Lectures on Optimal Transport. May, 2022. KAIST

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

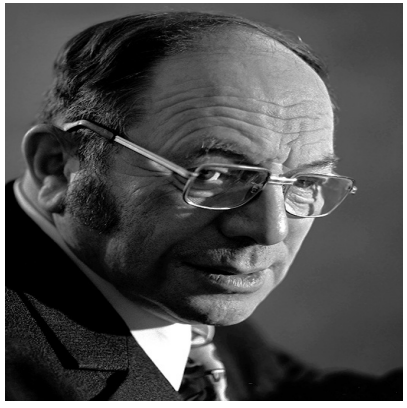
Lecture 1	The Monge-Kantorovich problem and duality. **Today!**
Lecture 2	Wasserstein geometry of the space of probability measures.
Lecture 3	Entropic regularization of optimal transport.
Seminar	Optimal Brownian stopping with free target and the supercooled Stefan problem.
Lecture 4	Application of optimal transport to developmental processes.
Lecture 5	Multimarginal optimal transport. Wasserstein barycentre.
Lecture 6	Optimal marginale 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/
  - Kim, Lavenant, Schiebinger, Zhang: 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

Optimal transport

Gaspard Monge
(in 19th century)



Leonid Kantorovich
(1975 Nobel prize in Economics)

Brenier, Caffarelli, McCann, Otto, **Villlani (2008 Fields Medal)**, Ambrosio, Trudinger, Wang, ... ,
Figalli (2018 Fields Medal), and many many others are working in this area.

Optimal transport considers certain optimization problems for measures.

Many connections and applications within and outside mathematics

In Mathematics:

- analysis/geometry/probability:
  - Ricci curvature, Ricci Flow
  - Analysis on singular spaces (metric-measure spaces)
  - geometric inequalities: isoperimetric, log-Sobolev
  - convex geometry
  - $\cdots$
- PDE
  - Monge-Ampère equation and geometric optics
  - nonlinear diffusion (porus-medium equation, $\cdots$)
  - fluid mechanics (Euler equation, $\cdots$)
  - free boundaries; e.g. the Stefan problem (of ice melting or water freezing). $\cdots$)
- $\cdots$.

Villani: *"In my opinion this abundance only reflects the fact that optimal transport is a simple, meaningful, natural and therefore universal concept."*

Practical applications of Optimal Transport

- Image processing

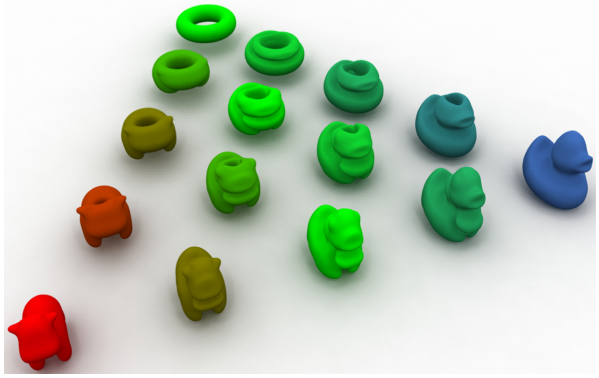
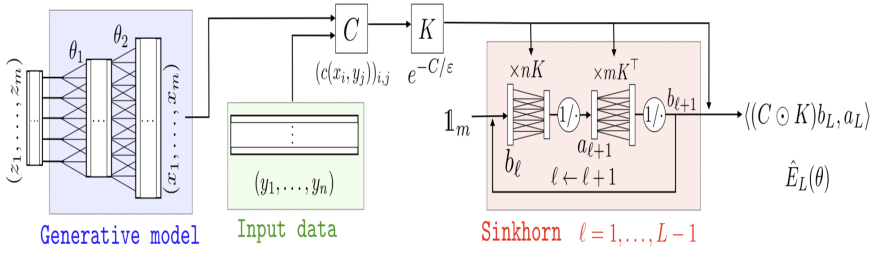


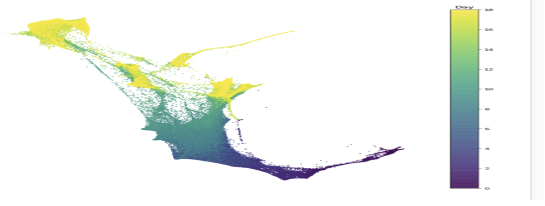
Image provided by Marco Cuturi

- Machine learning



from [Genevay-Peyr'e-Cuturi 2017]

- Stem cell research

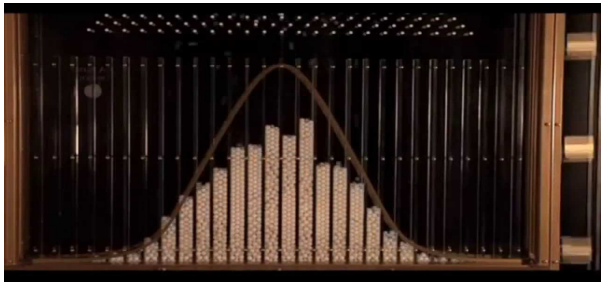


(Image from G. Schiebinger)

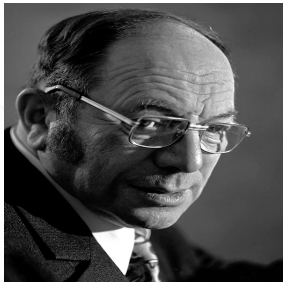
- Economics
- Finance
- Data Science/Statistics

Probability measures in practice

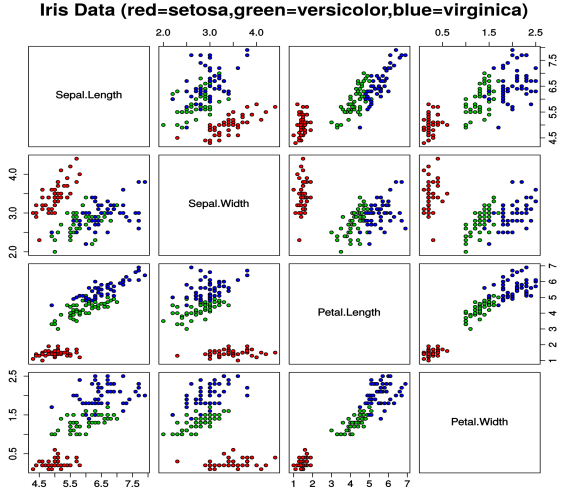
- Randomness



- Distributions
  - distribution of mass, temperature, etc.
  - Images/shapes



- Data sets



A basic problem of optimal transport, called **Monge-Kantorovich problem**, concerns matching two measures, say, μ ands ν on a certain measurable spaces X, Y, respectively.

Example

- μ=distribution of buyers over preferences in X
- ν= distribution of sellers over locations in Y.

Example

- μ=distributions of students over locations in X.
- ν= distribution of capacities for schools in Y.

Example

- μ= pixels in one photo
- ν= pixels in another photo

Example

- μ= a bacteria population today
- ν= the bacteria population tomorrow.

We are interested in studying **matchings π between measures μ and ν.**

Example

- μ= distribution of buyers,
- ν= distribution of sellers,
- π = how to match the buyers and sellers

Example

- μ= pixels in one photo
- ν= pixels in another photo
- π=how to match these pixels in different photos. (questions in image processing)

Example

- μ= a bacteria population today
- ν= a bacteria population tomorrow
- π = how the bacterias change from today to tomorrow.

- To make matching two measures μ and ν be rigorous, we assume that they have the same total mass.
- Equivalently, we assume that μ and ν are probability measures.

Transport plan π between probability measures μ, ν as a probability measure on the product space $X \times X$ with marginals μ and ν.

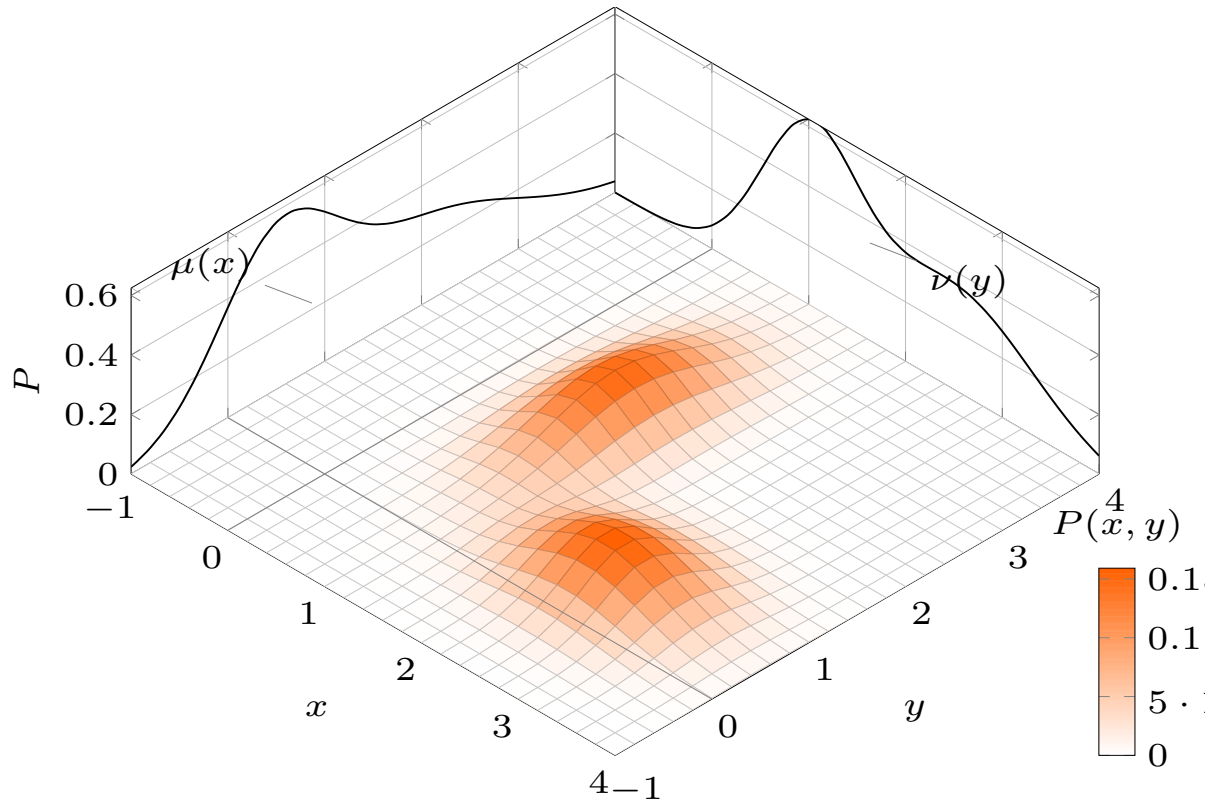


Image provided by Marco Cuturi

π has the marginals $\pi^1 = \mu, \quad \pi^2 = \nu$

Transport plans

- Let μ, ν be probability measures on X, Y, respectively.

A **transportation plan** π between μ and ν, is a measure on $X \times Y$, which satisfies

$$\int_Y d\pi = \mu \quad \text{(the first marginal of } \pi \text{ is } \mu),$$

$$\int_X d\pi = \nu \quad \text{(the second marginal of } \pi \text{ is } \nu),$$

$$\pi \geq 0.$$

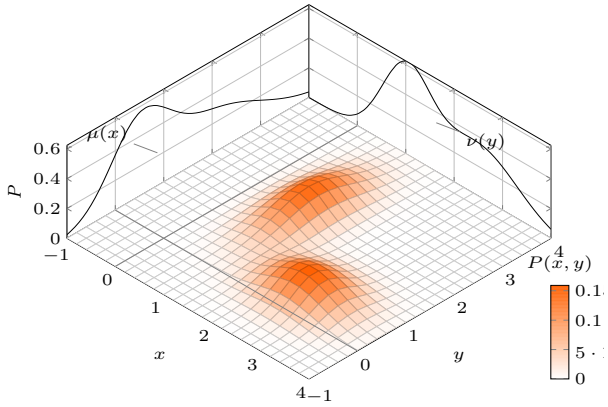


Image provided by Marco Cuturi

- Note that π itself is a probability measure:
  - $\int_{X \times Y} d\pi = 1.$

The set of transport plans between μ and ν:

$$\Pi(\mu, \nu) = \left\{ \pi \;\middle|\; \int_Y d\pi = \mu, \quad \int_X d\pi = \nu, \quad \pi \geq 0 \right\}$$

Existence of transport plans

Example

- $x, y \in \mathbb{R}^d$, $\mu = \delta_x$, $\nu = \delta_y$. Is there $\pi \in \Pi(\mu, \nu)$?

- $x \in \mathbb{R}^d$, $\mu = \delta_x$, $d\nu = \frac{1}{m(B_r(x))} \chi_{B_r(x)} dm$.
  Is there $\pi \in \Pi(\mu, \nu)$?

Q: Given probability measures μ and ν, does there exist $\pi \in \Pi(\mu, \nu)$, a transport plan between them?

Theorem

For any probability measures μ on X, ν on Y, there is a transport plan $\pi \in \Pi(\mu, \nu)$.

Proof.

Let $\pi = \mu \otimes \nu$, the product measure of μ and ν. $\square$

We are interested in **OPTIMAL** matchings (also called ‘transport’).
To discuss optimality, we introduce the cost of a transport plan.

Cost functions

Cost function: $c : X \times Y \to \mathbb{R}$.

- $c(x, y) =$
  the cost of moving a unit mass at location x to location y.

Example

For points $x \in \mathbb{R}^d$, $y \in \mathbb{R}^d$

- distance cost $c(x, y) = |x - y|$
- distance squared cost $c(x, y) = |x - y|^2$

Example

- μ= distribution of buyers,
- ν= distribution of sellers,
- π = how to match the buyers and sellers
- c= the “distance” between buyers and sellers
  - the ”distance” can be a distance in some abstract data space, not necessarily the physical distance.

Example

- μ= a bacteria population today
- ν= a bacteria population tomorrow
- π = how the bacterias change from today to tomorrow.
- c= the distance between cells
  - It can be some quantifier to measure the difference between two cells in some data space (e.g. the gene expression space).
    - There are many genes, so the gene expression space is of very high dimensions, e.g. $\geq$ 20,000.

Transportation cost

Given a transport plan π, **the transportation cost for** π:

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

Monge-Kantorovich problem

Given μ, ν and c,

$$MK(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y)d\pi(x,y).$$

That is,

$$\begin{array}{ll} \text{Minimize} & \int_{X\times Y} c(x,y)d\pi(x,y). \\ \text{subject to} & \int_Y d\pi = \mu \\ & \int_X d\pi = \nu \\ & \pi \geq 0 \end{array}$$

- This is an (infinite dimensional) Linear Programming (LP) problem!
- In discrete setting it is a finite dimensional LP:
  Given c_{ij}, $\mu = (\mu_1, ..., \mu_m), \nu = (\nu_1, ..., \nu_n)$,

$$\begin{array}{ll} \text{Minimize} & \sum_{i=1}^m \sum_{j=1}^n c_{ij}\pi_{ij} \\ \text{subject to} & \sum_{j=1}^n \pi_{ij} = \mu_i \\ & \sum_{i=1}^m \pi_{ij} = \nu_j \\ & \pi_{ij} \geq 0 \end{array}$$

- **Linear programming** is an optimization problem where the objective function is a linear function and constraints are given by linear equations or inequalities. It is a nontrivial problem in general, though there are effective ways to solve them, including the simplex method of G. Dantzig, for finite dimensional problems.

Example (distance squared cost example: two points to two points in $\mathbb{R}^1$.)

- $x_1 = 0, x_2 = 1, y_1 = 2, y_2 = 3 \in \mathbb{R}$.
- $\mu = (1/2, 1/2)$,
  mass $1/2$ at x_1 and mass $1/2$ at x_2.
- $\nu = (1/2, 1/2)$,
  mass $1/2$ at y_1 and mass $1/2$ at y_2.
- $c_{ij} = |x_i - y_j|^2$.

$$\min_{\pi \in \Pi(\mu,\nu)} \sum_{ij} c_{ij} \pi_{ij}.$$

What are the optimal solutions?

Distance squared cost example: two points to two points in $\mathbb{R}^1$

- $$[c_{ij}] = \begin{bmatrix} 2^2 & 3^2 \\ 1^2 & 2^2 \end{bmatrix}, \quad [\pi_{ij}] = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix}$$

- $$\sum_{ij} c_{ij}\pi_{ij} = 2^2\pi_{11} + 3^2\pi_{12} + \pi_{21} + 2^2\pi_{22}$$

- Thus we have the linear programming problem:

$$\begin{array}{ll} \text{Minimize} & 2^2\pi_{11} + 3^2\pi_{12} + \pi_{21} + 2^2\pi_{22} \\ \text{subject to} & \pi_{11} + \pi_{12} = 1/2 \\ & \pi_{21} + \pi_{22} = 1/2 \\ & \pi_{11} + \pi_{21} = 1/2 \\ & \pi_{12} + \pi_{22} = 1/2 \\ & \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22} \geq 0 \end{array}$$

Distace squared cost example: two points to two points in $\mathbb{R}^1$

$$
\begin{array}{ll}
\text{Minimize} & 2^2\pi_{11} + 3^2\pi_{12} + \pi_{21} + 2^2\pi_{22} \\
\text{subject to} & \pi_{11} + \pi_{12} = 1/2 \\
& \pi_{21} + \pi_{22} = 1/2 \\
& \pi_{11} + \pi_{21} = 1/2 \\
& \pi_{12} + \pi_{22} = 1/2 \\
& \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22} \geq 0
\end{array}
$$

is equivalent to

$$
\begin{array}{ll}
\text{Minimize} & 2^2\pi_{11} + 3^2(1/2 - \pi_{11}) + (1/2 - \pi_{11}) + 2^2\pi_{22} \\
\text{subject to} & \pi_{11} \leq 1/2 \\
& 1/2 - \pi_{11} + \pi_{22} = 1/2 \\
& \pi_{11} \leq 1/2 \\
& 1/2 - \pi_{11} + \pi_{22} = 1/2 \\
& \pi_{11}, \pi_{22} \geq 0
\end{array}
$$

Distance squared cost: two points to two points in $\mathbb{R}^1$

$$
\begin{array}{rl}
\text{Minimize} & 2^2\pi_{11} + 3^2(1/2 - \pi_{11}) + (1/2 - \pi_{11}) + 2^2\pi_{22} \\
\text{subject to} & \pi_{11} \leq 1/2 \\
& 1/2 - \pi_{11} + \pi_{22} = 1/2 \\
& \pi_{11} \leq 1/2 \\
& 1/2 - \pi_{11} + \pi_{22} = 1/2 \\
& \pi_{11}, \pi_{22} \geq 0
\end{array}
$$

is equivalent to

$$
\begin{array}{rl}
\text{Minimize} & 2^2\pi_{11} + 3^2(1/2 - \pi_{11}) + (1/2 - \pi_{11}) + 2^2\pi_{11} \\
\text{subject to} & \pi_{11} \leq 1/2 \\
& \pi_{11} \geq 0
\end{array}
$$

which is equivalent to

$$
\begin{array}{rl}
\text{Minimize} & 5 - 2\pi_{11} \\
\text{subject to} & 0 \leq \pi_{11} \leq 1/2 \\
& \pi_{11} \geq 0
\end{array}
$$

Distance squared cost: two points to two points in $\mathbb{R}^1$

$$\begin{array}{ll} \text{Minimize} & 5 - 2\pi_{11} \\ \text{subject to} & 0 \leq \pi_{11} \leq 1/2 \end{array}$$

What does this mean?

- Answer: Optimal solution is $\pi_{11} = 1/2$.
  So, from the constraints

$$\begin{aligned} \pi_{11} + \pi_{12} &= 1/2 \\ \pi_{21} + \pi_{22} &= 1/2 \\ \pi_{11} + \pi_{21} &= 1/2 \\ \pi_{12} + \pi_{22} &= 1/2 \end{aligned}$$

  we get
  $\pi_{12} = 0, \pi_{21} = 0, \pi_{22} = 1/2$.

Optimal transport introduces a geometric structure on the space of (probability) measures, that gives a natural extension of the geometry of the underlying space.

Optimal transport and Wasserstein distance

Let $(X, dist)$ be a metric space;
e.g. $X = \mathbb{R}^d$, $dist(x, y) = |x - y|$.

- A **distance** between probability measures μ, ν on X:

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

  called **the Wasserstein distance**.

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

We will see some aspects of this geometry in later lectures.

Solutions of Monge-Kantorovich problem

$$MK(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y)d\pi(x,y).$$

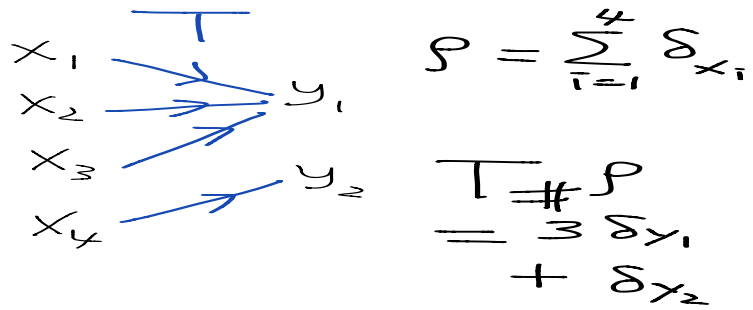
Before discuss solutions of MK-problem, let us recall the notion ‘push-forward’ of a measure by measurable mappings.

Push-forward

For a measurable mapping $T : X \to Y$ and a measure μ on X, the push-forward $T_{\#}\mu$ is a measure on Y defined as

$$T_{\#}\mu(E) = \mu(T^{-1}(E)) \quad \text{for each measurable set } E \subset Y.$$

Example



Example

$\lambda > 0$, $T : \mathbb{R}^d \to \mathbb{R}^d$, $T(x) = \lambda x$. $m =$ Lebesgue measure. Then, $T_{\#}m = \lambda^{-d}m$.

Solutions to the Monge-Kantorovich Problem

$$MK(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y)d\pi(x,y).$$

Kantorovich solution vs. Monge solution.

- **Kantorovich solution**: an **optimal** transport plan $\pi^* \in \Pi(\mu,\nu)$.
- **Monge solution**: a special type of Kantorovich solution. A measurable mapping $T : X \to Y$ defined μ-a.e. such that $T_{\#}\mu = \nu$ and the corresponding $\pi^* := (id \times T)_{\#}\mu$, is a Kanotorovich solution.

  An illustration:

$$MK(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y)d\pi(x,y).$$

Theorem (Brenier 80's)

Suppose

- $X = Y = \mathbb{R}^n$, μ, ν *are probability measures, compactly supported.*
- $\mu \ll$ *Lebesgue,*
  - *that is,* $d\mu = fdm$, *for the Lebesgue measure m, f measurable;*
- $c(x,y) = |x-y|^2$.

Then,

- *there exists unique optimal solution* π^* *to* $MK(\mu,\nu)$;
- π^* *is given by a measurable mapping* $T : \mathbb{R}^n \to \mathbb{R}^n$ *defined* μ*-a.e., that is,* $\pi^* = (id \times T)_{\#}\mu$;
- T *is given by a convex function* $\phi : \mathbb{R}^n \to \mathbb{R}$ *in the sense that* $T(x) = \nabla\phi(x)$ *for* μ*-a.e.* x.

Dualtiy of optimal transport

Duality

Primal:

$$\begin{array}{ll} \text{Minimize} & \int_{X\times Y} c(x,y)d\pi(x,y). \\ \text{subject to} & \int_Y d\pi = \mu \\ & \int_X d\pi = \nu \\ & \pi \geq 0 \end{array}$$

$$\begin{array}{ll} \text{Minimize} & \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\pi_{ij} \\ \text{subject to} & \sum_{j=1}^{n} \pi_{ij} = \mu_i \\ & \sum_{i=1}^{m} \pi_{ij} = \nu_j \\ & \pi_{ij} \geq 0 \\ & \forall i = 1,...,m, \forall j = 1,...,n \end{array}$$

Dual:

$$\begin{array}{ll} \text{Maximize} & \int_X \phi d\mu + \int_Y \psi d\nu \\ \text{subject to} & \phi(x) + \psi(y) \leq c(x,y)\ \forall x, y \end{array}$$

$$\begin{array}{ll} \text{Maximize} & \sum_{i=1}^{m} \phi_i \mu_i + \sum_{j=1}^{n} \psi_j \nu_j \\ \text{subject to} & \phi_i + \psi_j \leq c_{ij} \\ & \forall i = 1,...,m, \forall j = 1,...,n \end{array}$$

This follows from the linear programming duality:

- **Primal:**

$$\begin{array}{ll} \text{Minmize} & \vec{c}^T \vec{x} \\ \text{subject to} & A\vec{x} = \vec{b} \\ & \vec{x} \geq \vec{0} \end{array}$$

- **Dual:**

$$\begin{array}{ll} \text{Maximize} & \vec{b}^T \vec{y} \\ \text{subject to} & A^T \vec{y} \leq \vec{c} \end{array}$$

Explanation of duality

So, we can (roughly) see

Primal:

$$\begin{array}{ll} \text{Min.} & \vec{c}^T \vec{x} \\ \text{subj.} & A\vec{x} = \vec{b} \\ & \vec{x} \geq \vec{0} \end{array}$$

Primal:

$$\begin{array}{ll} \text{Min.} & \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \pi_{ij} \\ \text{subj.} & \sum_{j=1}^{n} \pi_{ij} = \mu_i \\ & \sum_{i=1}^{m} \pi_{ij} = \nu_j \\ & \pi_{ij} \geq 0 \\ & \forall i = 1, ..., m, \\ & \forall j = 1, ..., n \end{array}$$

Dual:

$$\begin{array}{ll} \text{Max.} & \vec{b}^T \vec{y} \\ \text{subj.} & A^T \vec{y} \leq \vec{c} \end{array}$$

Dual:

$$\begin{array}{ll} \text{Max} & \sum_{i=1}^{m} \phi_i \mu_i + \sum_{j=1}^{n} \psi_j \nu_j \\ \text{subj.} & \phi_i + \psi_j \leq c_{ij} \\ & \forall i = 1, ..., m, \\ & \forall j = 1, ..., n \end{array}$$

Another explanation of duality of OT

Use Lagrange multipliers:

- $$\begin{array}{ll} \text{Minimize} & \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\pi_{ij} \\ \text{subject to} & \sum_{j=1}^{n}\pi_{ij} = \mu_i \\ & \sum_{i=1}^{m}\pi_{ij} = \nu_j \\ & \pi_{ij} \geq 0 \end{array}$$

  is equivalent to

- $$\min_{\pi \geq \mathbf{0}} \max_{\phi,\psi} \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\pi_{ij} + \sum_i \phi_i \left[\mu_i - \sum_{j=1}^{n}\pi_{ij}\right] + \sum_j \psi_i \left[\nu_j - \sum_{i=1}^{m}\pi_{ij}\right]$$

Another explanation of duality: continued

- $$\min_{\pi \geq \mathbf{0}} \max_{\phi,\psi} \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\pi_{ij} + \sum_i \phi_i \left[\mu_i - \sum_{j=1}^{n} \pi_{ij}\right] + \sum_j \psi_i \left[\nu_j - \sum_{i=1}^{m} \pi_{ij}\right]$$
- is equivalent to
$$\min_{\pi \geq \mathbf{0}} \max_{\phi,\psi} \left[\sum_i \phi_i \mu_i + \sum_j \psi_i \nu_j + \sum_{i=1}^{m}\sum_{j=1}^{n} \left[c_{ij} - \phi_i - \psi_j\right] \pi_{ij}\right]$$
- In fact, in this case one can reorder min max to max min. (Using Min-Max theorem of Von Neumann in the finite dimensional case, and using Fenchel duality for general (infinite) dimensional case.)

Another explanation of duality: continued

- $$\max_{\phi,\psi} \min_{\pi \geq \mathbf{0}} \left[\sum_i \phi_i \mu_i + \sum_j \psi_i \nu_j + \sum_{i=1}^{m} \sum_{j=1}^{n} \left[c_{ij} - \phi_i - \psi_j \right] \pi_{ij} \right]$$
- If $c_{i'j'} - \phi_{i'} - \psi_{j'} < 0$ for some i', j', one can choose

$$\pi_{ij} = \begin{cases} \lambda & \text{for } i = i', j = j' \ , \\ 0 & \text{otherwise.} \end{cases}$$

  and such that the **minimum** $\to -\infty$ as $\lambda \to \infty$.
  - This will not count for the **maximum**. So, such case can be ignored.
- For those ϕ, ψ with $c_{ij} - \phi_i - \psi_j \geq 0$, the minimum occurs when $\pi = \mathbf{0}$.
- Thus, the problem can be reduced to

$$\max_{\phi_i + \psi_j \leq c_{ij}} \left[\sum_i \phi_i \mu_i + \sum_j \psi_i \nu_j \right] .$$

□

We have:

$$\min_{\pi \in \Pi(\mu,\nu)} \sum_i \sum_j c_{ij} \pi_{ij} = \max_{\phi_i + \psi_j \leq c_{ij}} \left[\sum_i \phi_i \mu_i + \sum_j \psi_i \nu_j \right] .$$

c-transforms

That condition $\phi \oplus \psi \leq c$ can be related to
***c*-transform**:

$$\phi^c(y) = \min_x [c(x, y) - \phi(x)]$$
$$\psi^c(x) = \min_y [c(x, y) - \psi(y)]$$

Or the discrete version:

$$\phi_j^c = \min_i \left[c_{ij} - \phi_i\right] \qquad \& \qquad \psi_i^c = \min_j \left[c_{ij} - \psi_j\right]$$

Special case: $c(x, y) = x \cdot y$ for $x, y \in \mathbb{R}^d$

$$\phi^*(y) = \min_{x \in \mathbb{R}^d} [x \cdot y - \phi(x)]$$

[This is a version of the so-called **Legendre transform** in convex analysis.]
The function ϕ^* is a concave function.
See board for a picture.

Example

For $x, y \in \mathbb{R}$.

- If $\phi(x) = 0$, then $\phi^*(y) = -\infty$ if $y \neq 0$, and $\phi^*(0) = 0$.
- If $\phi(x) = x$, then $\phi^*(y) = -\infty$ if $y \neq 1$, and $\phi^*(1) = 0$.
- (Exercise) If $\phi(x) = -\frac{1}{2}x^2$, then $\phi^*(y) = -\frac{1}{2}y^2$.

Properties of c-transform

$$\phi^c(y) = \min_x [c(x,y) - \phi(x)] \quad \& \quad \psi^c(x) = \min_y [c(x,y) - \psi(y)]$$

- Note for all ϕ, ψ,
  - $\phi(x) + \phi^c(y) \leq c(x,y) \quad \forall x, y$
  - $\psi^c(x) + \psi(y) \leq c(x,y) \quad \forall x, y$

- If $\phi(x) + \psi(y) \leq c(x,y)$ for all x, y, then
  - $\psi(y) \leq c(x,y) - \phi(x)$ and $\phi(y) \leq c(x,y) - \psi(x)$ for all x, y.
  - Therefore,

$$\psi(y) \leq \phi^c(y) \quad \forall y \quad \& \quad \phi(x) \leq \psi^c(x) \quad \forall x.$$

Reduction of the maximization problem

- $$\begin{aligned}\sum_i \phi_i \mu_i + \sum_j \psi_j \nu_j &\leq \sum_i \psi_i^c \mu_i + \sum_j \psi_j \nu_j \\ &\leq \sum_i \psi_i^c \mu_i + \sum_j \psi_j^{cc} \nu_j \qquad (\psi^{cc} := (\psi^c)^c)\end{aligned}$$

- Therefore

$$\max_{\phi_i + \psi_j \leq c_{ij}} \left[\sum_i \phi_i \mu_i + \sum_j \psi_i \nu_j \right] = \max_{(\psi_i^c, \psi_j^{cc})} \left[\sum_i \psi_i^c \mu_i + \sum_j \psi_j^{cc} \nu_j \right].$$

Consequence of the reduction

$$\max_{\phi_i+\psi_j\leq c_{ij}} \left[\sum_i \phi_i\mu_i + \sum_j \psi_i\nu_j\right] = \max_{(\psi_i^c,\psi_j^{cc})} \left[\sum_i \psi_i^c\mu_i + \sum_j \psi_j^{cc}\nu_j\right].$$

$$\max_{\phi\oplus\psi\leq c} \left[\int_X \phi d\mu + \int_Y \psi d\nu\right] = \max_{(\psi^c,\psi^{cc})} \left[\int_X \psi^c d\mu + \int_Y \psi^{cc} d\nu\right].$$

The dual **optimal** solution has the special property.

- It is given by a pair $(\bar{\psi}^c, \bar{\psi}^{cc})$ of functions on X and Y respectively.
- The functions $\bar{\psi}^c, \bar{\psi}^{cc}$ have special features:
  - They are *c*-**concave functions**.
  - For example, if $c(x, y) = \pm x \cdot y$, then they are concave functions.
    - $$\phi^c(y) = \min_x [\pm x \cdot y - \phi(x)] \quad \& \quad \psi^c(x) = \min_y [\pm x \cdot y - \psi(y)]$$

- When $c(x,y) = |x-y|^2/2 = |x|^2/2 - x\cdot y + |y|^2/2$,
  - $$\begin{aligned}\int c(x,y)d\pi(x,y) &= \int \left(|x|^2/2 - x\cdot y + |y|^2/2\right) d\pi(x,y)\\ &= \int \frac{|x|^2}{2} d\pi + \int (-x\cdot y)d\pi + \int \frac{|y|^2}{2} d\pi(x,y)\\ &= \int (-x\cdot y)d\pi + \int \frac{|x|^2}{2} d\mu(y) + \int \frac{|y|^2}{2} d\nu(y)\end{aligned}$$

    since $\int_Y d\pi = \mu, \quad \int_X d\pi = \nu.$
  - Therefore, minimizing $\int c(x,y)d\pi(x,y)$ is equivalent to minimizing $\int (-x\cdot y)d\pi$.

Weak duality, No duality gap & Existence of optimizers of OT.

Primal:

$$\begin{array}{ll} \text{Minimize} & \int_{X\times Y} c(x,y)d\pi(x,y). \\ \text{subject to} & \int_Y d\pi = \mu \\ & \int_X d\pi = \nu \\ & \pi \geq 0 \end{array}$$

Dual:

$$\begin{array}{ll} \text{Maximize} & \int_X \phi d\mu + \int_Y \psi d\nu \\ \text{subject to} & \phi(x) + \psi(y) \leq c(x,y)\ \forall x, y \end{array}$$

- **Weak duality:**
  Primal objective values $\geq$ Dual objective values.
  - It holds generally.
- **No duality gap:**
  Primal **optimal** objective value $=$ Dual **optimal** objective value.
  - It holds generally for finite dimensional linear programming
  - It holds for practically cases: for compact domains X, Y, continuous $c(x, y)$, etc.
- **Existence of primal and dual optimizaer**
  - It holds for practical cases: for compact domains X, Y, Lipschitz $c(x, y)$, etc.
  - Existence of dual optimizer is in general more difficult than that of the primal.

Complementary Slackness

Primal:

$$\begin{array}{ll} \text{Minimize} & \int_{X\times Y} c(x,y)d\pi(x,y). \\ \text{subject to} & \int_Y d\pi = \mu \\ & \int_X d\pi = \nu \\ & \pi \geq 0 \end{array}$$

Dual:

$$\begin{array}{ll} \text{Maximize} & \int_X \phi d\mu + \int_Y \psi d\nu \\ \text{subject to} & \phi(x) + \psi(y) \leq c(x,y)\ \forall x, y \end{array}$$

Suppose π^* and $(\bar{\psi}^c, \bar{\psi}^{cc})$ are optimal primal and dual solutions to the optimal transport problem. Then No Duality Gap

$$\int c d\pi^* = \int_X \bar{\psi}^c d\mu + \int_Y \bar{\psi}^{cc} d\nu$$

implies

$$\int (c(x,y) - \bar{\psi}^c(x) - \bar{\psi}^{cc}(y)) d\pi^*(x,y) = 0.$$

Since $c(x,y) - \bar{\psi}^c(x) - \bar{\psi}^{cc}(y)) \geq 0$ for all (x,y) by the constraint, we have **Complementary Slackness:**

$$c(x,y) - \bar{\psi}^c(x) - \bar{\psi}^{cc}(y)) = 0 \quad \pi^*\text{-a.e. } x, y.$$

Consequence of complementary slackness

- **[Complementary Slackness]**
  Suppose π^* and $(\bar{\psi}^c, \bar{\psi}^{cc})$ are optimal primal and dual solutions to the optimal transport problem.
  Then, π^* is concentrated on the set

$$\{(x, y) \mid c(x, y) - \bar{\psi}^c(x) - \bar{\psi}^{cc}(y) = 0\}$$

- This means the following:
  - By solving the dual problem, one can find properties of the primal optimal solution π^*.
  - As the functions $\bar{\psi}^c$, $\bar{\psi}^{cc}$ are *special*, the set of x, y with $c(x, y) - \bar{\psi}^c(x) - \bar{\psi}^{cc}(y) = 0$ has a special property.
  - We can derive special properties of π^*.

Duality and connection to partial differential equations (PDE)

- Consider the case the mass distributions μ, ν are continuously distributed on domains in $\mathbb{R}^d$.
- Then, the dual optimal solution $\bar{\psi}^c$ is given by the solution of the **Monge-Ampère type equation**:

$$\det\left[D^2\bar{\psi}^c(x) - D^2\mathcal{A}(x, D\bar{\psi}^c(x))\right] = f(x, D\bar{\psi}^c(x))$$

  for some $\mathcal{A}$ and f.
  - D, D^2 are the first and second partial differentiations.
- Understanding the optimal transport π^* is connected to understanding solutions of this partial differential equation.