Simultaneous Optimal Transport

Ruodu Wang

http://sas.uwaterloo.ca/~wang

Department of Statistics and Actuarial Science
University of Waterloo

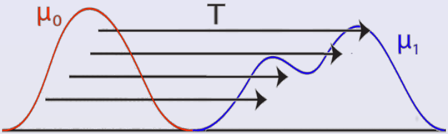
Economics Seminar, Bielefeld University
November 30, 2021 (online)

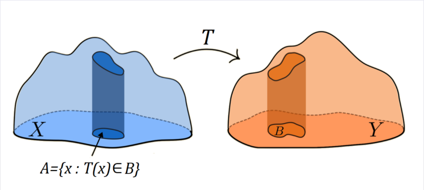
Agenda

Based on joint work with Zhenyuan Zhang (Stanford)

Transport theory

- Pure mathematics theory
- Important applications
  - economics
  - decision theory
  - finance
  - engineering
  - operations research
  - physics
- 1 Nobel Prize laureate
- 2 Fields medalists

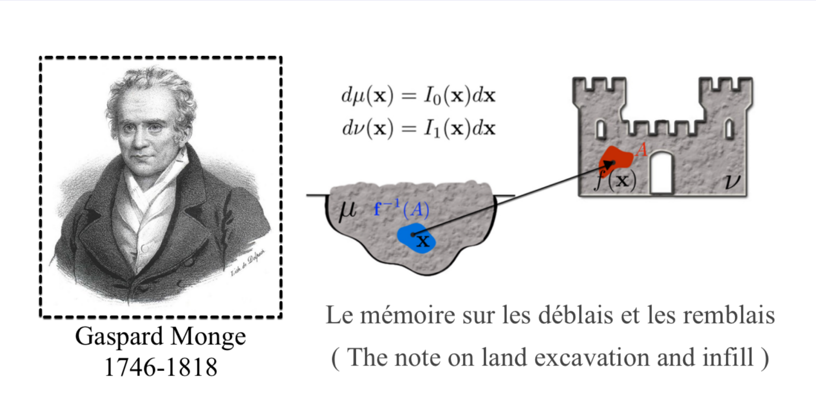




Monge's formulation

- Monge's problem: find a transport map $T : X \to Y$ that attains
$$\inf \left\{ \int_X c(x, T(x))\, \mathrm{d}\mu(x) \,\middle|\, T_\# \mu = \nu \right\}$$
where
  - X and Y are two Polish spaces (main example: $\mathbb{R}^d$)
  - Cost function $c : X \times Y \to [0, \infty]$ or $(-\infty, \infty]$
  - probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ are given
  - $T_\# \mu = \mu \circ T^{-1}$ is the push forward of μ by T
- Such T is an optimal transport map

Monge's formulation



Kantorovich's formulation

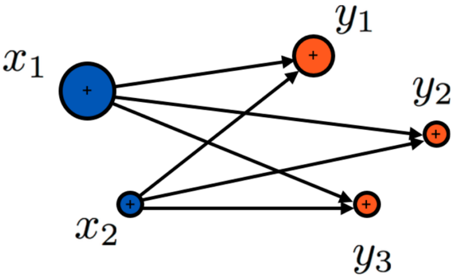
- Monge's formulation may be ill-posed (e.g., point masses)
- Kantorovich's problem: find a probability measure $\pi \in \mathcal{P}(X \times Y)$ that attains

$$\inf \left\{ \int_{X \times Y} c(x, y)\, \pi(\mathrm{d}x, \mathrm{d}y) \mid \pi \in \Pi(\mu, \nu) \right\},$$

  where $\Pi(\mu, \nu)$ is the set of probability measures on $X \times Y$ with marginals μ and ν
- $X \times Y = \mathbb{R} \times \mathbb{R}$: copulas and dependence
- Discrete version: linear programming

Kantorovich's formulation

Leonid Kantorovich
1912-1986



Resource allocation

Kernel formulation

- Kernel formulation: find a transport kernel $\kappa : X \to \mathcal{P}(Y)$ that attains

$$\inf \left\{ \int_{X \times Y} c(x,y)(\mu \otimes \kappa)(\mathrm{d}x, \mathrm{d}y) \mid \kappa \in \mathcal{K}(\mu,\nu) \right\}$$

where $\mathcal{K}(\mu,\nu)$ is the set of all stochastic kernels κ such that

$$\kappa_{\#}\mu := \int_X \kappa(x)\mu(\mathrm{d}x) = \nu$$

- $(\mu \otimes \kappa)(A) = \int_A \kappa(x, \mathrm{d}y)\mu(\mathrm{d}x)$
- $\mu \otimes \kappa = \pi \in \Pi(\mu,\nu)$
- $\kappa(x) = \nu$ for each $x \in X$: independent coupling

Transport duality

If c is non-negative and lower semi-continuous, then duality holds

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c \,\mathrm{d}\pi = \sup \left(\int_X \phi \,\mathrm{d}\mu + \int_Y \psi \,\mathrm{d}\nu \right),$$

where the supremum runs over all pairs of bounded and continuous functions $\phi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$ such that

$$\phi(x) + \psi(y) \leqslant c(x, y).$$

Economic interpretation

- $x \in X$: the vector of characteristics of a worker
- $y \in Y$: the vector of characteristics of a firm
- $g(x, y)$ the economic output (production) generated by worker x matched with firm y
- Social economic-output maximization

$$\sup \left\{ \int_{X \times Y} g \, \mathrm{d}\pi \mid \pi \in \Pi(\mu, \nu) \right\}$$

- Dual problem $g(x, y) \leqslant \phi(x) + \psi(y)$: social equilibrium
  - ϕ: the equilibrium wage function
  - ψ: the equilibrium profit function

On the cost function

Assume $X = Y = \mathbb{R}$.

- If c is submodular, i.e.,

$$c(x, y) + c(x', y') \leqslant c(x, y') + c(x', y) \quad \text{for } x \leqslant x' \text{ and } y \leqslant y',$$

  the optimal transport is comonotone. Examples:
  - $c(x, y) = (y - x)^2$
  - $c(x, y) = -\mathbb{1}_{\{(x,y) \leqslant (x_0, y_0)\}}$
  - $c(x, y) = f(x) + g(y) + h(y - x)$ where h is convex
- If c is supermodular, the optimal transport is antitone (counter-monotonic).
- $c(x, y) = \mathbb{1}_{\{y - x > d_0\}}$: probability of transport distance $> d_0$

Probabilistic formulation

For random variables $L \sim \mu$ and $R \sim \nu$

- Classic optimal transport (OT)

$$\inf_{L\sim\mu, R\sim\nu} \mathbb{E}[c(L,R)]$$

- Martingale optimal transport (MOT) require: $\mu \preceq_{\mathrm{cx}} \nu$

$$\inf_{L\sim\mu, R\sim\nu} \mathbb{E}[c(L,R)] : \; L = \mathbb{E}[R|L]$$

- Supermartingale optimal transport (SMOT) require: $\mu \succeq_{\mathrm{ssd}} \nu$

$$\inf_{L\sim\mu, R\sim\nu} \mathbb{E}[c(L,R)] : \; L \geqslant \mathbb{E}[R|L]$$

- Directional optimal transport (DOT) require: $\mu \preceq_{\mathrm{st}} \nu$

$$\inf_{L\sim\mu, R\sim\nu} \mathbb{E}[c(L,R)] : \; L \leqslant R$$

MOT: Beiglböck/Henry-Labordère/Penkner'13 F&S; Beiglböck/Juillet'16 AoP
SMOT: Nutz/Stebegg'18 AoP; DOT: Nutz/W.'21 AAP

Simultaneous transport

- $d \in \mathbb{N}$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d) \in \mathcal{P}(X)^d$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d) \in \mathcal{P}(Y)^d$
- A transport plan from $\boldsymbol{\mu}$ to $\boldsymbol{\nu}$ sends μ_j to ν_j for all $j \in \{1, \ldots, d\}$ simultaneously
- The set of all Monge transports from $\boldsymbol{\mu}$ to $\boldsymbol{\nu}$

$$\mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \{T : X \to Y \mid T_{\#}\boldsymbol{\mu} = \boldsymbol{\nu}\}$$

- The set of all transport kernels κ such that $\kappa_{\#}\boldsymbol{\mu} = \boldsymbol{\nu}$

$$\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \bigcap_{j=1}^{d} \mathcal{K}(\mu_j, \nu_j)$$

- Existence not guaranteed; Kantorovich formulation unclear

All equalities and inequalities are component-wise

Motivating example 1: rocket planning

m Mars bases and n space stations

several types of resources to be transported by rockets

Motivating example 1: rocket planning

- Each base j supplies $\mu_A(j)$ units of A and $\mu_B(j)$ units of B
- Each station k needs $\nu_A(k)$ units of A and $\nu_B(k)$ units of B
- Assume supply-demand clearance

$$\sum_{j=1}^{m} \mu_A(j) = \sum_{k=1}^{n} \nu_A(k) = \sum_{j=1}^{m} \mu_B(j) = \sum_{k=1}^{n} \nu_B(k) = 1$$

- A transport plan is an arrangement to send resources from bases to stations to meet their needs
- Single trips: cannot transport among stations or among bases
- Transport costs: rockets and fuel (propellant)
- $\mathcal{T}((\mu_A, \mu_B), (\nu_A, \nu_B))$: one base supplies only one station

Motivating example 1: rocket planning

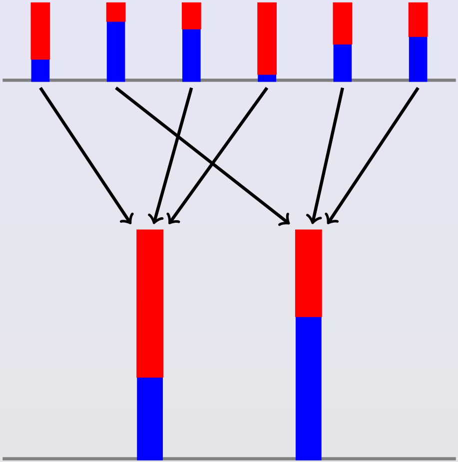


Figure: Simultaneous transport in the Monge setting; red and blue represent different types of resources

Motivating example 2: product distribution

Transport several types of products from factories to retailers

- Kernel setting $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$
- Allowing one factory to supply multiple retailers

Assumptions

- supply-demand clearance
- products are bundled and can only divided proportionally
  - e.g., personnel, skills, boxed packages, cargo specification

Motivating example 2: product distribution

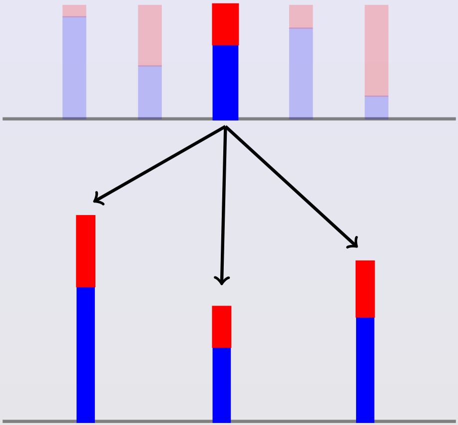


Figure: Simultaneous transport in the kernel setting

Cost

How do we model the cost?

- Cost function $c : X \times Y \to [0, +\infty]$
- Classic setting: for $T \in \mathcal{T}(\mu, \nu)$ or $\kappa \in \mathcal{K}(\mu, \nu)$,

$$\mathcal{C}(T) = \int_X c(x, T(x))\mu(\mathrm{d}x)$$
$$\mathcal{C}(\kappa) = \int_{X \times Y} c \,\mathrm{d}(\mu \otimes \kappa)$$

- Simultaneous transport: what should take the place of μ?
- Choose $\bar{\mu} := \frac{1}{d} \sum_{j=1}^{d} \mu_j$?
- Baseline measure $\eta \in \mathcal{M}(X)$, $\eta \ll \bar{\mu}$ (no transport $\Rightarrow$ no cost)

Cost

- Transport cost for $T \in \mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu})$ or $\kappa \in \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$,

$$\mathcal{C}_{\eta}(T) = \int_X c(x, T(x))\eta(\mathrm{d}x)$$

$$\mathcal{C}_{\eta}(\kappa) = \int_{X\times Y} c\,\mathrm{d}(\eta \otimes \kappa)$$

- η may not be linear in $\boldsymbol{\mu}$, e.g., petrol cost is nonlinear in weights
- If $\mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu})$ or $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is empty, then the cost is set to ∞
- Special case: $\eta = \bar{\mu}$
- Optimal transport:

$$\inf_{T\in\mathcal{T}(\boldsymbol{\mu},\boldsymbol{\nu})} \mathcal{C}_{\eta}(T) \quad \text{or} \quad \inf_{\kappa\in\mathcal{K}(\boldsymbol{\mu},\boldsymbol{\nu})} \mathcal{C}_{\eta}(\kappa)$$

Motivating example 3: risk measures

Scenario-based risk measures (W.-Ziegel'21 F&S)

- $\mathcal{X}$ the space of random variables on Ω and $\mu_1, \dots, \mu_d \in \mathcal{P}(\Omega)$
- A $\boldsymbol{\mu}$-based risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is such that $\rho(L)$ is determined by the distribution of L under $\boldsymbol{\mu}$
  - $d = 1$: law-invariant under μ
- Define $\mathcal{X}_{\boldsymbol{\mu}}(L) = \{R \in \mathcal{X} : R_\# \boldsymbol{\mu} = L_\# \boldsymbol{\mu}\}$
- $R_\# \boldsymbol{\mu} = L_\# \boldsymbol{\mu}$ means $R \overset{\text{law}}{=} L$ under each μ_i for $i = 1, \dots, d$
- Any $\boldsymbol{\mu}$-based risk measure is constant on each $\mathcal{X}_{\boldsymbol{\mu}}(L)$

Motivating example 3: risk measures

- For $\eta \ll \bar{\mu}$, the mapping

$$\rho(L) = \sup \{\mathbb{E}^{\eta}[R] : R \in \mathcal{X}_{\boldsymbol{\mu}}(L)\}, \qquad L \in \mathcal{X}$$

is a $\boldsymbol{\mu}$-based risk measure (coherent distortion if $d = 1$)
  - $d = 1$: Kusuoka'01 representation of coherent risk measures
- The optimization problem

$$\rho(L) = \sup \{\mathbb{E}^{\eta}[R] : R_{\#}\boldsymbol{\mu} = L_{\#}\boldsymbol{\mu}\}$$

is a Monge transport problem with cost $c(x, y) = -y$, baseline measure η, and $\boldsymbol{\nu} = L_{\#}\boldsymbol{\mu}$

Motivating example 4: cost-efficient payoffs

- η is a pricing measure
- d agents need to jointly purchase a payoff R
- Agent j needs R to have a distribution ν_j and uses model μ_j
- Problem: find the cheapest R

$$\min \{\mathbb{E}^{\eta}[R] : R_{\#}\boldsymbol{\mu} = \boldsymbol{\nu}\}$$

- $d = 1$: Fréchet-Hoeffding; Dybvig'88 JB

Motivating example 5: Markovian embedding

- An $\mathbb{R}^d$-valued Markov process $\xi = (\xi_t)_{t=1,\dots,T}$
- ξ has marginal distributions $\mu_1, \dots, \mu_T \in \mathcal{P}(\mathbb{R}^d)$
- The Markov kernel κ_t of ξ: $(\xi_{t+1} \mid \xi_t = x) \overset{\text{law}}{\sim} \kappa_t(x)$
- Time-homogeneity: $\kappa = \kappa_t$ does not depend on t
- $\kappa \in \mathcal{K}(\mu_t, \mu_{t+1})$ for $t = 1, \dots, T-1$
- $\kappa \in \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ with $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{T-1})$ and $\boldsymbol{\nu} = (\mu_2, \dots, \mu_T)$
  - with fixed marginals, each κ corresponds to a time-homogeneous Markov process

Inequalities

- Δ_d: the standard simplex in $\mathbb{R}^d$
- Each κ transports each $\boldsymbol{\lambda} \cdot \boldsymbol{\mu}$ to $\boldsymbol{\lambda} \cdot \boldsymbol{\nu}$ for $\boldsymbol{\lambda} \in \Delta_d$

$$\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu}) \subseteq \bigcap_{\boldsymbol{\lambda} \in \Delta_d} \mathcal{K}(\boldsymbol{\lambda} \cdot \boldsymbol{\mu}, \boldsymbol{\lambda} \cdot \boldsymbol{\nu})$$

$$\inf_{\kappa \in \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})} \mathcal{C}_\eta(\kappa) \geqslant \sup_{\boldsymbol{\lambda} \in \Delta_d} \inf_{\kappa \in \mathcal{K}(\boldsymbol{\lambda} \cdot \boldsymbol{\mu}, \boldsymbol{\lambda} \cdot \boldsymbol{\nu})} \mathcal{C}_\eta(\kappa) \geqslant \inf_{\kappa \in \mathcal{K}(\bar{\mu}, \bar{\nu})} \mathcal{C}_\eta(\kappa)$$

- not sharp in general

Inequalities

- $d = 2$
- $\tilde{\mu}_1 = (\mu_1 - \mu_2)_+$, and similar

$$\inf_{\kappa \in \mathcal{K}(\boldsymbol{\mu},\boldsymbol{\nu})} \mathcal{C}_\eta(\kappa) \geqslant \inf_{\substack{\kappa \in \mathcal{K}(\mu,\tilde{\nu}_1) \\ \mu \leqslant \tilde{\mu}_1}} \mathcal{C}_\eta(\kappa) + \inf_{\substack{\kappa \in \mathcal{K}(\mu,\tilde{\nu}_2) \\ \mu \leqslant \tilde{\mu}_2}} \mathcal{C}_\eta(\kappa)$$

Figure: shaded $\tilde{\mu}_1$ covers shaded $\tilde{\nu}_1$; grey $\tilde{\mu}_2$ covers grey $\tilde{\nu}_2$

Inequalities

- Both inequalities are sharp if μ is mutually singular
- Transport problem is back to $d = 1$ if
  - μ is mutually singular, or
  - μ and ν both have identical components
- Generally, there is no symmetry between X and Y

The role of η

If $c'(x,y) = c(x,y) + \phi(x) + \boldsymbol{\psi}(y) \cdot \frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}\eta}(x)$, then

$$\int_{X\times Y} c' \,\mathrm{d}(\eta \otimes \kappa) = \int_{X\times Y} c \,\mathrm{d}(\eta \otimes \kappa) + \underbrace{\int_X \phi \,\mathrm{d}\eta + \int_Y \boldsymbol{\psi}^\top \mathrm{d}\boldsymbol{\nu}}_{\text{does not depend on } \kappa}$$

Classic setting ($d = 1$): if $c(x,y) = \phi(x) + \psi(y)$ then $\int c \,\mathrm{d}(\mu \otimes \kappa)$ does not depend on κ

Example

- $X = Y = [0,1]$
- $\eta = \bar{\mu} =$ Lebesgue, $\mathrm{d}\mu_1/\mathrm{d}\bar{\mu} = 2x$, $\boldsymbol{\nu}$ arbitrary
- Assume $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is nonempty

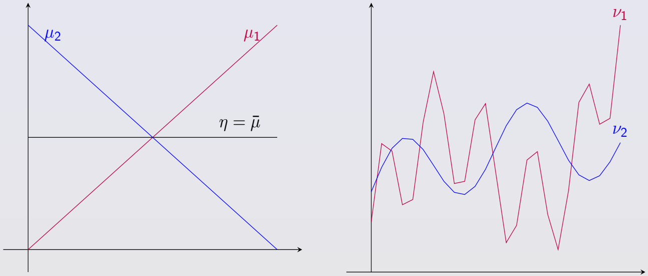


Figure: An example of densities of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, $d = 2$

Examples

- $c(x, y) = (x - y)^2$
- take $\phi(x) = x^2$, $\psi_1(y) = -y$, $\psi(y) = y^2$

Decomposition holds

$$c(x, y) = \phi(x) + \underbrace{\psi_1(y)\frac{\mathrm{d}\mu_1}{\mathrm{d}\bar{\mu}}(x)}_{-2xy} + \psi(y)$$

$$\implies \int c \,\mathrm{d}(\bar{\mu} \otimes \kappa) = \int_X x^2 \bar{\mu}(\mathrm{d}x) - \int_Y y\nu_1(\mathrm{d}y) + \int_Y y^2 \bar{\nu}(\mathrm{d}y)$$

- all transports have the same cost
- no optimal comonotone transport
- the two inequalities are not sharp

Existence

Definition 1 (Shen-Shen-Wang-W.'19 F&S)

Let $\boldsymbol{\mu} \in \mathcal{P}(X)^d$ and $\boldsymbol{\nu} \in \mathcal{P}(Y)^d$.

- Write $\boldsymbol{\mu} \succeq_{\mathrm{h}} \boldsymbol{\nu}$ (heterogeneity) if there exist $\mu \gg \boldsymbol{\mu}$ and $\nu \gg \boldsymbol{\nu}$ such that $\mathrm{d}\boldsymbol{\mu}/\mathrm{d}\mu \geqslant_{\mathrm{cx}} \mathrm{d}\boldsymbol{\nu}/\mathrm{d}\nu$ where $\geqslant_{\mathrm{cx}}$ is the convex order.
- $\boldsymbol{\mu}$ is jointly atomless if there exist $\mu \gg \boldsymbol{\mu}$ and a random variable L such that under μ, L is continuously distributed and independent of $\mathrm{d}\boldsymbol{\mu}/\mathrm{d}\mu$.

- $d = 1$ recovers classic non-atomicity
- called conditional atomless by Shen-Shen-Wang-W.'19
- μ, ν can be chosen as $\bar{\mu}, \bar{\nu}$

Existence

Proposition 1 (Torgersen'91; Shen-Shen-Wang-W.'19)

Let $\boldsymbol{\mu} \in \mathcal{P}(X)^d$ and $\boldsymbol{\nu} \in \mathcal{P}(Y)^d$.

(i) *The set $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is nonempty if and only if $\boldsymbol{\mu} \succeq_{\mathrm{h}} \boldsymbol{\nu}$.*

(ii) *Assume that $\boldsymbol{\mu}$ is jointly atomless. The set $\mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is nonempty if and only if $\boldsymbol{\mu} \succeq_{\mathrm{h}} \boldsymbol{\nu}$.*

- $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$, $\mathcal{K}(\boldsymbol{\nu}, \boldsymbol{\eta})$ nonempty $\Longrightarrow$ $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\eta})$ nonempty

Joint non-atomicity

Remarks.

- Let $\boldsymbol{\mu}' = \mathrm{d}\boldsymbol{\mu}/\mathrm{d}\bar{\mu}$ and $\boldsymbol{\nu}' = \mathrm{d}\boldsymbol{\nu}/\mathrm{d}\bar{\nu}$
- $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$, $\mathcal{K}(\boldsymbol{\nu}, \boldsymbol{\mu})$ nonempty $\iff \boldsymbol{\mu}' \overset{\text{law}}{=} \boldsymbol{\nu}'$ (wrt $\bar{\mu}$ and $\bar{\nu}$ resp.)
- $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ nonempty and $\boldsymbol{\mu}$ identical $\implies \boldsymbol{\nu}$ identical
- $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ nonempty and $\boldsymbol{\mu}$ equivalent $\implies \boldsymbol{\nu}$ equivalent
- $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ nonempty and $\boldsymbol{\nu}$ mutually singular $\implies \boldsymbol{\mu}$ mutually singular
- $\boldsymbol{\nu}$ identical $\implies \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ nonempty for all $\boldsymbol{\mu}$
- $\boldsymbol{\mu}$ mutually singular $\implies \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ nonempty for all $\boldsymbol{\nu}$

Joint non-atomicity

Definition 2 (Delbaen'21 F&S)

Let $(\Omega, \mathcal{G}, \mu)$ be a measure space. We say that $(\mathcal{G}, \mu)$ is atomless conditionally to the sub-σ-field $\mathcal{F} \subseteq \mathcal{G}$, if for all $A \in \mathcal{G}$ with $\mu(A) > 0$, there exists $A' \subseteq A$, $A' \in \mathcal{G}$, such that

$$\mathbb{E}^{\mu}[\mathbb{1}_A|\mathcal{F}] > 0 \implies 0 < \mathbb{E}^{\mu}[\mathbb{1}_{A'}|\mathcal{F}] < \mathbb{E}^{\mu}[\mathbb{1}_A|\mathcal{F}].$$

Lemma 1 (Delbaen'21)

Let μ be any strictly positive convex combination of $\boldsymbol{\mu} \in \mathcal{P}(X)^d$. Then $\boldsymbol{\mu}$ is jointly atomless if and only if $(\mathcal{B}(X), \mu)$ is atomless conditionally to $\sigma(\mathrm{d}\boldsymbol{\mu}/\mathrm{d}\mu)$.

Kantorovich formulation

- Define
$$\Pi_\eta(\boldsymbol{\mu},\boldsymbol{\nu}) = \{\eta\otimes\kappa \mid \kappa\in\mathcal{K}(\boldsymbol{\mu},\boldsymbol{\nu})\}$$
- If $\eta\sim\bar{\mu}$, it is
$$\left\{\pi\in\mathcal{P}(X\times Y) \mid \pi_X=\eta,\ \int_X \frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}\eta}(x)\pi(\mathrm{d}x,\mathrm{d}y)=\boldsymbol{\nu}(\mathrm{d}y)\right\}$$
where π_X is the first marginal of π
- Omit η if $\eta=\bar{\mu}$: $\Pi(\boldsymbol{\mu},\boldsymbol{\nu})=\Pi_{\bar{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})$
- Cost
$$\mathcal{C}(\pi):=\int_{X\times Y} c\,\mathrm{d}\pi=\mathcal{C}_\eta(\kappa)$$
- π no longer has specified second marginal, unless η is a linear function of $\boldsymbol{\mu}$

Monge vs Kantorovich

Theorem 1

Suppose that X and Y are compact, $\boldsymbol{\mu}$ is jointly atomless, and c is continuous. Then

$$\inf_{\kappa \in \mathcal{K}(\boldsymbol{\mu},\boldsymbol{\nu})} \mathcal{C}_\eta(\kappa) = \inf_{T \in \mathcal{T}(\boldsymbol{\mu},\boldsymbol{\nu})} \mathcal{C}_\eta(T).$$

Remarks on joint non-atomicity.

- classic non-atomicity ($d = 1$) $\iff$ $\exists$ uniform rv
  joint non-atomicity $\iff$ $\exists$ uniform rv independent of $\mathrm{d}\boldsymbol{\mu}/\mathrm{d}\bar{\mu}$
- In both settings of non-atomicity
  - $\exists$ Monge transport $\iff$ $\exists$ Kantorovich transport
  - Monge infimum $=$ Kantorovich infimum

Duality

Theorem 2

Suppose that X, Y are compact, $\eta \sim \bar{\mu}$, and $c : X \times Y \to [0, +\infty]$ is lower semi-continuous. Duality holds as

$$\min_{\pi \in \Pi_\eta(\boldsymbol{\mu}, \boldsymbol{\nu})} \int_{X \times Y} c \,\mathrm{d}\pi = \sup_{(\phi, \boldsymbol{\psi}) \in \Phi_c} \int_X \phi \,\mathrm{d}\eta + \int_Y \boldsymbol{\psi}^\top \,\mathrm{d}\boldsymbol{\nu},$$

where

$$\Phi_c = \left\{ (\phi, \boldsymbol{\psi}) \in C(X) \times C(Y)^d \mid \phi(x) + \boldsymbol{\psi}(y) \cdot \frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}\eta}(x) \leqslant c(x, y) \right\}.$$

- $d = 1$ and $\eta = \bar{\mu}$: classic duality (but with compactness)
- duality holds for $X = Y = \mathbb{R}^N$ if $\mathrm{d}\eta / \mathrm{d}\bar{\mu}$ is bounded

Uniqueness of the transport

- $\boldsymbol{\mu}' = \mathrm{d}\boldsymbol{\mu}/\mathrm{d}\bar{\mu}$ and $\boldsymbol{\nu}' = \mathrm{d}\boldsymbol{\nu}/\mathrm{d}\bar{\nu}$

Theorem 3

Suppose that both $\Pi(\boldsymbol{\mu},\boldsymbol{\nu})$ and $\Pi(\boldsymbol{\nu},\boldsymbol{\mu})$ are nonempty and $\boldsymbol{\mu}'$ is injective on the support of $\boldsymbol{\mu}$.

(i) *There exist a unique $\pi \in \Pi(\boldsymbol{\mu},\boldsymbol{\nu})$ and a unique $\tilde{\pi} \in \Pi(\boldsymbol{\nu},\boldsymbol{\mu})$.*

(ii) $\pi(A \times B) = \tilde{\pi}(B \times A)$ *for all* $(A, B) \in \mathcal{B}(X) \times \mathcal{B}(Y)$.

(iii) $\pi(\{(x, y) \mid \boldsymbol{\mu}'(x) \neq \boldsymbol{\nu}'(y)\}) = 0$.

Remarks.

- $\boldsymbol{\mu}'$ injective $\iff \mathcal{B}(X) = \sigma(\boldsymbol{\mu}') \implies \boldsymbol{\mu}$ not jointly atomless
- $d = 1$: $\boldsymbol{\mu}' = 1$ injective $\iff X = \{x\}$

Proof of the uniqueness

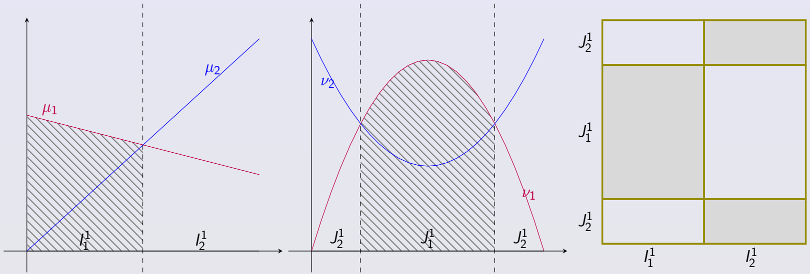


Figure: Idea of the proof in case $d = 2$. The transports are divided into shaded (I_1^1, J_1^1) and unshaded parts (I_2^1, J_2^1); π must be supported in the gray area

Uniqueness of the transport

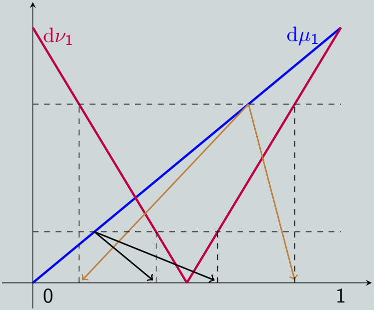
Example 1

$X = Y = [0,1]$, $\mu_2 = \nu_2 = \text{Unif}$, $\mathrm{d}\mu_1 = 2x\mathrm{d}x$, $\mathrm{d}\nu_1 = |2-4x|\mathrm{d}x$

- μ_1' strictly increasing
- a transport kernel

$$\kappa(x) = \frac{1}{2}\delta_{(1+x)/2} + \frac{1}{2}\delta_{(1-x)/2}$$

- $\Pi(\boldsymbol{\mu},\boldsymbol{\nu})$, $\Pi(\boldsymbol{\nu},\boldsymbol{\mu})$ nonempty
- unique transport kernel
- no Monge, $\mathcal{T}(\boldsymbol{\mu},\boldsymbol{\nu}) = \emptyset$



Wasserstein distance

- $X = Y$ equipped with a metric ρ; $p \geqslant 1$
- Define

$$\mathcal{P}(X)_{p,\rho} = \left\{ \mu \in \mathcal{P}(X) \mid \exists x_0 \in X,\ \int_X \rho(x, x_0)^p \mu(\mathrm{d}x) < \infty \right\}$$

- Classic Wasserstein distance between μ and ν in $\mathcal{P}(X)_{p,\rho}$

$$\mathcal{W}_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{X^2} \rho^p \, \mathrm{d}\pi \right)^{1/p}$$

- Example: $X = \mathbb{R}^d$, $\rho =$ Euclidean and $p = 2$
- Wasserstein distance between $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ in $\mathcal{P}(X)^d_{p,\rho}$?

Wasserstein distance

- $\eta = \bar{\mu}$
- For $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{P}(X)^d_{p,\rho}$, define the "Wasserstein distance"

$$\mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \left(\inf_{\pi \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})} \int_{X^2} \rho^p \, \mathrm{d}\pi \right)^{1/p}$$

- Triangle inequality OK; symmetry NO
  - not a metric on $\mathcal{P}(X)^d_{p,\rho}$
- Find $\mathcal{E} \subseteq \mathcal{P}(X)^d_{p,\rho}$ such that $\mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \mathcal{W}_p(\boldsymbol{\nu}, \boldsymbol{\mu})$ on $\mathcal{E}$

Wasserstein distance

Theorem 4

Let $\boldsymbol{\mu} \in \mathcal{P}(X)^d$ and $\boldsymbol{\nu} \in \mathcal{P}(Y)^d$. Suppose that both $\Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$ and $\Pi(\boldsymbol{\nu}, \boldsymbol{\mu})$ are nonempty. Then

$$\inf_{\pi \in \Pi(\boldsymbol{\mu},\boldsymbol{\nu})} \int_{X \times Y} c(x, y)\pi(\mathrm{d}x, \mathrm{d}y) = \inf_{\tilde{\pi} \in \Pi(\boldsymbol{\nu},\boldsymbol{\mu})} \int_{Y \times X} c(x, y)\tilde{\pi}(\mathrm{d}y, \mathrm{d}x).$$

$\Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$, $\Pi(\boldsymbol{\nu}, \boldsymbol{\mu})$ nonempty $\iff \boldsymbol{\mu}' \overset{\text{law}}{=} \boldsymbol{\nu}'$

- If $X = Y$ and c is symmetric, then

$$\mathcal{I}_c(\boldsymbol{\mu}, \boldsymbol{\nu}) := \inf_{\pi \in \Pi(\boldsymbol{\mu},\boldsymbol{\nu})} \mathcal{C}(\pi) = \inf_{\tilde{\pi} \in \Pi(\boldsymbol{\nu},\boldsymbol{\mu})} \mathcal{C}(\tilde{\pi}) = \mathcal{I}_c(\boldsymbol{\nu}, \boldsymbol{\mu})$$

- $\mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \mathcal{W}_p(\boldsymbol{\nu}, \boldsymbol{\mu})$

Wasserstein distance

- An equivalence relation: $\boldsymbol{\mu} \simeq \boldsymbol{\nu}$ if $\bar{\mu} \circ (\boldsymbol{\mu}')^{-1} = \bar{\nu} \circ (\boldsymbol{\nu}')^{-1}$
  - both $\Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$ and $\Pi(\boldsymbol{\nu}, \boldsymbol{\mu})$ are nonempty
- $\mathcal{E}_P$: equivalent class under $\simeq$ where $P = \bar{\mu} \circ (\boldsymbol{\mu}')^{-1} \in \mathcal{P}(\mathbb{R}_+)^d$
- $\mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu})$ is indeed a distance on each $\mathcal{E}_P$
- For $\kappa \in \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$,

$$\int_{X^2} c \, \mathrm{d}(\bar{\mu} \otimes \kappa) = \frac{1}{d} \sum_{j=1}^{d} \int_{X^2} c \, \mathrm{d}(\mu_j \otimes \kappa)$$

$$\implies \quad \mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu})^p \geqslant \frac{1}{d} \sum_{j=1}^{d} \mathcal{W}_p(\mu_j, \nu_j)^p$$

Wasserstein distance

Degenerate cases

- $\boldsymbol{\mu}$ identical and $\boldsymbol{\nu}$ identical $\implies$ the optimal transport from μ_1 to ν_1 is optimal from $\boldsymbol{\mu}$ to $\boldsymbol{\nu}$

$$\implies \quad \mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu})^p = \frac{1}{d} \sum_{j=1}^{d} \mathcal{W}_p(\mu_j, \nu_j)^p = \mathcal{W}_p(\mu_1, \nu_1)^p$$

- $d = 1 \implies \Pi(\mu, \nu)$, $\Pi(\nu, \mu)$ nonempty $\implies \mathcal{W}_p$ is the classic Wasserstein distance on $\mathcal{P}(X)_{p,\rho}$
- $\boldsymbol{\mu}$ mutually singular $\implies \mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu})^p = \frac{1}{d} \sum_{j=1}^{d} \mathcal{W}_p(\mu_j, \nu_j)^p$

Wasserstein distance

- Decompose (P-a.s.) $\bar{\mu} = \int \bar{\mu}_{\mathbf{z}} P(\mathrm{d}\mathbf{z})$ where $\{\mu_{\mathbf{z}}\}_{\mathbf{z}\in\mathbb{R}^d_+}$ is given by $\bar{\mu}_{\mathbf{z}}(\{x \in X : \boldsymbol{\mu}' = z\}) = 1$
- Similarly $\bar{\nu} = \int \bar{\nu}_{\mathbf{z}} P(\mathrm{d}\mathbf{z})$

Theorem 5

For $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{E}_P$ and $\kappa \in \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$, the following are equivalent:

(i) *κ is an optimal transport from $\boldsymbol{\mu}$ to $\boldsymbol{\nu}$;*

(ii) *κ is an optimal transport from $\bar{\mu}_{\mathbf{z}}$ to $\bar{\nu}_{\mathbf{z}}$ for each P-a.s. $\mathbf{z}$;*

(iii) $\mathcal{C}_{\bar{\mu}}(\kappa) = \int_{\mathbb{R}^d_+} \mathcal{I}_c(\bar{\mu}_{\mathbf{z}}, \bar{\nu}_{\mathbf{z}}) P(\mathrm{d}\mathbf{z})$.

Wasserstein distance

- Let $\kappa_{\mathbf{z}}$ be an optimal transport from $\bar{\mu}_{\mathbf{z}}$ to $\bar{\nu}_{\mathbf{z}}$ for each $\mathbf{z}$
- Define $\kappa(x) := \kappa_{\mu'(x)}(x) \implies \kappa$ is optimal and

$$\mathcal{I}_c(\boldsymbol{\mu}, \boldsymbol{\nu}) = \int_{\mathbb{R}_+^d} \mathcal{I}_c(\bar{\mu}_{\mathbf{z}}, \bar{\nu}_{\mathbf{z}}) P(\mathrm{d}\mathbf{z})$$

$$\mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu})^p = \int_{\mathbb{R}_+^d} \mathcal{W}_p(\bar{\mu}_{\mathbf{z}}, \bar{\nu}_{\mathbf{z}})^p P(\mathrm{d}\mathbf{z})$$

- For $1 \leqslant p < \infty$, $(\mathcal{E}_P, \mathcal{W}_p)$ is a Polish space
- Topology induced by $\mathcal{W}_p$ is yet unclear

Wasserstein distance

Corollary 1

Let $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{E}_P$, $X = \mathbb{R}$ *and* $c : \mathbb{R}^2 \to [0, \infty]$ *be submodular. Then*

$$\mathcal{I}_c(\boldsymbol{\mu}, \boldsymbol{\nu}) = \int_{\mathbb{R}_+^d} \int_0^1 c(F_{\mathbf{z}}^{-1}(t), G_{\mathbf{z}}^{-1}(t)) \mathrm{d}t \, P(\mathrm{d}\mathbf{z}),$$

where $F_{\mathbf{z}}^{-1}, G_{\mathbf{z}}^{-1}$ *are the distribution functions of* $\mu_{\mathbf{z}}, \nu_{\mathbf{z}}$ *respectively.*

If $\rho =$ Euclidean on $\mathbb{R}$, then

$$\mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu})^p = \int_{\mathbb{R}_+^d} \int_0^1 |F_{\mathbf{z}}^{-1}(t) - G_{\mathbf{z}}^{-1}(t)|^p \mathrm{d}t \, P(\mathrm{d}\mathbf{z})$$

An equilibrium model

- Assume $\eta \sim \sum_{i=1}^{d} \mu_i$, X and Y are compact, and $g : X \times Y \to [-\infty, \infty)$ is upper semi-continuous

Duality holds

$$\max_{\pi \in \Pi_\eta(\mu, \boldsymbol{\nu})} \int_{X \times Y} g \, \mathrm{d}\pi = \inf_{(\phi, \psi) \in \Phi_g} \int_X \phi \, \mathrm{d}\eta + \int_Y \boldsymbol{\psi}^\top \, \mathrm{d}\boldsymbol{\nu},$$

where

$$\Phi_g = \left\{ (\phi, \boldsymbol{\psi}) \in C(X) \times C^d(Y) : \ \phi(x) + \boldsymbol{\psi}(y) \frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}\eta}(x) \geqslant g(x, y) \right\}$$

An equilibrium model

- $x \in X$ represents worker labels (characteristics)
- $y \in Y$ represent firms
- η: the distribution of the workers labelled with $x \in X$
  - discrete $\eta(x) = 1/n$: each worker has their own label
- d types of skills
  - workers with the same label have the same skills
- μ_i: supply of type-i skill provided by the workers
  - discrete $\mu_i(x)$: type-i skill provided by each worker label x
  - $\boldsymbol{\mu}' = \mathrm{d}\boldsymbol{\mu}/\mathrm{d}\eta$: (per-worker) skill vector
- ν_i: demand of type-i skill from the firms
  - discrete $\nu_i(y)$: type-i skill demanded by each firm y

An equilibrium model

- Assume that demand and supply of each skill are equal
  - $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are normalized
- A matching is $\kappa \in \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$; $\pi = \eta \otimes \kappa$
- $g(x, y)$: the production of firm y hiring worker x (per unit)
  - total production: $\int g \, \mathrm{d}\pi$
- $w : X \to \mathbb{R}$: wage function
  - $w(x)$ is the wage of worker x
- $\mathbf{p} : Y \to \mathbb{R}^d$: profit-per-skill function
  - assumption: profit is linear in skills employed
  - if firm y employs a skill vector $\mathbf{q} \in \mathbb{R}_+^d$, its profit is $\mathbf{p}(y) \cdot \mathbf{q}$
  - the profit generated from hiring worker x is $\mathbf{p}(y) \cdot \boldsymbol{\mu}'(x)$
- $(w, \mathbf{p})$: a social plan

An equilibrium analysis

The total profit of all firms is

$$\int_{X\times Y} \mathbf{p}(y)\cdot \boldsymbol{\mu}'(x)\mathrm{d}\pi(\mathrm{d}x,\mathrm{d}y) = \int_Y \mathbf{p}^\top \mathrm{d}\boldsymbol{\nu}.$$

For worker x, their objective is to choose a firm to maximize their wage

$$y_x^* = \arg\max_{y\in Y} \left\{g(x,y) - \mathbf{p}(y)\cdot \boldsymbol{\mu}'(x)\right\}.$$

For firm y, its objective is to hire workers to maximize its profit

$$x_y^* = \arg\max_{x\in X} \left\{g(x,y) - w(x)\right\}.$$

An equilibrium analysis

For a social plan $(w, \mathbf{p})$ and a matching $\kappa \in \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$, an equilibrium is attained if

(a) the social plan is optimal

$$w(x) = \max_{y \in Y} \{g(x, y) - \mathbf{p}(y) \cdot \boldsymbol{\mu}'(x)\}$$

$$\mathbf{p}(y) \cdot \boldsymbol{\mu}'(x_y^*) = \max_{x \in X} \{g(x, y) - w(x)\}$$

(b) the total production covers the total wage plus the total profit

$$\int_{X \times Y} g \, \mathrm{d}\pi \geqslant \int_X w \, \mathrm{d}\eta + \int_Y \mathbf{p}^\top \mathrm{d}\boldsymbol{\nu}$$

An equilibrium analysis

$$
\begin{aligned}
\text{(a)} \quad &\Longrightarrow \quad w(x) + \mathbf{p}(y)\cdot \boldsymbol{\mu}'(x) \geqslant g(x,y) \\
+ \text{ integrate} \quad &\Longrightarrow \quad \int_X w \, \mathrm{d}\eta + \int_Y \mathbf{p}^\top \mathrm{d}\boldsymbol{\nu} \geqslant \int_{X\times Y} g \, \mathrm{d}\pi \\
+ \text{ (b)} \quad &\Longrightarrow \quad \int_X w \, \mathrm{d}\eta + \int_Y \mathbf{p}^\top \mathrm{d}\boldsymbol{\nu} = \int_{X\times Y} g \, \mathrm{d}\pi \\
&\Longrightarrow \quad \text{duality holds and attained}
\end{aligned}
$$

- Equilibrium exists $\Longleftrightarrow$ duality holds and attained
- Discrete setting: equilibrium exists $\Longleftrightarrow$ duality holds

Homogeneous Gaussian Markov process

Proposition 2

Suppose that $\mu_t = \mathrm{N}(0, \sigma_t^2)$, $\sigma_t > 0$, $t = 1, \ldots, T$. If there exists a time-homogeneous Markov process with the above marginals, then the mapping $t \mapsto \sigma_t$ is increasing log-concave or decreasing log-convex. If $T \leqslant 3$, then the above condition is also sufficient.

- General result?
- Optimal Markov process?
- Existence and optimality of simultaneous transport between two vectors of Gaussian measures on $\mathbb{R}^N$?

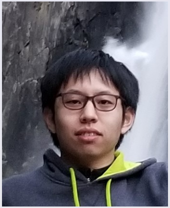
Future directions

- infinite dimension $d = \infty$
  - μ has no dominating measure
- multi-marginal transports $\mu^1, \dots, \mu^n$
- capacities instead of probabilities μ, ν nonlinear
- nonlinear cost in the probability η nonlinear
- constrained transport shrink $\mathcal{K}(\mu, \nu)$, $\mathcal{T}(\mu, \nu)$
  - martingale simultaneous transport
  - directional simultaneous transport
- Imperfect matching problems $\kappa_{\#}\mu \geqslant \nu$
  - requires $\mu(X) \geqslant \nu(Y)$ instead of $\mu(X) = \nu(Y)$

Thank you

Thank you for your attention!

Based on (on-going) joint work with



Zhenyuan Zhang
(Stanford)