Martingale Transports, Monge Maps, and Strassen's Theorem

Ruodu Wang

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

Department of Statistics and Actuarial Science
University of Waterloo

Optimal Transport and Distributional Robustness
Banff, March 25–29, 2024

Agenda

Based on joint work with Marcel Nutz (Columbia) and Zhenyuan Zhang (Stanford)

Main result

$X, Y \in L^1$: $X \leqslant_{\mathrm{cx}} Y$ means $\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)]$ for all convex ϕ

We will prove

A refinement of Strassen's theorem

On an atomless probability space, for $X, Y \in L^1$,

$$X \leqslant_{\mathrm{cx}} Y \iff X \overset{\mathrm{law}}{=} \mathbb{E}[Y|\mathcal{G}] \text{ for some } \sigma\text{-field } \mathcal{G}$$

... and some other related results

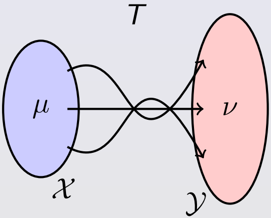
Monge's problem

Monge's problem: find a transport map $T : \mathcal{X} \to \mathcal{Y}$ that minimizes

$$\int_{\mathcal{X}} c(x, T(x))\,\mathrm{d}\mu(x) : T_{\#}\mu = \nu$$

where

- $\mathcal{X}$ and $\mathcal{Y}$ are two Polish spaces (e.g., $\mathbb{R}^d$)
- Cost function $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ or $(-\infty, \infty]$
- Probabilities μ on $\mathcal{X}$ and ν on $\mathcal{Y}$ are given
- $T_{\#}\mu = \mu \circ T^{-1}$ is the push forward of μ by T
- Such T is an optimal transport map



Kantorovich's problem

Kantorovich's problem

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

- $\Pi(\mu,\nu)$: probabilities on $\mathcal{X}\times\mathcal{Y}$ with marginals μ and ν

Kernel formulation

$$\min \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)(\mu\otimes\kappa)(\mathrm{d}x,\mathrm{d}y) : \kappa \in \mathcal{K}(\mu,\nu)$$

- $\mathcal{K}(\mu,\nu)$: kernels κ satisfying $\kappa_{\#}\mu := \int_{\mathcal{X}} \kappa_x \mu(\mathrm{d}x) = \nu$

Probabilistic formulation

$$\min \ \mathbb{E}[c(X,Y)] : \ X \overset{\text{law}}{\sim} \mu; \ Y \overset{\text{law}}{\sim} \nu$$

Martingale transport

- Martingale optimal transport

$$\min \ \mathbb{E}[c(X,Y)]: \ X \overset{\text{law}}{\sim} \mu; \ Y \overset{\text{law}}{\sim} \nu; \ X = \mathbb{E}[Y|X]$$

- Equivalently

$$\min \ \int c(x,y)(\mu \otimes \kappa)(\mathrm{d}x, \mathrm{d}y) : \kappa \in \mathcal{K}(\mu,\nu); \ e[\kappa_x] = x \ (\mu\text{-a.e.})$$

where e is the mean of a probability

- $\mathcal{M}(\mu,\nu)$: martingale transports (MT) in $\Pi(\mu,\nu)$

Martingale transport

- Motivated by (Asian) option pricing: sup/inf of

$$\left\{\mathbb{E}[(X_1+X_2-K)_+] : X_1 \overset{\text{law}}{\sim} \mu_1;\ X_2 \overset{\text{law}}{\sim} \mu_2;\ \mathbb{E}[X_2|X_1]=X_1\right\}$$

  where $K \in \mathbb{R}$ and μ_1, μ_2 are calibrated from option prices
- On $\mathbb{R}$:
  - $c(x,y)=(x-y)^2$: all MT have the same cost
  - $c(x,y)=h(x-y)$: the first two derivatives do not matter, but the third does

---

MOT: Beiglböck/Henry-Labordère/Penkner'13 FS; Beiglböck/Juillet'16 AOP

Monge Martingale transport

- $X = \mathbb{E}[Y|X]$ and Y is function of $X \Longrightarrow Y = X$
- MT cannot be Monge in the forward direction unless trivial
- Backward direction is possible:
  - $X = \mathbb{E}[Y|X]$ and X is function of Y
  - Example: $Y \overset{\text{law}}{\sim} \mathrm{U}[-2, 2]$ and $X = \mathrm{sign}(Y)$
- Monge martingale transports (MMT; omit "backward")
- $\mathcal{M}_M(\mu, \nu)$: set of MMT
- MMT is useful for:
  - identifying worst-case probabilities
  - some settings of matching problems

Strassen's theorem

- From now on $\mathcal{X} = \mathcal{Y} = \mathbb{R}$
- $\mathcal{P}(\mathbb{R})$: Borel probabilities on $\mathbb{R}$ with finite first moment
- $\mu \leqslant_{\mathrm{cx}} \nu$: $\int \phi \mathrm{d}\mu \leqslant \int \phi \mathrm{d}\nu$ for all convex ϕ

Strassen's Theorem

For $\mu, \nu \in \mathcal{P}(\mathbb{R})$, $\mu \leqslant_{\mathrm{cx}} \nu$ if and only if $\mathcal{M}(\mu, \nu)$ is non-empty.

- Increasing in risk — Rothschild-Stiglitz'70 JET

Monge martingale transport

Theorem 1 (Existence)

Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ satisfy $\mu \leqslant_{\mathrm{cx}} \nu$. There exists $\pi \in \mathcal{M}(\mu, \nu)$ and a Borel function $h : \mathbb{R} \to \mathbb{R}$ such that $\pi(T_{\mathrm{rg}} \cup T_{\mathrm{atom}}) = 1$, where

(i) $T_{\mathrm{rg}} = \{(h(y), y) : y \in \mathbb{R}\}$;

(ii) $T_{\mathrm{atom}} = \{(x, y) : \nu(\{y\}) > 0\}$.

In particular, if ν is atomless, π is a Monge martingale transport.

$$\nu \text{ atomless and } \mu \leqslant_{\mathrm{cx}} \nu \implies \text{MMT exists}$$

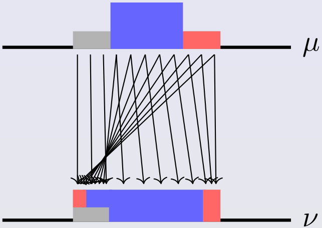
Main idea of the proof

- We need the left-curtain transport π_{lc} Beiglböck/Juillet'16 AOP
- Write $\mu \leqslant_{\mathrm{E}} \nu$ for finite measures μ, ν with finite first moment if $\int \phi \,\mathrm{d}\mu \leqslant \int \phi \,\mathrm{d}\nu$ for any nonnegative convex ϕ
  - if $\mu(\mathbb{R}) = \nu(\mathbb{R})$, this is $\mu \leqslant_{\mathrm{cx}} \nu$
  - if $\mu \leqslant \nu$ (set-wise) then $\mu \leqslant_{\mathrm{E}} \nu$
- Given $\mu \leqslant_{\mathrm{E}} \nu$, the shadow $S^\nu(\mu)$ of μ in ν is defined as

$$S^\nu(\mu) = \min_{\leqslant_{\mathrm{cx}}}\{\eta : \mu \leqslant_{\mathrm{cx}} \eta \leqslant \nu\},$$

  - always well-posed
- Given $\mu \leqslant_{\mathrm{cx}} \nu$, the left-curtain (LC) transport $\pi_{\mathrm{lc}} \in \mathcal{M}(\mu, \nu)$ is uniquely defined by the property that it transports $\mu|_{(-\infty,x]}$ to its shadow $S^\nu(\mu|_{(-\infty,x]})$ for every $x \in \mathbb{R}$

Main idea of the proof



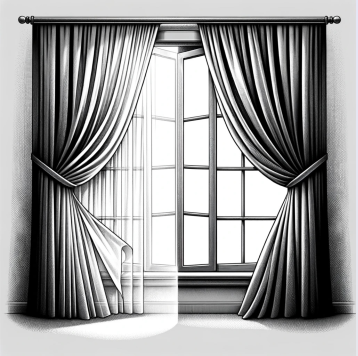


Figure: An example of the left-curtain transport (not MMT)

- The LC transport uniquely minimizes $\int c\mathrm{d}\pi$: $\pi \in \mathcal{M}(\mu,\nu)$ for $c(x,y) = h(y-x)$ with h' strictly convex

Main idea of the proof

- We will create a barcode transport by decomposing μ and ν into countably many mutually singular Monge parts
- Define the densities

$$d_\mu = \frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\nu)} \quad \text{and} \quad d_\nu = \frac{\mathrm{d}\nu}{\mathrm{d}(\mu+\nu)}$$

The barcode transport is defined by a sequential construction:

- Take the part on $\mathbb{R}$ with $d_\mu \geqslant 1/2$ "$\mathrm{d}\mu \geqslant \mathrm{d}\nu$"
- Apply the left-curtain transport on this part with its shadow
- Remove the matched parts from both μ and ν
- Repeat on the rest
- This procedure converges

Main idea of the proof

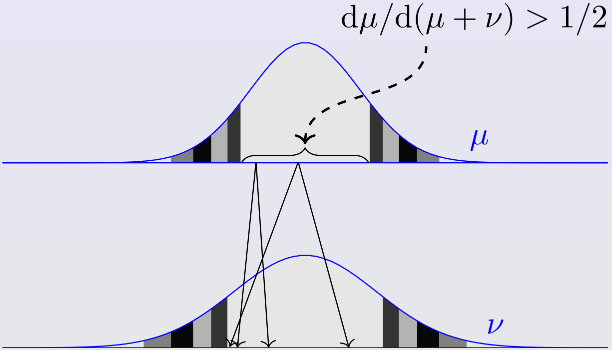


Figure: An example of the barcode transport between Gaussian marginals

Main idea of the proof

Proposition 1 (Structure of π_{lc})

Let $\mu \leqslant_{\mathrm{cx}} \nu$. There exists a Borel function $h : \mathbb{R} \to \mathbb{R}$ such that the LC transport π_{lc} satisfies $\pi_{\mathrm{lc}}(S_{\mathrm{rg}} \cup S_{\mathrm{diag}} \cup S_{\mathrm{atom}}) = 1$, where

(i) $S_{\mathrm{rg}} = \{(h(y), y) : y \in \mathbb{R}\}$;

(ii) $S_{\mathrm{diag}} = \{(x, x) : x \in \mathbb{R}\}$;

(iii) $S_{\mathrm{atom}} = \{(x, y) : \nu(\{y\}) > 0\}$.

If $d_\mu \geqslant 1/2$ μ-a.e., then $\pi_{\mathrm{lc}}(S_{\mathrm{rg}} \cup S_{\mathrm{atom}}) = 1$. In particular, if in addition ν is atomless, then $\pi_{\mathrm{lc}} \in \mathcal{M}_M(\mu, \nu)$.

$$\nu \text{ atomless and } d_\mu \geqslant 1/2 \implies \pi_{\mathrm{lc}} \text{ is MMT}$$

Let us prove Proposition 1

Lemma 1 (LC is left-monotone; Beiglböck/Juillet'16 AOP)

The LC transport $\pi_{\mathrm{lc}} \in \mathcal{M}(\mu, \nu)$ *satisfies* $\pi_{\mathrm{lc}}(\Gamma) = 1$ *for some* $\Gamma \subseteq \mathbb{R} \times \mathbb{R}$ *that is a left-monotone set; i.e.,*

$$(x, y^-), (x, y^+), (x', y') \in \Gamma \text{ with } x < x' \text{: it forbids } y^- < y' < y^+.$$

Moreover, $\pi_{\mathrm{lc}} \in \mathcal{M}(\mu, \nu)$ *is uniquely characterized by that property.*

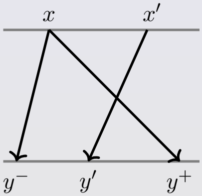


Figure: Forbidden configuration for left-monotonicity: the legs of a point x' cannot step into the legs of another point x to the left of x'

Let us prove Proposition 1

Lemma 2 (Support of LC; Beiglböck/Juillet'16 AOP)

There exist two functions $T_{\mathrm{d}},\ T_{\mathrm{u}} : \mathbb{R} \to \mathbb{R}$ *such that* $\pi_{\mathrm{lc}}(R_{\mathrm{legs}} \cup R_{\mathrm{atom}}) = 1$, *where*

(a) R_{legs} *is the union of the graphs of* $T_{\mathrm{d}},\ T_{\mathrm{u}}$ *over the first marginal;*

(b) $R_{\mathrm{atom}} = \{(x, y) : \mu(\{x\}) > 0\}$.

Let us prove Proposition 1

Denote by $\kappa_x(\mathrm{d}y)$ the disintegration of π_{lc} by μ:

$$\pi_{\mathrm{lc}}(\mathrm{d}x, \mathrm{d}y) = \mu(\mathrm{d}x) \otimes \kappa_x(\mathrm{d}y)$$

Lemma 3

We have $d_\mu \leqslant d_\nu$ μ-a.e. on $\{x \in \mathbb{R} : \kappa_x = \delta_x\}$.

Figure: For the left-curtain transport, $d_\mu \leqslant d_\nu$ μ-a.e. on $\{x \in \mathbb{R} : \kappa_x = \delta_x\}$

Let us prove Proposition 1

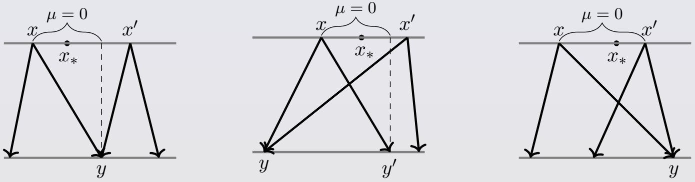
Let us prove

π_{lc} is supported on S_{rg} if $d_\mu \geqslant d_\nu$ μ-a.e. and ν is atomless

- ν is atomless $\implies$ if $\mu(\{x\}) > 0$ then $\kappa_x(\{x\}) = 0$
- Lemma 2 $\implies$ μ-a.e. if $\mu(\{x\}) = 0$ then either $\kappa_x = \delta_x$ or κ_x is supported on two points $T_{\mathrm{d}}(x) < x < T_{\mathrm{u}}(x)$.
- Lemma 3 $\implies$ μ-a.e. if $\kappa_x(\{x\}) > 0$ then $d_\mu(x) \leqslant d_\nu(x)$
- $d_\mu \geqslant d_\nu$ μ-a.e. $\implies$ μ-a.e. if $\kappa_x(\{x\}) > 0$ then $d_\mu(x) = d_\nu(x)$
- π_{lc} is id on $S := \{x \in \mathbb{R} : \kappa_x(\{x\}) > 0\}$ and Monge
- Safely remove S and assume $\kappa_x(\{x\}) = 0$ μ-a.e. to continue

Let us prove Proposition 1

- Lemma 1 $\Longrightarrow \pi_{\mathrm{lc}}$ is supported on a left-monotone set Γ
- $\Gamma \subseteq \mathrm{supp}(\mu) \times \mathrm{supp}(\nu)$ and $\Gamma \subseteq R_{\mathrm{legs}} \cup R_{\mathrm{atom}}$
- Previous step $\Longrightarrow \Gamma \cap \{(x,x) : x \in \mathbb{R}\} = \varnothing$
- Suppose $(x,y),(x',y) \in \Gamma$ with $y \notin \{x,x'\}$
  - (a) $x < y < x' \Longrightarrow \mu((x,y)) = 0$
  - (b) $y < x < x' \Longrightarrow \mu((x, y' \wedge x')) = 0$ for y' the right leg of x
  - (c) $x < x' < y \Longrightarrow \mu((x,x')) = 0$

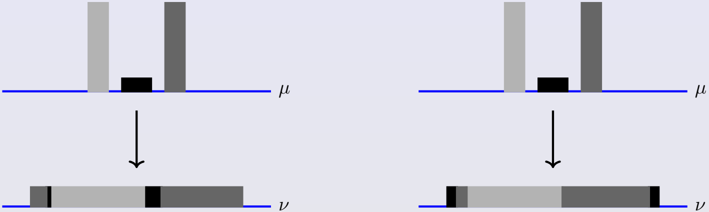


(a) The case $x < y < x'$ (b) The case $y < x < x'$ (c) The case $x < x' < y$

Let us prove Proposition 1

- As $\mathrm{supp}(\mu)$ is closed, its complement can be written as a countable disjoint union of open intervals
- Each pair of $(x, y), (x', y) \in \Gamma$ with $x \neq x'$ corresponds to an endpoint of one of the open intervals
- There are at most countably many y that do this
- ν atomless $\Longrightarrow$ such y has ν-measure 0 $\Longrightarrow \pi_{\mathrm{lc}}(S_{\mathrm{rg}}) = 1$
- Verify that the transport map in S_{rg} can be required Borel

Main idea of the proof



(a) Left-curtain transport (b) Barcode transport

Figure: It is possible that the left-curtain transport is MMT, but it is not necessarily equal to the barcode transport

Strassen's theorem on random variables

Consider the statement

$$X \leqslant_{\mathrm{cx}} Y \iff X \overset{\mathrm{law}}{=} \mathbb{E}[Y|\mathcal{G}] \text{ for some } \sigma\text{-field } \mathcal{G}?$$

- Jensen's inequality gives $\Leftarrow$
- Is $\Rightarrow$ true?

Example 1

Let $\Omega = \{1,2\}$, uniform probability, $X = (1,2)$ and $Y = (0,3)$

- $X \leqslant_{\mathrm{cx}} Y$ but $X \overset{\mathrm{law}}{=} \mathbb{E}[Y|\mathcal{G}]$ does not hold for any $\mathcal{G}$

Strassen's theorem on random variables

Reverse Jensen's:

$$X \leqslant_{\mathrm{cx}} Y \Longrightarrow X \overset{\mathrm{law}}{=} \mathbb{E}[Y|\mathcal{G}] \text{ for some } \sigma\text{-field } \mathcal{G} \qquad (\star)$$

What about an atomless space?

- If $(\star)$ holds true for $\sigma(Y) = \mathcal{F}$ then necessarily

$$X \leqslant_{\mathrm{cx}} Y \Longrightarrow X \overset{\mathrm{law}}{=} f(Y) \text{ for some measurable } f$$

- $(\star)$ requires a Monge martingale transport to exist!

A refinement of Strassen's theorem

Theorem 2 (Refinement of Strassen's Theorem)

For random variables X and Y on $(\Omega, \mathcal{F}, \mathbb{P})$ that is atomless, $X \leqslant_{\mathrm{cx}} Y$ if and only if $X \overset{\mathrm{law}}{=} \mathbb{E}[Y|\mathcal{G}]$ for some σ-field $\mathcal{G} \subseteq \mathcal{F}$.

- It is a refinement to the following version

Strassen's in the form of Theorem 3.A.4 of Shaked/Shantikumar'07

Two random variables X and Y satisfy $X \leqslant_{\mathrm{cx}} Y$ if and only if there exist random variables X', Y' on an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying $X' \overset{\mathrm{law}}{=} X$, $Y' \overset{\mathrm{law}}{=} Y$ and $X' = \mathbb{E}[Y'|X']$.

Denseness

Theorem 3 (MMTs are dense)

Let $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless. Then $\mathcal{M}_M(\mu,\nu)$ is weakly dense in $\mathcal{M}(\mu,\nu)$. If μ is discrete, it is also dense for the ∞-Wasserstein topology.

Corollary 1 (Optimal MT cost = optimal MMT cost)

Let $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless. If $c : \mathbb{R}^2 \to \mathbb{R}$ is continuous with $|c(x,y)| \leqslant a(x) + b(y)$ for some $a \in L^1(\mu)$ and $b \in L^1(\nu)$, then

$$\inf_{\pi \in \mathcal{M}_M(\mu,\nu)} \int c \mathrm{d}\pi = \inf_{\pi \in \mathcal{M}(\mu,\nu)} \int c \mathrm{d}\pi.$$

Denseness

Lemma 4

Let $\nu \in \mathcal{P}(\mathbb{R})$ be atomless. Given any decomposition $\nu = \sum_{i=1}^{\infty} \nu_i$ of ν, there exist mutually singular $\hat{\nu}_i$, $i \in \mathbb{N}$ such that $\nu = \sum_{i=1}^{\infty} \hat{\nu}_i$ and $\nu_1 \leqslant_{\mathrm{cx}} \hat{\nu}_1$ and $\mathrm{bary}(\nu_i) = \mathrm{bary}(\hat{\nu}_i)$ for $i \geqslant 2$.

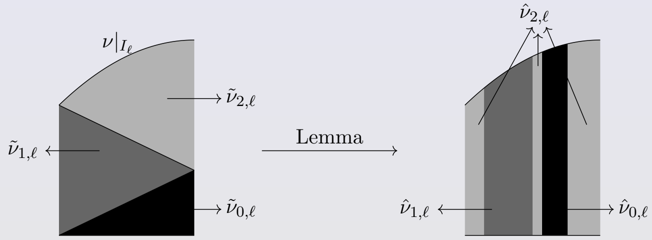


Figure: An illustration of the main idea to prove Theorem 3

Uniqueness

Theorem 4 (Uniqueness)

Let $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless. The following are equivalent:

(i) *The MT from μ to ν is unique.*

(ii) *The MMT from μ to ν is unique.*

(iii) *Let $\mu_a := \sum_{j\in\mathbb{N}} a_j \delta_{x_j}$ be the atomic part of μ, where $\{x_j\}_{j\in\mathbb{N}}$ are distinct. Then the shadows $S^\nu(a_j\delta_{x_j})$, $j \in \mathbb{N}$ are mutually singular and $\mu - \mu_a = \nu - \sum_{j\in\mathbb{N}} S^\nu(a_j\delta_{x_j})$.*

- μ and ν are both atomless: uniqueness $\iff$ $\mu = \nu$

Uniqueness

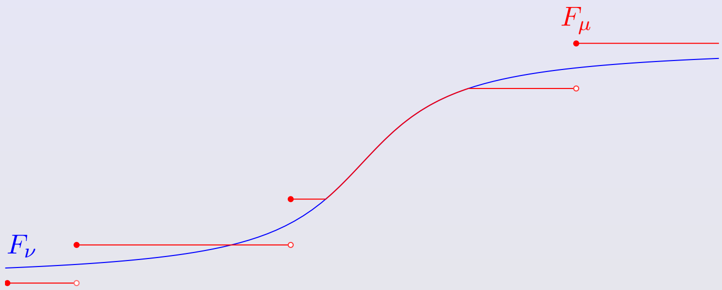


Figure: Distribution functions of μ, ν where the MMT (and MT) from μ to ν is unique

Uniqueness

Example 2 (MT exists; MMT does not)

Let μ and ν be two-point distributions satisfying $\mu \leqslant_{\mathrm{cx}} \nu$. Then there is a unique MT but there is no MMT unless $\mu = \nu$.

- In general, if μ, ν are discrete and $\mathrm{card}(\cdot)$ denotes the cardinality of the support, the existence of an MMT implies $(2\,\mathrm{card}((\mu-\nu)_+)) \vee \mathrm{card}(\mu) \leqslant \mathrm{card}(\nu)$

Example 3 (MMT is unique; MT is not)

Let μ be uniform on $\{2, 5\}$ and ν be uniform on $\{0, 3, 4, 7\}$. The unique MMT is given by transporting $\{2\}$ to $\{0, 4\}$ and $\{5\}$ to $\{3, 7\}$, while it is easy to see that there exist many MTs.

Optimizer

Proposition 2

Consider $\mu \leqslant_{\mathrm{cx}} \nu$ with ν atomless. For any strictly convex $f, g : \mathbb{R} \to \mathbb{R}$,

$$\mathcal{M}_M(\mu,\nu) = \underset{(X,Y)\in\Pi(\mu,\nu)}{\arg\min} \ \mathbb{E}\left[f(\mathbb{E}[Y|X] - X) - g(\mathbb{E}[X|Y])\right]. \quad (\Diamond)$$

- $f(x) = g(x) = x^2$: $(\Diamond)$ is equivalent to

$$\mathbb{E}\left[\mathbb{E}[Y|X]^2 - \mathbb{E}[X|Y]^2 - 2\mathbb{E}[XY]\right]$$

- This cost is not symmetric in X and Y
- The term $-2\mathbb{E}[XY]$ is essential: $\mathcal{M}_M(\mu,\nu)$ does not minimize $\mathbb{E}[\mathbb{E}[Y|X]^2 - \mathbb{E}[X|Y]^2]$ unless X is a constant

Backward deterministic martingales

Definition 1

A stochastic process $(X_n)_{n\in\mathbb{N}}$ is backward deterministic if $(X_j)_{j=1}^n$ is $\sigma(X_n)$-measurable for all $n \in \mathbb{N}$.

- $\sigma(X_n)$ is non-decreasing in n.
- A backward deterministic process is Markovian
- Perfect memory: its time-n value records all its history up to time n

Backward deterministic martingales

Corollary 2

Given any martingale $(Y_n)_{n\in\mathbb{N}}$ with atomless marginals, there exists a backward deterministic martingale $(X_n)_{n\in\mathbb{N}}$ such that $X_n \stackrel{\text{law}}{=} Y_n$ for all $n \in \mathbb{N}$.

- Although rare, the class of backward deterministic martingale is surprisingly "rich"

Open questions

Backward deterministic fake Brownian motions

- Let $(W_t)_{t\in[0,T]}$ be a martingale with marginal distribution $W_t \overset{\text{law}}{=} \mathrm{N}(0,t)$
- Fake Brownian motions
  - (continuous path) Beiglböck/Lowther/Pammer/Schachermayer'23 FS
- Does there exist such W that is also backward deterministic?
  - $(W_s)_{s\in[0,t]}$ is $\sigma(W_t)$-measurable for each $t\in[0,T]$

Open questions

Generalization to $\mathbb{R}^d$

- Seems highly challenging

Conjecture 1

Let μ, ν be probability measures on $\mathbb{R}^d$ satisfying $\mu \leqslant_{\mathrm{cx}} \nu$, and $\{\nu_x : x \in \mathbb{R}^d\}$ is a decomposition into irreducible components of (μ, ν). Suppose that ν_x is atomless for μ-a.e. $x \in \mathbb{R}^d$. Then $\mathcal{M}_M(\mu, \nu)$ is non-empty and weakly dense in $\mathcal{M}(\mu, \nu)$. If μ is discrete, it is also dense for the ∞-Wasserstein topology.

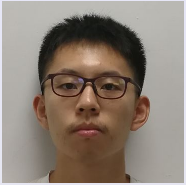
Open questions

Monge supermartingale/directional transport

- Existence of Monge directional transport: clear Nutz-W.'22 AAP
- Existence of Monge supermartingale transport: unclear
- Denseness: unclear

Thank you for your kind attention

Marcel Nutz
(Columbia)



Zhenyuan Zhang
(Stanford)

Takeaway 1. On an atomless probability space,

$$X \leqslant_{\mathrm{cx}} Y \iff X \overset{\mathrm{law}}{=} \mathbb{E}[Y|\mathcal{G}] \text{ for some } \sigma\text{-field } \mathcal{G}$$

Takeaway 2. For atomless ν and continuous and "non-exploding" c,

$$\text{optimal MT cost} = \text{optimal MMT cost}$$