Martingale Optimal Transport and Robust Finance

Marcel Nutz
Columbia University

(with Mathias Beiglböck and Nizar Touzi)

April 2015

Outline

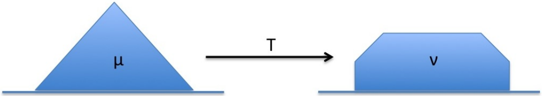
1. Monge–Kantorovich Optimal Transport

2. Martingale Optimal Transport

Monge Optimal Transport

Given:

- Probabilities μ, ν on $\mathbb{R}$.
- Reward (cost) function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.



Objective:

- Find a map $T : \mathbb{R} \to \mathbb{R}$ satisfying $\nu = T^{-1} \circ \mu$ such as to maximize the total reward,

$$\max_T \int f(x, T(x))\, \mu(dx).$$

Monge–Kantorovich Optimal Transport

Relaxation:

- Find a probability P on $\mathbb{R} \times \mathbb{R}$ with marginals μ, ν such as to maximize the reward:

$$\max_{P \in \Pi(\mu,\nu)} E^P[f(X,Y)], \quad \text{where} \quad \Pi(\mu,\nu) := \{P : \; P_1 = \mu, \; P_2 = \nu\}$$

  and $(X, Y) = \mathrm{Id}_{\mathbb{R}\times\mathbb{R}}$.

- $P \in \Pi(\mu,\nu)$ is a Monge transport if of the form $P = \mu \otimes \delta_{T(x)}$.

Example: Hoeffding–Frechet Coupling

Theorem: Let $f(x,y) = g(y-x)$ where g is strictly concave and sufficiently integrable. Then the optimal P is given by the Hoeffding–Frechet Coupling:

- P is the law of $((F_\mu)^{-1}, (F_\nu)^{-1})$ under the uniform measure on $[0,1]$.
- If μ is diffuse, P is of Monge type with $T = (F_\nu)^{-1} \circ F_\mu$.
- P is characterized by monotonicity:

$$\text{if } (x,y),(x',y') \in \operatorname{supp}(P) \text{ and if } x < x', \text{ then } y \leq y'.$$

Kantorovich Duality

- Buy $\varphi(X)$ at price $\mu(\varphi) := E^{\mu}[\varphi]$ and $\psi(Y)$ at $\nu(\psi)$ to superhedge,

$$f(X,Y) \leq \varphi(X) + \psi(Y).$$

- Then for all $P \in \Pi(\mu,\nu)$,

$$E^P[f(X,Y)] \leq E^P[\varphi(X) + \psi(Y)] = \mu(\varphi) + \nu(\psi).$$

- **Theorem** (Kantorovich, Kellerer): For any measurable $f \geq 0$,

$$\sup_{P \in \Pi(\mu,\nu)} E^P[f(X,Y)] = \inf_{\varphi,\psi} \mu(\varphi) + \nu(\psi)$$

and dual optimizers $\hat{\varphi}$, $\hat{\psi}$ exist.

Kantorovich Duality

- Buy $\varphi(X)$ at price $\mu(\varphi) := E^{\mu}[\varphi]$ and $\psi(Y)$ at $\nu(\psi)$ to superhedge,
$$f(X,Y) \leq \varphi(X) + \psi(Y).$$
- Then for all $P \in \Pi(\mu,\nu)$,
$$E^P[f(X,Y)] \leq E^P[\varphi(X) + \psi(Y)] = \mu(\varphi) + \nu(\psi).$$
- **Theorem** (Kantorovich, Kellerer): For any measurable $f \geq 0$,
$$\sup_{P\in\Pi(\mu,\nu)} E^P[f(X,Y)] = \inf_{\varphi,\psi} \mu(\varphi) + \nu(\psi)$$
and dual optimizers $\hat{\varphi}$, $\hat{\psi}$ exist.

Application: Fundamental Theorem of Optimal Transport

Let $\Gamma = \{(x,y) : \hat{\varphi}(x) + \hat{\psi}(y) = f(x,y)\}$ and $P \in \Pi(\mu,\nu)$. TFAE:

(1) P is optimal.

(2) $P(\Gamma) = 1$.

(3) $\mathrm{supp}(P)$ is f-cyclically monotone P-a.s.; i.e.,

$$\sum_{i=1}^{n} f(x_i, y_i) \geq \sum_{i=1}^{n} f(x_i, y_{\sigma(i)}) \quad \forall\, (x_i, y_i) \in \mathrm{supp}(P), \quad \sigma \in \mathrm{Perm}(n).$$

(1)(2) If $P(\Gamma) < 1$, then P charges $\{(x,y) : \hat{\varphi}(x) + \hat{\psi}(y) > f(x,y)\}$ and thus $\mu(\hat{\varphi}) + \nu(\hat{\psi}) > E^P[f(X,Y)]$.

(2)(1) If $P(\Gamma) = 1$, then $\mu(\hat{\varphi}) + \nu(\hat{\psi}) = E^P[f(X,Y)]$, hence $P, \hat{\varphi}, \hat{\psi}$ are optimal.

(2)(3) This argument even shows: if $\tilde{P}(\Gamma) = 1$, then $\tilde{P}$ is an optimal transport between its own marginals. Apply this with discrete $\tilde{P}$ $\Rightarrow \Gamma$ is cyclically monotone.

Application: Fundamental Theorem of Optimal Transport

Let $\Gamma = \{(x,y) : \hat{\varphi}(x) + \hat{\psi}(y) = f(x,y)\}$ and $P \in \Pi(\mu,\nu)$. TFAE:

(1) P is optimal.

(2) $P(\Gamma) = 1$.

(3) $\mathrm{supp}(P)$ is f-cyclically monotone P-a.s.; i.e.,

$$\sum_{i=1}^{n} f(x_i, y_i) \geq \sum_{i=1}^{n} f(x_i, y_{\sigma(i)}) \quad \forall\, (x_i, y_i) \in \mathrm{supp}(P), \quad \sigma \in \mathrm{Perm}(n).$$

(1)(2) If $P(\Gamma) < 1$, then P charges $\{(x,y) : \hat{\varphi}(x) + \hat{\psi}(y) > f(x,y)\}$ and thus $\mu(\hat{\varphi}) + \nu(\hat{\psi}) > E^P[f(X,Y)]$.

(2)(1) If $P(\Gamma) = 1$, then $\mu(\hat{\varphi}) + \nu(\hat{\psi}) = E^P[f(X,Y)]$, hence $P, \hat{\varphi}, \hat{\psi}$ are optimal.

(2)(3) This argument even shows: if $\tilde{P}(\Gamma) = 1$, then $\tilde{P}$ is an optimal transport between its own marginals. Apply this with discrete $\tilde{P}$ $\Rightarrow \Gamma$ is cyclically monotone.

Application: Fundamental Theorem of Optimal Transport

Let $\Gamma = \{(x,y) : \hat{\varphi}(x) + \hat{\psi}(y) = f(x,y)\}$ and $P \in \Pi(\mu,\nu)$. TFAE:

(1) P is optimal.

(2) $P(\Gamma) = 1$.

(3) $\mathrm{supp}(P)$ is f-cyclically monotone P-a.s.; i.e.,

$$\sum_{i=1}^{n} f(x_i, y_i) \geq \sum_{i=1}^{n} f(x_i, y_{\sigma(i)}) \quad \forall\, (x_i, y_i) \in \mathrm{supp}(P), \quad \sigma \in \mathrm{Perm}(n).$$

(1)(2) If $P(\Gamma) < 1$, then P charges $\{(x,y) : \hat{\varphi}(x) + \hat{\psi}(y) > f(x,y)\}$ and thus $\mu(\hat{\varphi}) + \nu(\hat{\psi}) > E^P[f(X,Y)]$.

(2)(1) If $P(\Gamma) = 1$, then $\mu(\hat{\varphi}) + \nu(\hat{\psi}) = E^P[f(X,Y)]$, hence $P, \hat{\varphi}, \hat{\psi}$ are optimal.

(2)(3) This argument even shows: if $\tilde{P}(\Gamma) = 1$, then $\tilde{P}$ is an optimal transport between its own marginals. Apply this with discrete $\tilde{P}$ $\Rightarrow$ Γ is cyclically monotone.

Application: Fundamental Theorem of Optimal Transport

Let $\Gamma = \{(x,y) : \hat{\varphi}(x) + \hat{\psi}(y) = f(x,y)\}$ and $P \in \Pi(\mu,\nu)$. TFAE:

(1) P is optimal.

(2) $P(\Gamma) = 1$.

(3) $\operatorname{supp}(P)$ is f-cyclically monotone P-a.s.; i.e.,

$$\sum_{i=1}^{n} f(x_i, y_i) \geq \sum_{i=1}^{n} f(x_i, y_{\sigma(i)}) \quad \forall\, (x_i, y_i) \in \operatorname{supp}(P), \quad \sigma \in \operatorname{Perm}(n).$$

(1)(2) If $P(\Gamma) < 1$, then P charges $\{(x,y) : \hat{\varphi}(x) + \hat{\psi}(y) > f(x,y)\}$ and thus $\mu(\hat{\varphi}) + \nu(\hat{\psi}) > E^P[f(X,Y)]$.

(2)(1) If $P(\Gamma) = 1$, then $\mu(\hat{\varphi}) + \nu(\hat{\psi}) = E^P[f(X,Y)]$, hence $P, \hat{\varphi}, \hat{\psi}$ are optimal.

(2)(3) This argument even shows: if $\tilde{P}(\Gamma) = 1$, then $\tilde{P}$ is an optimal transport between its own marginals. Apply this with discrete $\tilde{P}$ $\Rightarrow$ Γ is cyclically monotone.

Outline

1. Monge–Kantorovich Optimal Transport

2. Martingale Optimal Transport

Dynamic Hedging

- Dynamically tradable underlying $S = (S_0, S_1, S_2)$.
- Semi-static superhedge:

$$f((S_t)_t) \leq \varphi(S_1) + \psi(S_2) + H_0(S_1 - S_0) + H_1(S_2 - S_1).$$

- With $S_0 = 0$, $S_1 = X \sim \mu$, $S_2 = Y \sim \nu$ and normalization $H_0 = 0$:

$$f(X, Y) \leq \varphi(X) + \psi(Y) + h(X)(Y - X).$$

- Formally, duality with $P \in \Pi(\mu, \nu)$ satisfying the constraint that

$$E^P[h(X)(Y - X)] = 0 \quad \forall h; \quad \text{i.e.} \quad E^P[Y|X] = X.$$

Dynamic Hedging

- Dynamically tradable underlying $S = (S_0, S_1, S_2)$.
- Semi-static superhedge:

$$f((S_t)_t) \leq \varphi(S_1) + \psi(S_2) + H_0(S_1 - S_0) + H_1(S_2 - S_1).$$

- With $S_0 = 0$, $S_1 = X \sim \mu$, $S_2 = Y \sim \nu$ and normalization $H_0 = 0$:

$$f(X, Y) \leq \varphi(X) + \psi(Y) + h(X)(Y - X).$$

- Formally, duality with $P \in \Pi(\mu, \nu)$ satisfying the constraint that

$$E^P[h(X)(Y - X)] = 0 \quad \forall h; \quad \text{i.e.} \quad E^P[Y|X] = X.$$

Martingale Transport

- Set of martingale transports:

$$\mathcal{M}(\mu,\nu) = \{P \in \Pi(\mu,\nu) : E^P[Y|X] = X\}.$$

- **Theorem** (Strassen): $\mathcal{M}(\mu,\nu)$ is nonempty iff $\mu \leq_c \nu$; i.e.,

$$\mu(\phi) \leq \nu(\phi) \quad \forall\, \phi \text{ convex.}$$

- Martingale Optimal Transport problem: Given $\mu \leq_c \nu$,

$$\sup_{P \in \mathcal{M}(\mu,\nu)} E^P[f(X,Y)].$$

- Beiglböck, Henry-Labordère, Penkner; Galichon, Henry-Labordère, Touzi; Hobson; Beiglböck, Juillet; Acciaio, Bouchard, Brown, Cheridito, Cox, Davis, Dolinsky, Fahim, Huang, Källblad, Kupper, Lassalle, Martini, Neuberger, Obłój, Rogers, Schachermayer, Soner, Stebegg, Tan, Tangpi, Zaev, . . .

Martingale Transport

- Set of martingale transports:

$$\mathcal{M}(\mu,\nu) = \{P \in \Pi(\mu,\nu) :\ E^P[Y|X] = X\}.$$

- **Theorem** (Strassen): $\mathcal{M}(\mu,\nu)$ is nonempty iff $\mu \leq_c \nu$; i.e.,

$$\mu(\phi) \leq \nu(\phi) \quad \forall\, \phi \text{ convex.}$$

- Martingale Optimal Transport problem: Given $\mu \leq_c \nu$,

$$\sup_{P\in\mathcal{M}(\mu,\nu)} E^P[f(X,Y)].$$

- Beiglböck, Henry-Labordère, Penkner; Galichon, Henry-Labordère, Touzi; Hobson; Beiglböck, Juillet; Acciaio, Bouchard, Brown, Cheridito, Cox, Davis, Dolinsky, Fahim, Huang, Källblad, Kupper, Lassalle, Martini, Neuberger, Obłój, Rogers, Schachermayer, Soner, Stebegg, Tan, Tangpi, Zaev, . . .

Example: Beiglböck–Juillet Coupling

- **Theorem** (Beiglböck, Juillet): Let $f(x,y) = g(y-x)$ where g is differentiable, sufficiently integrable, and g' is strictly concave. Then the optimal P is given by the Left-Courtain Coupling:

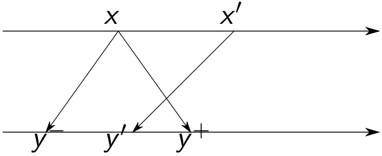
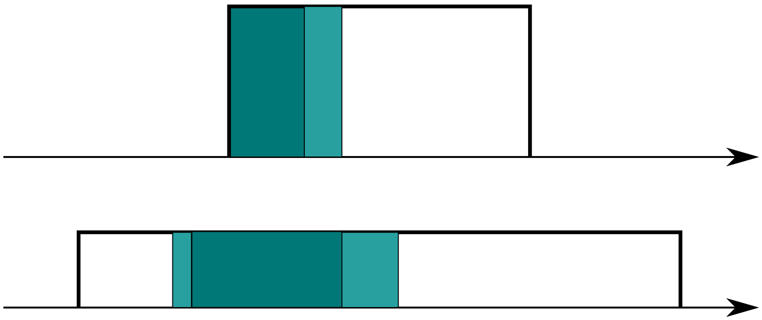


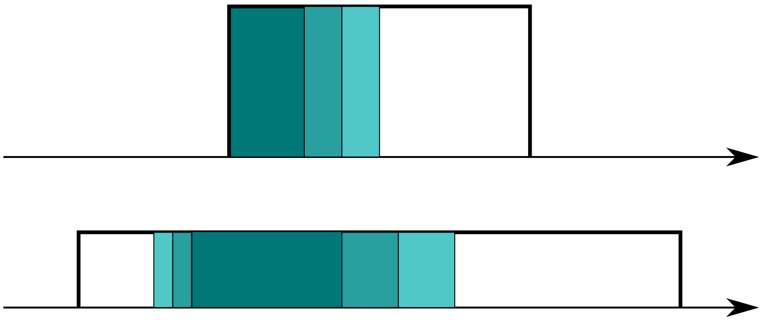
Figure : Forbidden Configuration.

Beiglböck–Juillet Coupling for Uniform Marginals

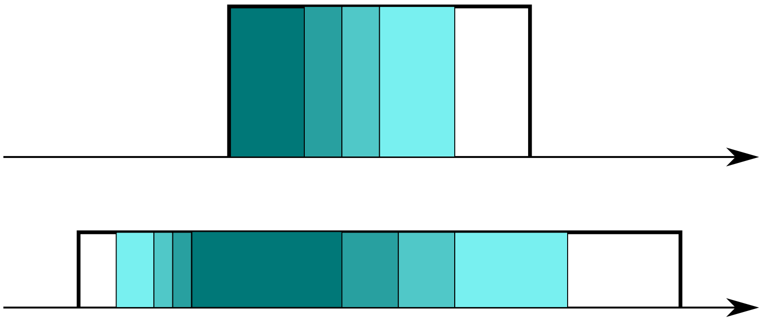
Beiglböck–Juillet Coupling for Uniform Marginals



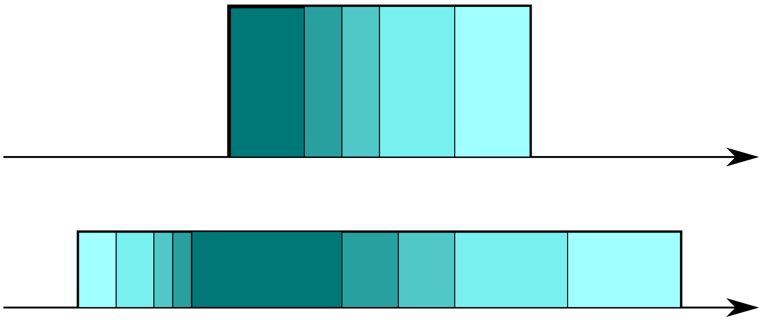
Beiglböck–Juillet Coupling for Uniform Marginals



Beiglböck–Juillet Coupling for Uniform Marginals



Beiglböck–Juillet Coupling for Uniform Marginals



Duality for Martingale Optimal Transport

In analogy to Monge–Kantorovich duality we want:

(1) No duality gap:

$$\sup_{P \in \mathcal{M}(\mu,\nu)} E^P[f(X,Y)] = \inf_{\varphi,\psi,h} \mu(\varphi) + \nu(\psi).$$

(2) Dual existence: $\hat{\varphi}$, $\hat{\psi}$, $\hat{h}$.

Theorem (Beiglböck, Henry-Labordère, Penkner):

- For upper semicontinuous $f \leq 0$, there is no duality gap.
- Dual existence fails in general, even if f is bounded, continuous and μ, ν are compactly supported.

Duality for Martingale Optimal Transport

In analogy to Monge–Kantorovich duality we want:

(1) No duality gap:

$$\sup_{P\in\mathcal{M}(\mu,\nu)} E^P[f(X,Y)] = \inf_{\varphi,\psi,h} \mu(\varphi)+\nu(\psi).$$

(2) Dual existence: $\hat{\varphi}$, $\hat{\psi}$, $\hat{h}$.

Theorem (Beiglböck, Henry-Labordère, Penkner):

- For upper semicontinuous $f\leq 0$, there is no duality gap.
- Dual existence fails in general, even if f is bounded, continuous and μ,ν are compactly supported.

An Example with Duality Gap

- Let f be the bounded, lower semicontinuous function

$$f(x,y) = \mathbf{1}_{x \neq y} = \begin{cases} 0 & \text{on the diagonal,} \\ 1 & \text{off the diagonal.} \end{cases}$$

- Let $\mu = \nu =$ Lebesgue measure on $[0,1]$.
- There exists a unique martingale transport P, concentrated on the diagonal ($T(x) = x$).
- Primal value: $\sup_{P \in \mathcal{M}(\mu,\nu)} E^P[f(X,Y)] = 0$.
- Dual optimizers exist, $\hat{\varphi} = 1$, $\hat{\psi} = 0$, $\hat{h} = 0$ but
- there is a duality gap: dual value $= 1 > 0$.

An Example with Duality Gap

- Let f be the bounded, lower semicontinuous function

$$f(x,y) = \mathbf{1}_{x \neq y} = \begin{cases} 0 & \text{on the diagonal,} \\ 1 & \text{off the diagonal.} \end{cases}$$

- Let $\mu = \nu =$ Lebesgue measure on $[0,1]$.

- There exists a unique martingale transport P, concentrated on the diagonal ($T(x) = x$).
- Primal value: $\sup_{P \in \mathcal{M}(\mu,\nu)} E^P[f(X,Y)] = 0$.
- Dual optimizers exist, $\hat{\varphi} = 1$, $\hat{\psi} = 0$, $\hat{h} = 0$ but
- there is a duality gap: dual value $= 1 > 0$.

Financial Intuition

Earlier work considered model uncertainty over a set $\mathcal{P}$ of real-world models (with finitely many traded options). $\mathcal{P}$-q.s. version of the FTAP:

Theorem (Bouchard, N.):
Let $\mathcal{M} = \{\text{calibrated martingale measures } Q \ll \mathcal{P}\}$. Then

$$\text{No arbitrage NA}(\mathcal{P}) \iff \mathcal{P} \text{ and } \mathcal{M} \text{ have same polar sets}$$

and under this condition, quasi-sure duality holds with existence.

Reverse-engineered:

To avoid arbitrage, neglect events not seen by $\mathcal{M}$.

Financial Intuition

Earlier work considered model uncertainty over a set $\mathcal{P}$ of real-world models (with finitely many traded options). $\mathcal{P}$-q.s. version of the FTAP:

Theorem (Bouchard, N.):
Let $\mathcal{M} = \{\text{calibrated martingale measures } Q \ll \mathcal{P}\}$. Then

$$\text{No arbitrage NA}(\mathcal{P}) \iff \mathcal{P} \text{ and } \mathcal{M} \text{ have same polar sets}$$

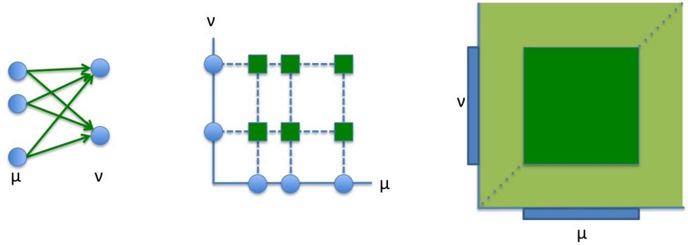
and under this condition, quasi-sure duality holds with existence.

Reverse-engineered:

To avoid arbitrage, neglect events not seen by $\mathcal{M}$.

Ordinary and Martingale OT: What is the Difference?

- In ordinary OT, all roads $x \to y$ can be used.
- E.g., in the discrete case, $\mu \times \nu$ already has full support.

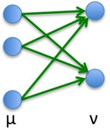


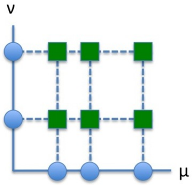
- **Theorem** (Kellerer): $A \subseteq \mathbb{R} \times \mathbb{R}$ is $\Pi(\mu,\nu)$-polar if and only if

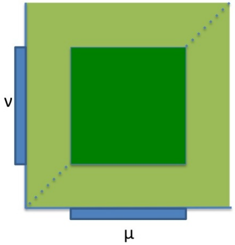
$$A \subseteq (N_1 \times \mathbb{R}) \cup (\mathbb{R} \times N_2), \quad \text{where} \quad \mu(N_1) = \nu(N_2) = 0.$$

Ordinary and Martingale OT: What is the Difference?

- In ordinary OT, all roads $x \to y$ can be used.
- E.g., in the discrete case, $\mu \times \nu$ already has full support.





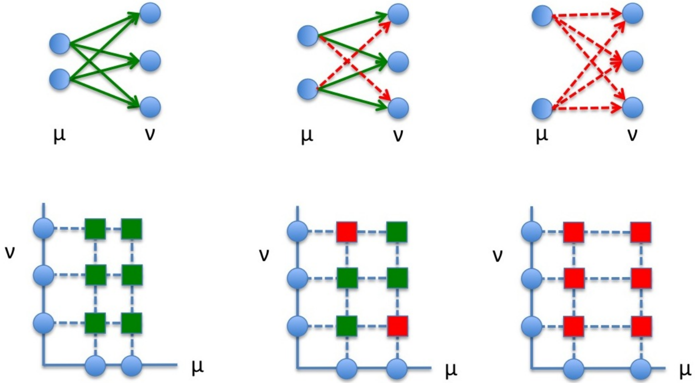


- **Theorem** (Kellerer): $A \subseteq \mathbb{R} \times \mathbb{R}$ is $\Pi(\mu,\nu)$-polar if and only if

$$A \subseteq (N_1 \times \mathbb{R}) \cup (\mathbb{R} \times N_2), \quad \text{where} \quad \mu(N_1) = \nu(N_2) = 0.$$

Obstructions for Martingale Transport

- In martingale OT, some roads $x \to y$ can be blocked.



Potential Functions

Potential $u_\mu(x) := \int |t-x|\,\mu(dt) = E[|X-x|]$ under any $P \in \mathcal{M}(\mu,\nu)$.

- $\mu \leq_c \nu \iff u_\mu \leq u_\nu$.
- If

  $$u_\mu(x) = u_\nu(x); \quad \text{i.e.} \quad E[|X-x|] = E[|Y-x|] \qquad (*),$$

  then x is a barrier for any martingale transport:

1. Jensen: $|X-x| = |E[Y|X]-x| = |E[Y-x|X]| \leq E[|Y-x|\,|X]$
2. Under $(*)$, it follows that $|X-x| = E[|Y-x|\,|X]$ a.s. Hence,

   $$E[|Y-x|\mathbf{1}_{X\geq x}] = E[|X-x|\mathbf{1}_{X\geq x}] = E[(X-x)\mathbf{1}_{X\geq x}] = E[(Y-x)\mathbf{1}_{X\geq x}]$$

   so that $Y \geq x$ a.s. on $\{X \geq x\}$.

$\rightarrow$ Partition $\mathbb{R}$ into intervals $\{u_\mu < u_\nu\}$.

Potential Functions

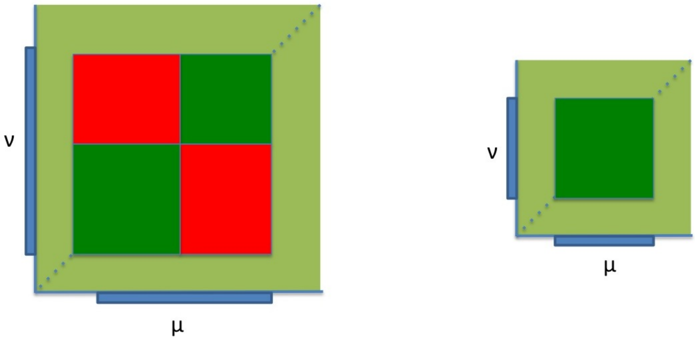
Potential $u_\mu(x) := \int |t-x|\,\mu(dt) = E[|X-x|]$ under any $P \in \mathcal{M}(\mu,\nu)$.

- $\mu \leq_c \nu \iff u_\mu \leq u_\nu$.
- If
$$u_\mu(x) = u_\nu(x); \quad \text{i.e.} \quad E[|X-x|] = E[|Y-x|] \qquad (*),$$
then x is a barrier for any martingale transport:

1. Jensen: $|X-x| = |E[Y|X]-x| = |E[Y-x|X]| \leq E[|Y-x|\,|X]$
2. Under $(*)$, it follows that $|X-x| = E[|Y-x|\,|X]$ a.s. Hence,
$$E[|Y-x|\mathbf{1}_{X\geq x}] = E[|X-x|\mathbf{1}_{X\geq x}] = E[(X-x)\mathbf{1}_{X\geq x}] = E[(Y-x)\mathbf{1}_{X\geq x}]$$
so that $Y \geq x$ a.s. on $\{X \geq x\}$.

$\rightarrow$ Partition $\mathbb{R}$ into intervals $\{u_\mu < u_\nu\}$.

Structure of $\mathcal{M}(\mu,\nu)$-polar Sets



Theorem: "These are precisely the $\mathcal{M}(\mu,\nu)$-polar sets."

Duality Result

Theorem

Let $f \geq 0$ be measurable and consider the quasi-sure relaxation of the dual problem:

$$f(X,Y) \leq \varphi(X) + \psi(Y) + h(X)(Y - X) \quad \mathcal{M}(\mu,\nu)\text{-q.s.}$$

Then,

(1) *there is no duality gap,*

(2) *dual optimizers $\hat{\varphi}$, $\hat{\psi}$, $\hat{h}$ exist.*

- The superhedge is pointwise on each component (e.g., $\mu = \delta_{x_0}$).
- Dual existence in the pointwise formulation typically fails as soon as there is more than one component.
- Application as in the FTOT.

Key Idea for the Proof

Core step: make almost-optimal φ_n, ψ_n converge.

- Control φ_n, ψ_n by a single, convex function χ_n.
- Suppose there is only one component; thus $\nu - \mu >_c 0$.
- After a normalization, $\chi_n(0) = \chi_n'(0) = 0$ and

$$0 \leq \int \chi_n \, d(\nu - \mu) \leq \text{const.}$$

- This bounds the convexity of χ_n.

$\Rightarrow$ Relative compactness of (χ_n).

$\Rightarrow$ Relative compactness of (φ_n, ψ_n); Komlos.

Conclusion

- The quasi-sure formulation ("model uncertainty") seems to be a natural setup for the Martingale Optimal Transport problem.

- One may expect this to be true for a larger class of problems; in particular, if discontinuous reward functions are involved or attainment is desired.

Thank you.

Conclusion

- The quasi-sure formulation ("model uncertainty") seems to be a natural setup for the Martingale Optimal Transport problem.
- One may expect this to be true for a larger class of problems; in particular, if discontinuous reward functions are involved or attainment is desired.

Thank you.