Caffarelli, Urbas,...
- ▶ More general costs or on Riemannian manifolds
  - ▶ Ma, Trudinger, Wang
  - ▶ Loeper,
  - ▶ Figalli, K. & McCann
  - ▶ ...

Nonsmooth only in a small set ("Partial Regularity")

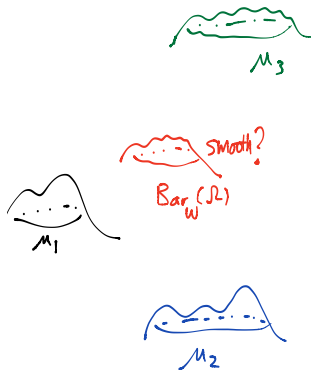
- ▶ .....
- ▶ Figalli & K.
- ▶ DePhilippis & Figalli
- ▶ ...

## More than two measures?

- ▶  $\mu_1, \mu_2, \mu_3$  smooth positive probability measures on  $\mathbb{R}^n$ .

$$\Omega = \frac{1}{3}[\delta_{\mu_1} + \delta_{\mu_2} + \delta_{\mu_3}] \in P(P(\mathbb{R}^n)).$$

Is  $\text{Bar}_W(\Omega)$  smooth on  $\mathbb{R}^n$ ?





## More than two measures?

### Future direction:

System of Monge-Ampere

$$\rho_i(\nabla u_i(x)) \det(\nabla^2 u_i(x)) = \bar{\rho}(x), \quad i = 1, \dots, k.$$

with balance condition:

$$\sum_i \nabla u_i(x) = 0 \quad \text{for } \bar{\rho}\text{-a.e. } x.$$