# Numerical Computation of Martingale Optimal Transport

*Gaoyue Guo, University of Oxford*

**joint work with Jan Obłój**

November 14, 2017 @ Marseilles

# Outline

A numerical scheme

An algorithm

# Objective

We aim to solve the martingale optimal transport (MOT) problem:

$$\mathsf{P}(\mu,\nu) \quad := \quad \sup_{\pi \in \mathcal{M}(\mu,\nu)} \quad \mathbb{E}_\pi\left[c(X,Y)\right],$$

where $\mu$ and $\nu$ are probability measures on $\mathcal{X} \equiv \mathbb{R}^d$ and $c : \mathcal{X}^2 \to \mathbb{R}$ is measurable.

# Objective

We aim to solve the martingale optimal transport (MOT) problem:

$$\mathsf{P}(\mu,\nu) \quad := \quad \sup_{\pi\in\mathcal{M}(\mu,\nu)} \quad \mathbb{E}_\pi\left[c(X,Y)\right],$$

where $\mu$ and $\nu$ are probability measures on $\mathcal{X} \equiv \mathbb{R}^d$ and $c : \mathcal{X}^2 \to \mathbb{R}$ is measurable.

- If $\mu = \sum_{i=1}^m \alpha_i\delta_{x_i}(dx)$ and $\nu = \sum_{j=1}^n \beta_j\delta_{y_j}(dy)$, then $\mathsf{P}(\mu,\nu)$ reduces to a linear programming (LP) problem;

# Objective

We aim to solve the martingale optimal transport (MOT) problem:

$$\mathsf{P}(\mu,\nu) \quad := \quad \sup_{\pi\in\mathcal{M}(\mu,\nu)} \quad \mathbb{E}_\pi\left[c(X,Y)\right],$$

where $\mu$ and $\nu$ are probability measures on $\mathcal{X} \equiv \mathbb{R}^d$ and $c : \mathcal{X}^2 \to \mathbb{R}$ is measurable.

- If $\mu = \sum_{i=1}^m \alpha_i \delta_{x_i}(dx)$ and $\nu = \sum_{j=1}^n \beta_j \delta_{y_j}(dy)$, then $\mathsf{P}(\mu,\nu)$ reduces to a linear programming (LP) problem;
- If $d = 1$ and $c(x,y) = h(x-y)$ or $c(x,y) = \varphi(x)\psi(y)$, then the map $(\mu,\nu) \mapsto \mathsf{P}(\mu,\nu)$ is continuous w.r.t. some topology of Wasserstein kind.

# A relaxed MOT problem

Let $\boldsymbol{X} = (X_k)_{1 \leq k \leq N}$ be the coordinate process on $\mathcal{X}^N$.

# A relaxed MOT problem

Let $\boldsymbol{X} = (X_k)_{1\le k\le N}$ be the coordinate process on $\mathcal{X}^N$.

For a vector of probability measures $\boldsymbol{\mu} = (\mu_k)_{1\le k\le N}$ and $\varepsilon \in \mathbb{R}_+$, we consider

$$\mathcal{M}_\varepsilon(\boldsymbol{\mu}) := \left\{\pi \in \mathcal{P}(\mathcal{X}^N) : \quad X_k \overset{\pi}{\sim} \mu_k \quad \text{and} \quad \left|\mathbb{E}_\pi\left[X_{k+1}\middle|X_1, \cdots X_k\right] - X_k\right| \le \varepsilon\right\},$$

OXFORD

European Research Council

# A relaxed MOT problem

Let $\boldsymbol{X} = (X_k)_{1 \le k \le N}$ be the coordinate process on $\mathcal{X}^N$.

For a vector of probability measures $\boldsymbol{\mu} = (\mu_k)_{1 \le k \le N}$ and $\varepsilon \in \mathbb{R}_+$, we consider

$$\mathcal{M}_\varepsilon(\boldsymbol{\mu}) := \left\{ \pi \in \mathcal{P}(\mathcal{X}^N) : \quad X_k \overset{\pi}{\sim} \mu_k \quad \text{and} \quad \left| \mathbb{E}_\pi \left[ X_{k+1} \middle| X_1, \cdots X_k \right] - X_k \right| \le \varepsilon \right\},$$

and define the corresponding optimization problem in the case of $c : \mathcal{X}^N \to \mathbb{R}$

$$\mathrm{P}_\varepsilon(\boldsymbol{\mu}) \quad := \quad \sup_{\pi \in \mathcal{M}_\varepsilon(\boldsymbol{\mu})} \mathbb{E}_\pi \left[ c(\boldsymbol{X}) \right] \quad \Big( \mathcal{M}_0(\boldsymbol{\mu}) \equiv \mathcal{M}(\boldsymbol{\mu}) \quad \text{and} \quad \mathrm{P}_0(\boldsymbol{\mu}) \equiv \mathrm{P}(\boldsymbol{\mu}) \Big).$$

# A relaxed MOT problem

Let $\boldsymbol{X} = (X_k)_{1 \le k \le N}$ be the coordinate process on $\mathcal{X}^N$.

For a vector of probability measures $\boldsymbol{\mu} = (\mu_k)_{1 \le k \le N}$ and $\varepsilon \in \mathbb{R}_+$, we consider

$$\mathcal{M}_\varepsilon(\boldsymbol{\mu}) \;:=\; \left\{ \pi \in \mathcal{P}(\mathcal{X}^N) : \quad X_k \overset{\pi}{\sim} \mu_k \quad \text{and} \quad \left| \mathbb{E}_\pi \left[ X_{k+1} \middle| X_1, \cdots X_k \right] - X_k \right| \;\le\; \varepsilon \right\},$$

and define the corresponding optimization problem in the case of $c : \mathcal{X}^N \to \mathbb{R}$

$$\mathrm{P}_\varepsilon(\boldsymbol{\mu}) \quad := \quad \sup_{\pi \in \mathcal{M}_\varepsilon(\boldsymbol{\mu})} \mathbb{E}_\pi \left[ c(\boldsymbol{X}) \right] \quad \Big( \mathcal{M}_0(\boldsymbol{\mu}) \equiv \mathcal{M}(\boldsymbol{\mu}) \quad \text{and} \quad \mathrm{P}_0(\boldsymbol{\mu}) \equiv \mathrm{P}(\boldsymbol{\mu}) \Big).$$

GOAL: COMPUTE NUMERICALLY $\mathrm{P}(\boldsymbol{\mu})$.

# Strassen's theorem

## Definition

*For any $\mu$ and $\nu$ admitting finite first moment, we say $\mu \preceq_\varepsilon \nu$ if*

$$\int_{\mathbb{R}^d} \left( \min_{z:\ |z-x|\leq\varepsilon} \psi(z) \right) \mu(dx) \quad \leq \quad \int_{\mathbb{R}^d} \psi(x)\nu(dx)$$

*holds for all convex function $\psi : \mathbb{R}^d \to \mathbb{R}$.*

# Strassen's theorem

## Definition

*For any $\mu$ and $\nu$ admitting finite first moment, we say $\mu \preceq_\varepsilon \nu$ if*

$$\int_{\mathbb{R}^d} \left( \min_{z:\ |z-x| \le \varepsilon} \psi(z) \right) \mu(dx) \quad \le \quad \int_{\mathbb{R}^d} \psi(x) \nu(dx)$$

*holds for all convex function $\psi : \mathbb{R}^d \to \mathbb{R}$.*

## Theorem

*Let $\mu_k$ have finite first moment for $k = 1, \cdots, N$. Then $\mathcal{M}_\varepsilon(\boldsymbol{\mu}) \neq \emptyset$ iff $\mu_k \preceq_\varepsilon \mu_{k+1}$ for $k = 1, \cdots, N-1$. In particular, $\boldsymbol{\mu}$ is called a PCOC if $\varepsilon = 0$.*

OXFORD

European Research Council

# A stability result

For $\boldsymbol{\mu} = (\mu_k)_{1\le k\le N}$ and $\boldsymbol{\nu} = (\nu_k)_{1\le k\le N}$, define

$$\mathcal{W}_1^{\oplus}(\boldsymbol{\mu}, \boldsymbol{\nu}) \quad := \quad \sum_{1\le k\le N} \mathcal{W}_1(\mu_k, \nu_k).$$

# A stability result

For $\boldsymbol{\mu} = (\mu_k)_{1\le k\le N}$ and $\boldsymbol{\nu} = (\nu_k)_{1\le k\le N}$, define

$$\mathcal{W}_1^{\oplus}(\boldsymbol{\mu},\boldsymbol{\nu}) \quad := \quad \sum_{1\le k\le N} \mathcal{W}_1(\mu_k,\nu_k).$$

## Theorem

*Let $(\boldsymbol{\mu}^n)_{n\ge 1}$ be a sequence converging to a PCOC $\boldsymbol{\mu}$ under $\mathcal{W}_1^{\oplus}$. Set $d_n := \mathcal{W}_1^{\oplus}(\boldsymbol{\mu}^n,\boldsymbol{\mu})$, then one has $\mathcal{M}_{d_n}(\boldsymbol{\mu}^n) \neq \emptyset$. If further $c$ is $L-$Lipschitz, then*

$$\mathsf{P}(\boldsymbol{\mu}) \quad \le \quad \mathsf{P}_{d_n}(\boldsymbol{\mu}^n) + Ld_n \quad \le \quad \mathsf{P}_{2d_n}(\boldsymbol{\mu}) + 2Ld_n.$$

*In particular* $\lim_{n\to\infty} \mathsf{P}_{d_n}(\boldsymbol{\mu}^n) = \mathsf{P}(\boldsymbol{\mu})$.

OXFORD

European Research Council

# Convergence rate: $\mu_1 \equiv \mu$ and $\mu_2 \equiv \nu$

## Theorem

*Let c be Lipschitz and satisfy* $\sup_{(x,y)\in\mathbb{R}^2} |\partial^2_{yy} c(x,y)| < +\infty$*. If* $\nu$ *has finite second moment, then there exists* $C > 0$ *s.t.*

$$\left|\mathrm{P}_{d_n}(\mu^n, \nu^n) \; - \; \mathrm{P}(\mu,\nu)\right| \quad \leq \quad C \inf_{R\in\mathbb{R}_+} \lambda_n(R),$$

*where*

$$\lambda_n(R) \quad := \quad Rd_n \; + \; \left(\int_{(R,+\infty)} (y-R)^2 \nu(dy) + \int_{(-\infty,-R)} (y+R)^2 \nu(dy)\right).$$

European Research Council

# Convergence rate: $\mu_1 \equiv \mu$ and $\mu_2 \equiv \nu$

## Theorem

*Let c be Lipschitz and satisfy* $\sup_{(x,y)\in\mathbb{R}^2} |\partial^2_{yy} c(x,y)| < +\infty$*. If* $\nu$ *has finite second moment, then there exists* $C > 0$ *s.t.*

$$\left|\mathsf{P}_{d_n}(\mu^n, \nu^n) - \mathsf{P}(\mu, \nu)\right| \leq C \inf_{R\in\mathbb{R}_+} \lambda_n(R),$$

*where*

$$\lambda_n(R) := R d_n + \left( \int_{(R,+\infty)} (y-R)^2 \nu(dy) + \int_{(-\infty,-R)} (y+R)^2 \nu(dy) \right).$$

## Remark

$\left|\mathsf{P}(\mu^n, \nu^n) - \mathsf{P}(\mu, \nu)\right| \leq \tilde{C} \inf_{R\in\mathbb{R}_+} \lambda_n(R)$ *holds if* $(\mu^n, \nu^n)$ *is a PCOC.*

OXFORD

European Research Council

# An explicit construction of PCOC

Define for $k \in \mathbb{Z}$

$$\mu^n(\{k/n\}) := \int_{[(k-1)/n,k/n)} (nx + 1 - k)d\mu + \int_{[k/n,(k+1)/n)} (1 + k - nx)d\mu,$$

$$\nu^n(\{k/n\}) := \int_{[(k-1)/n,k/n)} (nx + 1 - k)d\nu + \int_{[k/n,(k+1)/n)} (1 + k - nx)d\nu.$$

# An explicit construction of PCOC

Define for $k \in \mathbb{Z}$

$$\mu^n(\{k/n\}) := \int_{[(k-1)/n,k/n)} (nx+1-k)d\mu + \int_{[k/n,(k+1)/n)} (1+k-nx)d\mu,$$
$$\nu^n(\{k/n\}) := \int_{[(k-1)/n,k/n)} (nx+1-k)d\nu + \int_{[k/n,(k+1)/n)} (1+k-nx)d\nu.$$

## Lemma

(i) $(\mu^n, \nu^n)$ *are PCOCs supported on* $\{k/n\}_{k\in\mathbb{Z}}$*;*

(ii) $d_n \leq 2/n$.

OXFORD

erc European Research Council

# LP problem: $\mathrm{supp}(\nu) \subset [0, 1]$

Set $\alpha_k = \mu^n(\{k/n\})$, $\beta_k = \nu^n(\{k/n\})$ and $c_{i,j} = c(i/n, j/n)$. Then

## LP problem: $\text{supp}(\nu) \subset [0,1]$

Set $\alpha_k = \mu^n(\{k/n\})$, $\beta_k = \nu^n(\{k/n\})$ and $c_{i,j} = c(i/n, j/n)$. Then

$$\mathsf{P}(\mu^n, \nu^n) = \max_{p=(p_{i,j})_{0\le i,j\le n}} \sum_{i,j=0}^{n} p_{i,j} c_{i,j}$$

$$\text{s.t. } p \in \mathcal{C}_1 := \left\{ \sum_{j=0}^{n} p_{k,j} = \alpha_k, \text{ for } 0 \le k \le n \right\},$$

$$p \in \mathcal{C}_2 := \left\{ \sum_{i=0}^{n} p_{i,k} = \beta_k, \text{ for } 0 \le k \le n \right\},$$

$$p \in \mathcal{C}_3 := \left\{ \sum_{j=0}^{n} p_{k,j} j/n = \alpha_k k/n, \text{ for } 0 \le k \le n \right\}.$$

# Entropic regularization

Set $E(p) := \sum_{i,j=0}^{n} p_{i,j} c_{i,j}$ and $E_\varepsilon(p) = E(p) - \varepsilon \sum_{i,j=0}^{n} p_{i,j}\big(\log(p_{i,j}) - 1\big)$. In particular, $E_\varepsilon(p) := \varepsilon \mathrm{KL}(p|q)$, where

$$\mathrm{KL}(p|q) := \sum_{i,j=0}^{n} p_{i,j}\left[1 - \log\left(\frac{p_{i,j}}{q_{i,j}}\right)\right] \text{ with } q_{i,j} = e^{c_{i,j}/\varepsilon} \text{ for } 0 \le i,j \le n.$$

# Entropic regularization

Set $E(p) := \sum_{i,j=0}^{n} p_{i,j} c_{i,j}$ and $E_\varepsilon(p) = E(p) - \varepsilon \sum_{i,j=0}^{n} p_{i,j}\big(\log(p_{i,j}) - 1\big)$. In particular, $E_\varepsilon(p) := \varepsilon \mathrm{KL}(p|q)$, where

$$\mathrm{KL}(p|q) := \sum_{i,j=0}^{n} p_{i,j}\left[1 - \log\left(\frac{p_{i,j}}{q_{i,j}}\right)\right] \text{ with } q_{i,j} = e^{c_{i,j}/\varepsilon} \text{ for } 0 \leq i,j \leq n.$$

Consider

$$\max_{p \in \mathcal{M}(\mu^n,\nu^n)} E_\varepsilon(p) \quad = \quad \varepsilon \max_{p \in \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3} \mathrm{KL}(p|q).$$

## Proposition

*There exists* $0 < C \leq 1 + 2\log(n)$ *s.t.*

$$0 \quad \leq \quad \max_{p \in \mathcal{M}(\mu^n,\nu^n)} E_\varepsilon(p) - \mathsf{P}(\mu^n, \nu^n) \quad \leq \quad C\varepsilon.$$

# Example

$\mu = \mathcal{U}([-1,1])$, $\nu = \mathcal{U}([-2,2])$, $c(x,y) = |x-y|$. It follows from Hobson and Neuberger that

$$\pi^*(dx,dy) \quad = \quad \left(\frac{1}{2}\delta_{\xi_+(x)}(dy) \;+\; \frac{1}{2}\delta_{\xi_-(x)}(dy)\right)\mu(dx)$$

where $\xi_\pm(x) = x \pm 1$. In particular, $\mathsf{P}(\mu,\nu) = 1$.

OXFORD

European Research Council

# Example

$\mu = \mathcal{U}([-1,1])$, $\nu = \mathcal{U}([-2,2])$, $c(x,y) = |x-y|$. It follows from Hobson and Neuberger that

$$\pi^*(dx,dy) \quad = \quad \left(\frac{1}{2}\delta_{\xi_+(x)}(dy) \, + \, \frac{1}{2}\delta_{\xi_-(x)}(dy)\right)\mu(dx)$$

where $\xi_\pm(x) = x \pm 1$. In particular, $\mathsf{P}(\mu,\nu) = 1$.

- $\alpha_{-n} = 1/4n$, $\alpha_k = 1/2n$ for $-n \leq k < n$, $\alpha_n = 1/4n$;
- $\beta_{-2n} = 1/8n$, $\alpha_k = 1/4n$ for $-2n \leq k < 2n$, $\alpha_{2n} = 1/8n$;
- $d_n = 2/n$.

# Illustration

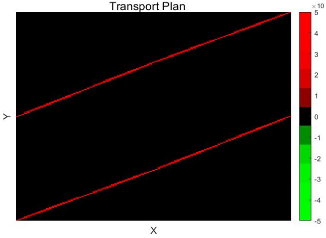


Figure 1: Optimal Transport Plan: $n = 50$

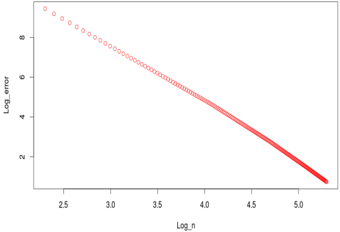


Figure 2: Logarithmic scale of error

# Conclusion

- Convergence of the iterative Bregman projection;
- Dual counterpart of the LP problem;
- Extension to $\boldsymbol{\mu} = (\mu_k)_{k \in I}$ for $I \subset \{1, \cdots, N\}$ and $\boldsymbol{\mu} = (\bar{\mu}_k)_{1 \leq k \leq N}$ for $\bar{\mu}_k = (\bar{\mu}_{k,i})_{1 \leq i \leq d}$.

**GG** & Jan Obłój: Computational Methods for Martingale Optimal Transport Problems. Preprint, arXiv: 1710.07911 [math.PR].

OXFORD

European Research Council

# Thank you very much!

OXFORD

erc European Research Council