Lecture 4: A glimpse of multi-marginal OT and applications

Luca Nenna
Matherials (INRIA) and CERMICS, ENPC, 02/02/2024, Champs sur Marne

(LMO) Université Paris-Saclay and INRIA-Saclay (ParMA)

Overview

1. A crash introduction to (multi-marginal) Optimal Transport
   - Quick recap on classical Optimal transport
   - Multi-marginal optimal transport
   - The real line case
2. Application I: MMOT for computing geodesics in the Wasserstein space
3. Application II: MMOT and the electron-electron repulsion
4. Entropic multi-marginal optimal transport
5. The asymptotics for the MOT_ε (with P. Pegon)
   - The upper bound
   - The lower bound
6. Another way to characterise (and solve) MOT_ε: an ODE approach (with B. Pass)
   - The ODE
   - The algorithm and some numerical results
   - An extension to general (entropic) multi-marginal problem

A crash introduction to (multi-marginal) Optimal Transport

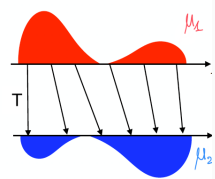
Classical Optimal Transportation Theory

Consider two probability measures μ_i on $X_i \subseteq \mathbb{R}^d$, and c a cost function (e.g. continuous or l.s.c.), the Optimal Transport (OT) problem is defined as follows

$$\mathrm{OT}_0 := \inf \left\{ \int_{\boldsymbol{X}} c(x_1, x_2) \mathrm{d}\gamma(x_1, x_2) \mid \gamma \in \Pi(\mu_1, \mu_2) \right\} \tag{1}$$

where $\Pi(\mu_1, \mu_2)$ denotes the set of couplings $\gamma(x_1, x_2) \in \mathcal{P}(\boldsymbol{X})$ having μ_1 and μ_2 as marginals.

- Solution à la Monge the transport plan γ is deterministic (or à la Monge) if $\gamma = (Id, T)_\sharp \mu_1$ where $T_\sharp \mu_1 = \mu_2$.



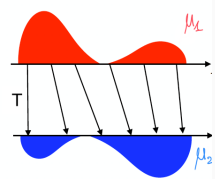
Classical Optimal Transportation Theory

Consider two probability measures μ_i on $X_i \subseteq \mathbb{R}^d$, and c a cost function (e.g. continuous or l.s.c.), the Optimal Transport (OT) problem is defined as follows

$$\mathrm{OT}_0 := \inf \left\{ \int_{\boldsymbol{X}} c(x_1, x_2) \mathrm{d}\gamma(x_1, x_2) \mid \gamma \in \Pi(\mu_1, \mu_2) \right\} \tag{1}$$

where $\Pi(\mu_1, \mu_2)$ denotes the set of couplings $\gamma(x_1, x_2) \in \mathcal{P}(\boldsymbol{X})$ having μ_1 and μ_2 as marginals.

- Solution à la Monge the transport plan γ is deterministic (or à la Monge) if $\gamma = (Id, T)_\sharp \mu_1$ where $T_\sharp \mu_1 = \mu_2$.



- Duality:

$$\sup \left\{ \mathcal{J}(\varphi_1, \varphi_2) \mid (\varphi_1, \varphi_2) \in \mathcal{K} \right\}. \tag{2}$$

where

$$\mathcal{J}(\varphi_1, \varphi_2) := \int_{X_1} \varphi_1 \mathrm{d}\mu_1 + \int_{X_2} \varphi_2 \mathrm{d}\mu_2$$

and $\mathcal{K}$ is the set of bounded and continuous functions (φ_1, φ_2) such that $\varphi_1(x_1) + \varphi(x_2) \leq c(x_1, x_2)$.

The Multi-Marginal Optimal Transportation

Take **(1)** m probability measures $\mu_i \in \mathcal{P}(X_i)$; **(2)** c a cost function. Then the multi-marginal OT problem reads as:

The Multi-Marginal Optimal Transportation

Take **(1)** m probability measures $\mu_i \in \mathcal{P}(X_i)$; **(2)** c a cost function. Then the multi-marginal OT problem reads as:

Multi-Marginal Optimal Transport problem

It reads as:

$$\mathrm{MOT}_0 := \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \int_{\boldsymbol{X}} c(x_1, \ldots, x_m) \mathrm{d}\gamma(x_1, \ldots, x_m) \tag{3}$$

where $\Pi(\mu_1, \ldots, \mu_m)$ denotes the set of couplings $\gamma(x_1, \ldots, x_m)$ having μ_i as marginals.

The Multi-Marginal Optimal Transportation

Take **(1)** m probability measures $\mu_i \in \mathcal{P}(X_i)$; **(2)** c a cost function. Then the multi-marginal OT problem reads as:

Multi-Marginal Optimal Transport problem

It reads as:

$$\mathrm{MOT}_0 := \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \int_{\boldsymbol{X}} c(x_1, \ldots, x_m) \mathrm{d}\gamma(x_1, \ldots, x_m) \tag{3}$$

where $\Pi(\mu_1, \ldots, \mu_m)$ denotes the set of couplings $\gamma(x_1, \ldots, x_m)$ having μ_i as marginals.

- Solution à la Monge: $\gamma = (Id, T_2, \ldots, T_m)_\sharp \mu_1$ where $T_{i\sharp}\mu_1 = \mu_i$.

The Multi-Marginal Optimal Transportation

Take **(1)** m probability measures $\mu_i \in \mathcal{P}(X_i)$; **(2)** c a cost function. Then the multi-marginal OT problem reads as:

Multi-Marginal Optimal Transport problem

It reads as:

$$\mathrm{MOT}_0 := \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \int_{\boldsymbol{X}} c(x_1, \ldots, x_m) \mathrm{d}\gamma(x_1, \ldots, x_m) \tag{3}$$

where $\Pi(\mu_1, \ldots, \mu_m)$ denotes the set of couplings $\gamma(x_1, \ldots, x_m)$ having μ_i as marginals.

- Solution à la Monge: $\gamma = (Id, T_2, \ldots, T_m)_\sharp \mu_1$ where $T_{i\sharp}\mu_1 = \mu_i$.
- Duality: Both 2 and m marginal OT problems admit a useful dual formulation

The Multi-Marginal Optimal Transportation

Take **(1)** m probability measures $\mu_i \in \mathcal{P}(X_i)$; **(2)** c a cost function. Then the multi-marginal OT problem reads as:

Multi-Marginal Optimal Transport problem

It reads as:

$$\mathrm{MOT}_0 := \inf_{\gamma \in \Pi(\mu_1, \dots, \mu_m)} \int_{\boldsymbol{X}} c(x_1, \dots, x_m) \mathrm{d}\gamma(x_1, \dots, x_m) \tag{3}$$

where $\Pi(\mu_1, \dots, \mu_m)$ denotes the set of couplings $\gamma(x_1, \dots, x_m)$ having μ_i as marginals.

- Solution à la Monge: $\gamma = (Id, T_2, \dots, T_m)_\sharp \mu_1$ where $T_{i\sharp}\mu_1 = \mu_i$.
- Duality: Both 2 and m marginal OT problems admit a useful dual formulation

Why is it a difficult problem to treat?
Example: $m = 3$, $d = 1$, $\mu_i = \mathcal{L}_{[0,1]}$ $\forall i$ and $c(x_1, x_2, x_3) = |x_1 + x_2 + x_3|^2$.

- Uniqueness fails **(Simone Di Marino, Gerolin, and Luca Nenna 2017)**;
- $\exists$ T_i optimal, are not differentiable at any point and they are fractal maps **ibid., Thm 4.6**

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **(Agueh and G. Carlier 2011)**): statistics, machine learning, image processing;

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **(Agueh and G. Carlier 2011)**): statistics, machine learning, image processing;
- Matching for teams problem (see **(Guillaume Carlier and Ekeland 2010)**): economics.

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **(Agueh and G. Carlier 2011)**): statistics, machine learning, image processing;
- Matching for teams problem (see **(Guillaume Carlier and Ekeland 2010)**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \ldots, x_m)$ returns the probability of finding electrons at position $x_1, \ldots, x_m$;

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **(Agueh and G. Carlier 2011)**): statistics, machine learning, image processing;
- Matching for teams problem (see **(Guillaume Carlier and Ekeland 2010)**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \ldots, x_m)$ returns the probability of finding electrons at position $x_1, \ldots, x_m$;
- Incompressible Euler Equations **(Brenier 1989)** : $\gamma(\omega)$ gives "the mass of fluid" which follows a path ω. See also **(Jean-David Benamou, Guillaume Carlier, and Luca Nenna 2018)**.

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **(Agueh and G. Carlier 2011)**): statistics, machine learning, image processing;
- Matching for teams problem (see **(Guillaume Carlier and Ekeland 2010)**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \ldots, x_m)$ returns the probability of finding electrons at position $x_1, \ldots, x_m$;
- Incompressible Euler Equations **(Brenier 1989)** : $\gamma(\omega)$ gives "the mass of fluid" which follows a path ω. See also **(Jean-David Benamou, Guillaume Carlier, and Luca Nenna 2018)**.
- Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **(Agueh and G. Carlier 2011)**): statistics, machine learning, image processing;
- Matching for teams problem (see **(Guillaume Carlier and Ekeland 2010)**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \ldots, x_m)$ returns the probability of finding electrons at position $x_1, \ldots, x_m$;
- Incompressible Euler Equations **(Brenier 1989)** : $\gamma(\omega)$ gives "the mass of fluid" which follows a path ω. See also **(Jean-David Benamou, Guillaume Carlier, and Luca Nenna 2018)**.
- Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;
- Risk measures **(Ennaji, Mérigot, Luca Nenna, and Pass 2022)**

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **(Agueh and G. Carlier 2011)**): statistics, machine learning, image processing;
- Matching for teams problem (see **(Guillaume Carlier and Ekeland 2010)**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \ldots, x_m)$ returns the probability of finding electrons at position $x_1, \ldots, x_m$;
- Incompressible Euler Equations **(Brenier 1989)** : $\gamma(\omega)$ gives "the mass of fluid" which follows a path ω. See also **(Jean-David Benamou, Guillaume Carlier, and Luca Nenna 2018)**.
- Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;
- Risk measures **(Ennaji, Mérigot, Luca Nenna, and Pass 2022)**
- Martingale transport, etc

The real line case

Given a probability measure $\mu \in \mathcal{P}(\mathbb{R})$, its inverse cumulative function (c.d.f) of as

$$F_\mu^{-1}(m) = \inf\{x \in \mathbb{R} \mid F_\mu(x) \geq m\} \text{ where } F_\mu(x) = \mu((-\infty, x]).$$

Theorem (see (Filippo Santambrogio 2015; Rachev and Rüschendorf 1998))

Let the cost function satisfies the condition

$$c(x', y') - c(x, y') - c(x', y) + c(x, y) \leq 0,$$

for $x' \geq x$, $y' \geq y$. Then the optimal transport plan γ is of the form $\gamma = (F_{\mu_1}^{-1}, F_{\mu_2}^{-1})_\sharp \mathrm{Leb}_{[0,1]}$

The real line case

Given a probability measure $\mu \in \mathcal{P}(\mathbb{R})$, its inverse cumulative function (c.d.f) of as

$$F_\mu^{-1}(m) = \inf\{x \in \mathbb{R} \mid F_\mu(x) \geq m\} \text{ where } F_\mu(x) = \mu((-\infty, x]).$$

Theorem (see (Filippo Santambrogio 2015; Rachev and Rüschendorf 1998))

Let the cost function satisfies the condition

$$c(x', y') - c(x, y') - c(x', y) + c(x, y) \leq 0,$$

for $x' \geq x$, $y' \geq y$. Then the optimal transport plan γ is of the form $\gamma = (F_{\mu_1}^{-1}, F_{\mu_2}^{-1})_\sharp \mathrm{Leb}_{[0,1]}$

- Question: can we extend this result to the multi-marginal case?

The real line case

Given a probability measure $\mu \in \mathcal{P}(\mathbb{R})$, its inverse cumulative function (c.d.f) of as

$$F_\mu^{-1}(m) = \inf\{x \in \mathbb{R} \mid F_\mu(x) \geq m\} \text{ where } F_\mu(x) = \mu((-\infty, x]).$$

Theorem (see (Filippo Santambrogio 2015; Rachev and Rüschendorf 1998))

Let the cost function satisfies the condition

$$c(x', y') - c(x, y') - c(x', y) + c(x, y) \leq 0,$$

for $x' \geq x$, $y' \geq y$. Then the optimal transport plan γ is of the form $\gamma = (F_{\mu_1}^{-1}, F_{\mu_2}^{-1})_\sharp \mathrm{Leb}_{[0,1]}$

- Question: can we extend this result to the multi-marginal case?

Theorem ((Guillaume Carlier 2003))

Given $c \in \mathcal{C}^2(X_1 \times \cdots \times X_m)$ and strict submodular cost, that is $\partial^2_{x_i x_j} c < 0$ for all $i \neq j$. Then the optimal γ is of the form $\gamma = (F_{\mu_1}^{-1}, \cdots, F_{\mu_m}^{-1})_\sharp \mathrm{Leb}_{[0,1]}$

Definition (Compatibility)

We will say that s is compatible if

$$\partial^2_{x_i x_j} c[\partial^2_{x_k x_j} c]^{-1} \partial^2_{x_k x_i} c(x_1, \ldots, x_m) < 0,$$

for each $i, j, k = 1, \ldots, m$ and each $(x_1, \ldots, x_m) \in X_1 \times \cdots \times X_m$.

Definition (Compatibility)

We will say that s is compatible if

$$\partial^2_{x_i x_j} c[\partial^2_{x_k x_j} c]^{-1} \partial^2_{x_k x_i} c(x_1, \ldots, x_m) < 0,$$

for each $i, j, k = 1, \ldots, m$ and each $(x_1, \ldots, x_m) \in X_1 \times \cdots \times X_m$.

Rmk: Note that if c compatible, the condition above implies that each $\partial^2_{x_i x_j} c \neq 0$ throughout thedomain $X_1 \times \ldots \times X_m$; continuity then yields that each $\partial^2_{x_i x_j} c$ is either always positive or always negative. Partition the set $\{1, 2, ...m\} = P_+ \cup P_-$ of indices into disjoint subsets P_+ and P_- such that $1 \in P_+$ and

- for each $i \neq j$, $\partial^2_{x_i x_j} c < 0$ throughout $X_1 \times \cdots \times X_m$ if either both i and j are in P_- or if both are in P_+;
- $\partial^2_{x_i x_j} c > 0$ throughout $X_1 \times \cdots \times X_m$ otherwise.

Comonotone plan for compatible cost

Definition

For a compatible c, we define the c - *comonotone coupling* by $\gamma = (G_1, G_2, \ldots, G_m)_{\#}\mathrm{Leb}_{[0,1]}$, where $G_1 = F^{-1}_{\mu_1}$ and for each $i = 1, 2, \ldots, m$

$$G_i(m) = \begin{cases} F^{-1}_{\mu_i}(m) & \text{if } i \in P_-, \\ F^{-1}_{\mu_i}(1-m) & \text{if } i \in P_+. \end{cases} \tag{4}$$

Comonotone plan for compatible cost

Definition

For a compatible c, we define the c - *comonotone coupling* by $\gamma = (G_1, G_2, \ldots, G_m)_{\#}\mathrm{Leb}_{[0,1]}$, where $G_1 = F_{\mu_1}^{-1}$ and for each $i = 1, 2, \ldots, m$

$$G_i(m) = \begin{cases} F_{\mu_i}^{-1}(m) & \text{if } i \in P_-, \\ F_{\mu_i}^{-1}(1-m) & \text{if } i \in P_+. \end{cases} \tag{4}$$

and in higher dimension?

For the quadratic cost or under some strong assumptions on the cost **(Gangbo and Swiech 1998; Pass 2012; Pass 2011; Pass 2015; Kim and Pass 2013)**.

Comonotone plan for compatible cost

Definition

For a compatible c, we define the c - *comonotone coupling* by $\gamma = (G_1, G_2, \ldots, G_m)_{\#}\mathrm{Leb}_{[0,1]}$, where $G_1 = F_{\mu_1}^{-1}$ and for each $i = 1, 2, \ldots, m$

$$G_i(m) = \begin{cases} F_{\mu_i}^{-1}(m) & \text{if } i \in P_-, \\ F_{\mu_i}^{-1}(1-m) & \text{if } i \in P_+. \end{cases} \tag{4}$$

and in higher dimension?

For the quadratic cost or under some strong assumptions on the cost **(Gangbo and Swiech 1998; Pass 2012; Pass 2011; Pass 2015; Kim and Pass 2013)**.

Good news: we can estimate the dimension of the support of the optimal plans.

A signature condition on the second mixed derivatives

- Consider **(1)** m probability measures $\mu_i \in \mathcal{P}(X_i)$ where X_i are $\mathcal{C}^2$ submanifolds of dimension d_i ; **(2)** $c \in \mathcal{C}^2(\mathrm{X})$; let **(3)** P be the set of partitions of $\{1, \ldots, m\}$ into two non empty disjoint subsets: $p := \{p_-, p_+\} \in P$;

A signature condition on the second mixed derivatives

- Consider **(1)** m probability measures $\mu_i \in \mathcal{P}(X_i)$ where X_i are $\mathcal{C}^2$ submanifolds of dimension d_i ; **(2)** $c \in \mathcal{C}^2(\mathrm{X})$; let **(3)** P be the set of partitions of $\{1, \ldots, m\}$ into two non empty disjoint subsets: $p := \{p_-, p_+\} \in P$;
- For each $p \in P$ we denote by g_p the bilinear form on X as

$$g_p = \sum_{i \in p_-, j \in p_+} D^2_{x_i, x_j} c + \sum_{i \in p_+, j \in p_-} D^2_{x_i, x_j} c.$$

A signature condition on the second mixed derivatives

- Consider **(1)** m probability measures $\mu_i \in \mathcal{P}(X_i)$ where X_i are $\mathcal{C}^2$ submanifolds of dimension d_i ; **(2)** $c \in \mathcal{C}^2(\mathrm{X})$; let **(3)** P be the set of partitions of $\{1, \ldots, m\}$ into two non empty disjoint subsets: $p := \{p_-, p_+\} \in P$;
- For each $p \in P$ we denote by g_p the bilinear form on X as

$$g_p = \sum_{i \in p_-, j \in p_+} D^2_{x_i, x_j} c + \sum_{i \in p_+, j \in p_-} D^2_{x_i, x_j} c.$$

- Define $G := \{\sum_{p \in P} t_p g_p \mid (t_p)_{p \in P} \in \Delta_P\}$ to be the convex hull generated by the g_p.

A signature condition on the second mixed derivatives

- Consider **(1)** m probability measures $\mu_i \in \mathcal{P}(X_i)$ where X_i are $\mathcal{C}^2$ submanifolds of dimension d_i ; **(2)** $c \in \mathcal{C}^2(\mathrm{X})$; let **(3)** P be the set of partitions of $\{1, \dots, m\}$ into two non empty disjoint subsets: $p := \{p_-, p_+\} \in P$;
- For each $p \in P$ we denote by g_p the bilinear form on X as

$$g_p = \sum_{i \in p_-, j \in p_+} D^2_{x_i, x_j} c + \sum_{i \in p_+, j \in p_-} D^2_{x_i, x_j} c.$$

- Define $G := \{\sum_{p \in P} t_p g_p \mid (t_p)_{p \in P} \in \Delta_P\}$ to be the convex hull generated by the g_p.

Theorem (Upper bound on the dimension of the support of the optimal plan (Pass 2011))

Let γ_0 a solution to MOT_0 *and suppose that at some point* $\mathrm{x} \in \mathrm{X}$, *the signature of some $g \in G$ is $(d^+(g), d^-(g), d^0(g))$. Then, there exists a neighbourhood N_{x} of* x *such that $N_{\mathrm{x}} \bigcap \mathrm{supp}(\gamma_0)$ is contained in a Lipschitz submanifold with dimension no greater than $\sum_i d_i - d^+(g)$.*

Application I: MMOT for computing geodesics in the Wasserstein space

The three formulations of quadratic Optimal Transport

Three formulations of Optimal Transport problem) with the quadratic cost :

- The static

$$\inf \left\{ \int_{X_1 \times X_2} \frac{1}{2} |x_1 - x_2|^2 d\gamma \mid \gamma \in \Pi(\mu_1, \mu_2) \right\}$$

- The dynamic (Lagrangian) ($C = H^1([0,1]; X)$ and $e_t : [0,1] \to X$)

$$\inf \left\{ \int_C \int_0^1 \frac{1}{2} |\dot{\omega}|^2 dt dQ(\omega) \mid Q \in \mathcal{P}(C),\ (e_0)_\sharp Q = \mu_1,\ (e_1)_\sharp Q = \mu_2 \right\}$$

- The dynamic (Eulerian), aka the Benamou-Brenier formulation

$$\inf \int \int \frac{1}{2} |v_t|^2 \rho_t dx dt \quad s.t.\ \partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$$

$$\rho(0, \cdot) = \mu_1,\ \rho(1, \cdot) = \mu_2$$

Some remarks and a MMOT formulation

Remarks:

- Consider the optimal solutions for the three formulations $\gamma^\star, Q^\star, \rho_t^\star$ then

$$\pi_t(x,y)_\sharp\gamma = (e_t)_\sharp Q = \rho_t^\star,$$

  where $\pi_t(x,y) = (1-t)x + ty$.

- if we discretise in time (let take $T+1$ time steps) the Lagrangian formulation and imposing that $\omega(t_i) = x_i$ ($t_i = i\frac{1}{T}$ for $i = 0, \cdots, T$) we get the following discrete (in time) MMOT problem

$$\inf \int_{X^T} \frac{1}{2T} \sum_{i=0}^{T} |x_{i+1} - x_i|^2 d\gamma(x_0, \cdots, x_T) \text{ s.t}$$
$$\gamma \in \mathcal{P}(X^{T+1}),\ \pi_{0,\sharp}\gamma = \mu_1,\ \pi_{T,\sharp}\gamma = \mu_2$$

The geodesic in 2D

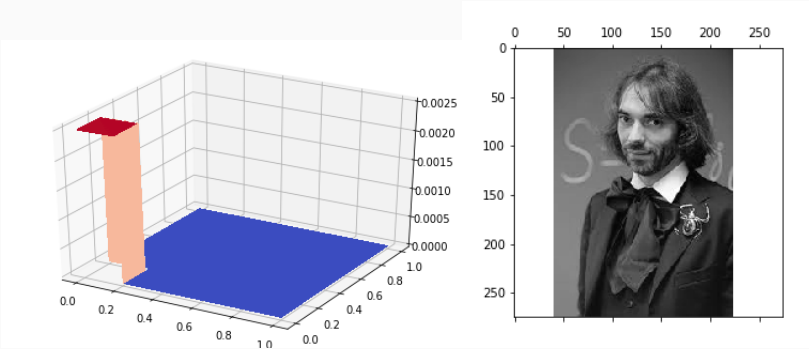


Figure 1: $t = 0$

The geodesic in 2D

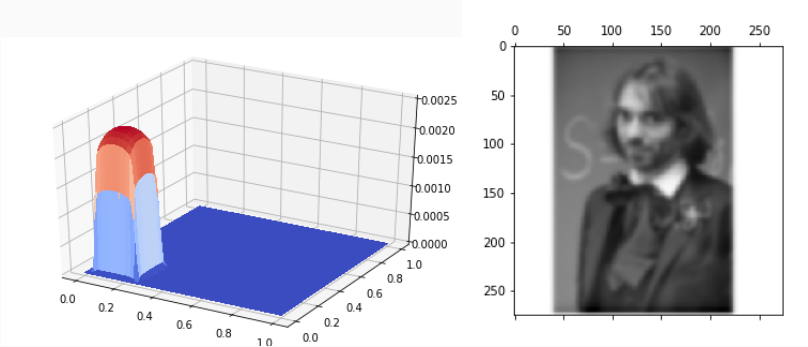


Figure 2: $t = \dfrac{1}{14}$

The geodesic in 2D

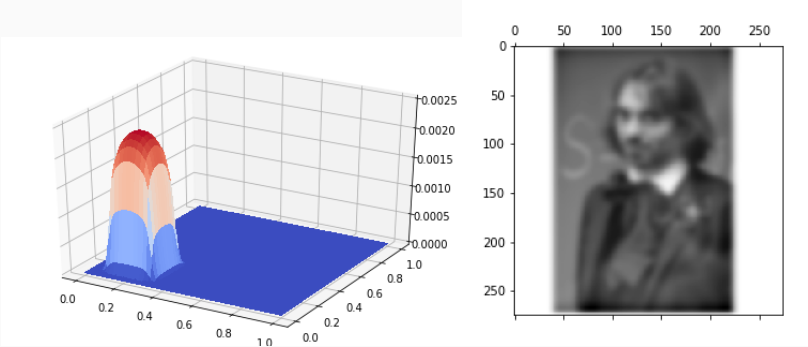


Figure 3: $t = \dfrac{2}{14}$

The geodesic in 2D

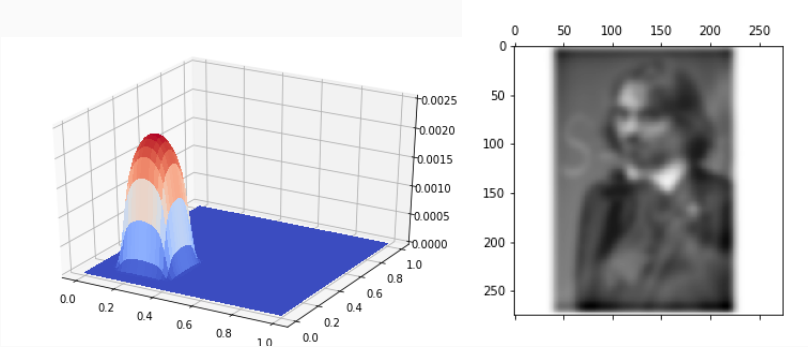


Figure 4: $t = \dfrac{3}{14}$

The geodesic in 2D

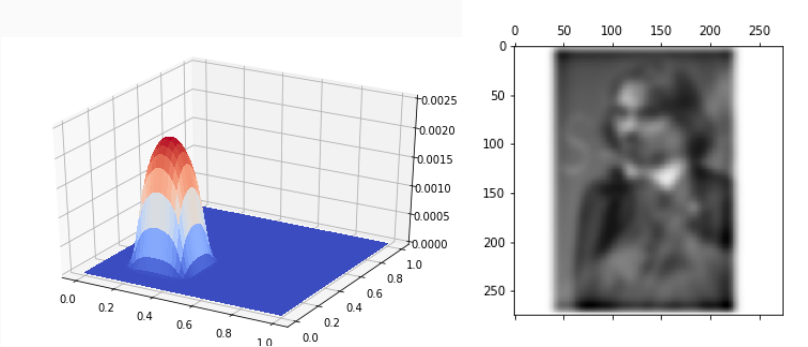


Figure 5: $t = \dfrac{4}{14}$

The geodesic in 2D

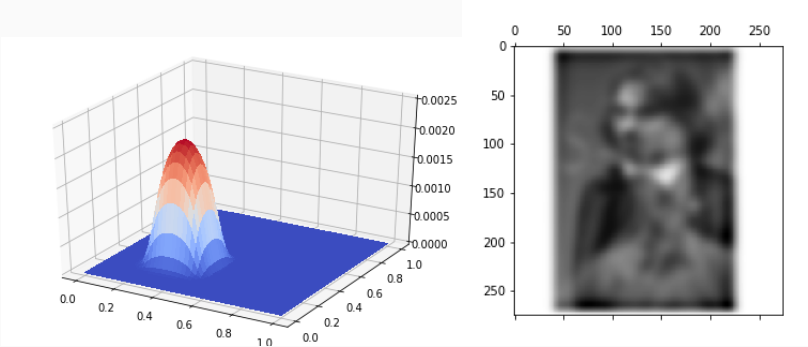


Figure 6: $t = \dfrac{5}{14}$

The geodesic in 2D

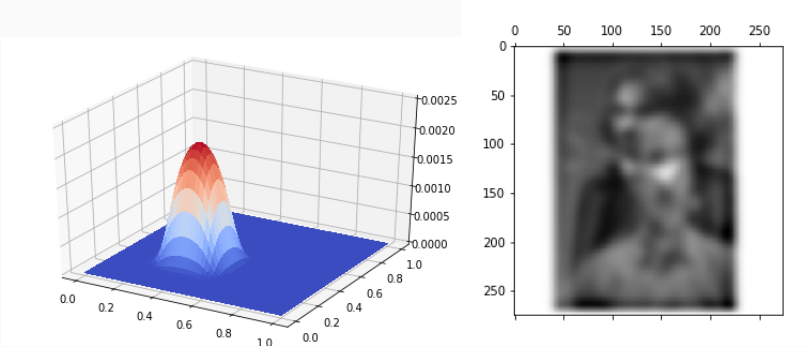


Figure 7: $t = \dfrac{6}{14}$

The geodesic in 2D

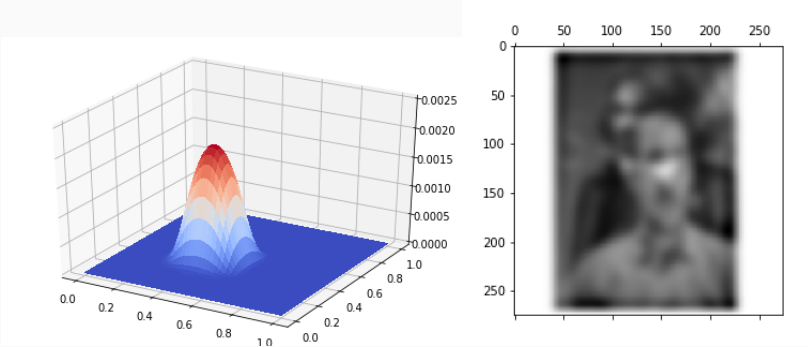


Figure 8: $t = \dfrac{7}{14}$

The geodesic in 2D

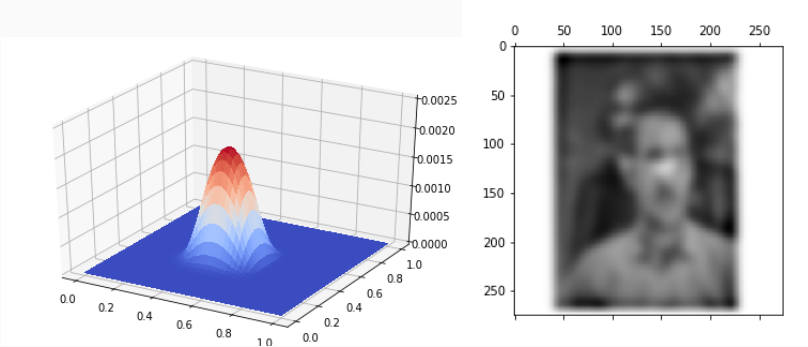


Figure 9: $t = \dfrac{8}{14}$

The geodesic in 2D

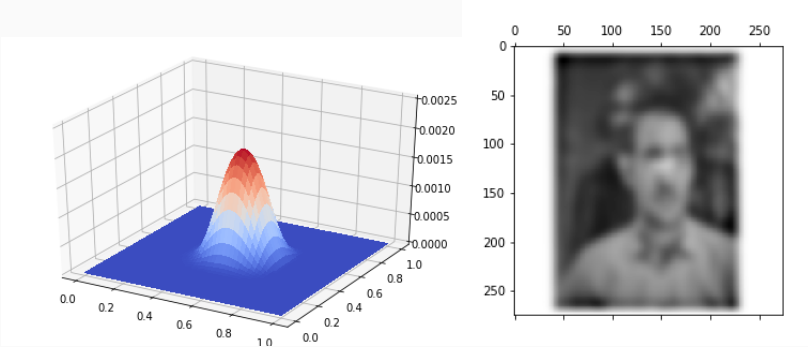


Figure 10: $t = \dfrac{9}{14}$

The geodesic in 2D

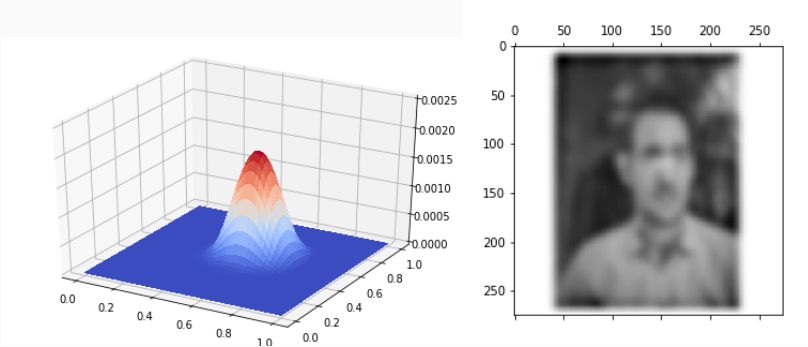


Figure 11: $t = \dfrac{10}{14}$

The geodesic in 2D

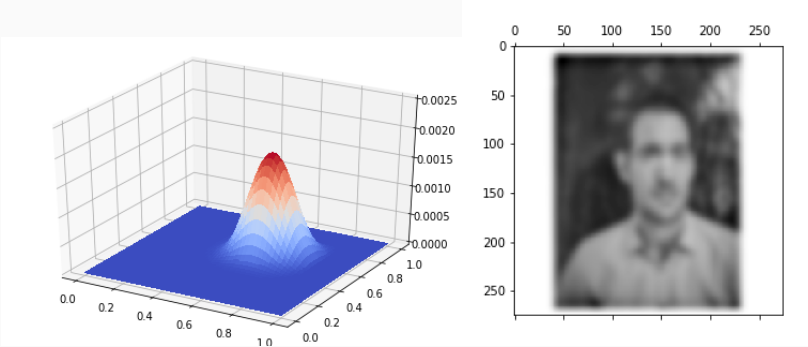


Figure 12: $t = \dfrac{11}{14}$

The geodesic in 2D

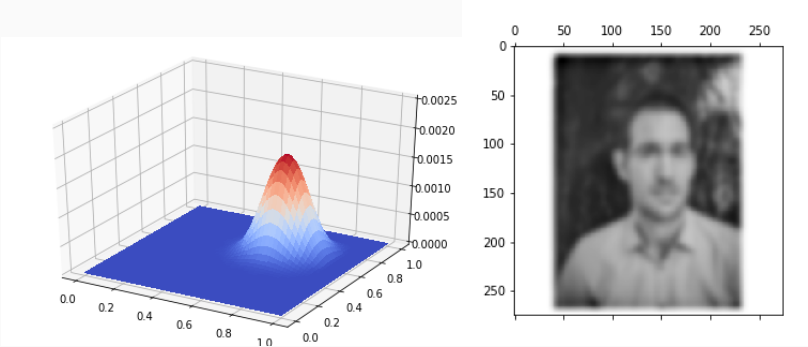


Figure 13: $t = \dfrac{12}{14}$

The geodesic in 2D

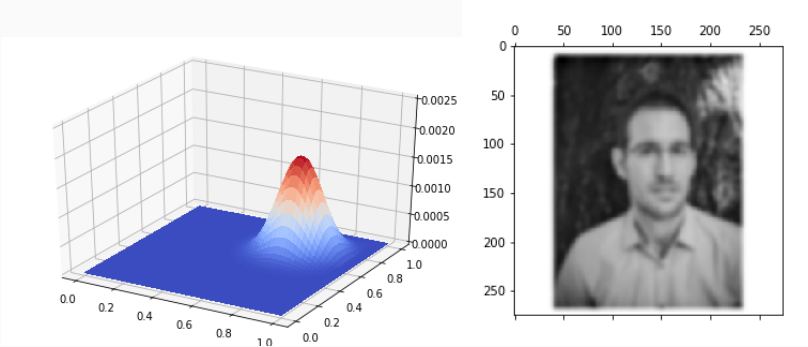


Figure 14: $t = \dfrac{13}{14}$

The geodesic in 2D

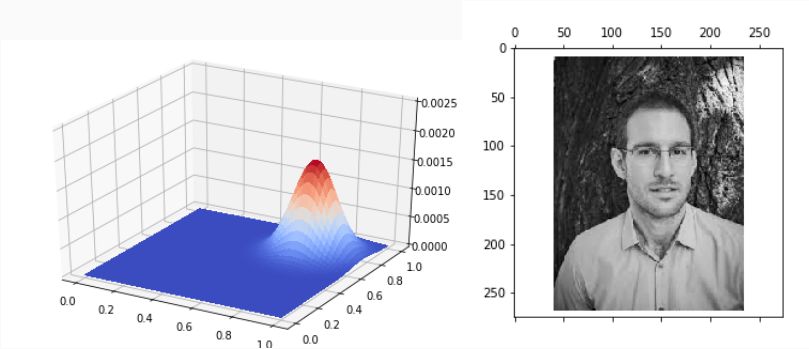


Figure 15: $t = 1$

Application II: MMOT and the electron-electron repulsion

The (optimal) electronic configuration via MMOT

Consider now the cost function

$$c(x_1, \cdots, x_m) = \sum_{i \neq j} \frac{1}{|x_i - x_j|},$$

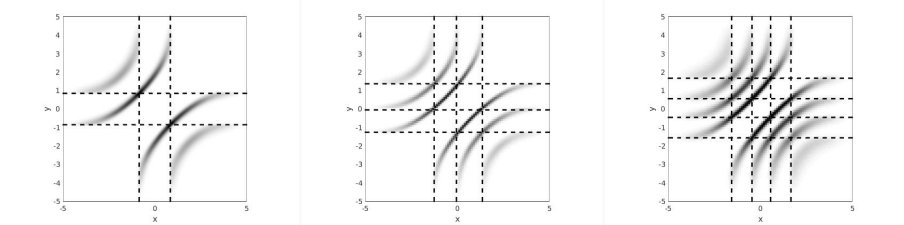
and $\mu_1 = \cdots = \mu_m = \rho$ (we refer to ρ as the electronic density) then the MMOT gives the electronic configuration (namely the optimal transport plan γ) which minimises the electron-electron repulsion.

Remarks:

- Since the cost is symmetric in the marginals then the dual problem reduces to look for only one potential;
- The cost is also radially symmetric which means that when ρ is radially symmetric then the $d = 3$ pb. reduces to a one dimensional pb;
- Existence of Monge solutions is still an open problem for $d > 1$;

Some simulations for $m = 3, 4, 5$ in 1D

We take the density $\rho(x) = \frac{m}{10}(1 + \cos(\frac{\pi}{5}x))$ and...



$N = 3$ $N = 4$ $N = 5$

Figure 16: *Support of the projected plan* $\gamma_{1,2}$

The transition from spread to deterministic plans for $m = 3$ and $d = 3$

Take $\rho_\alpha(r) = \alpha\rho_{Li}(r) + (1-\alpha)\rho_{exp}(r)$ and $\alpha \in [0,1]$ then...

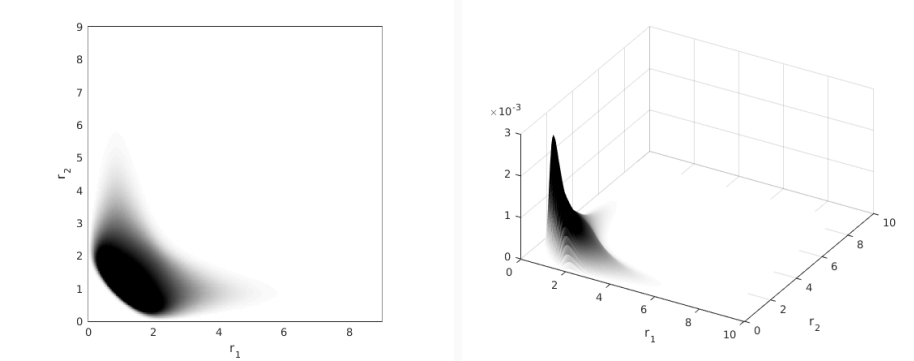


Figure 17: $\alpha = 0$

The transition from spread to deterministic plans for $m = 3$ and $d = 3$

Take $\rho_\alpha(r) = \alpha\rho_{Li}(r) + (1-\alpha)\rho_{exp}(r)$ and $\alpha \in [0,1]$ then...

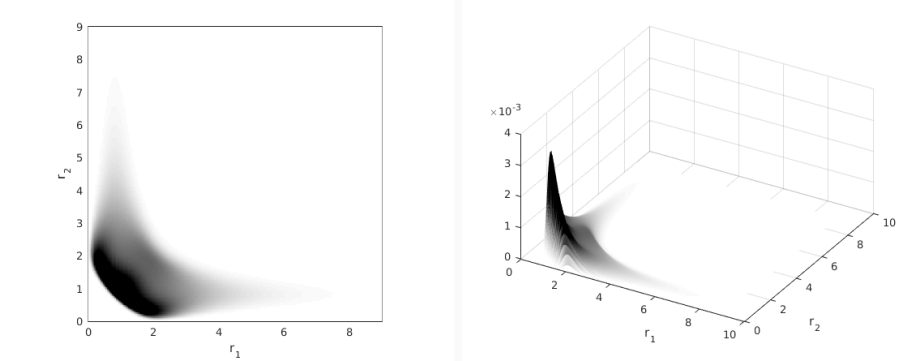


Figure 18: $\alpha = 0.1429$

The transition from spread to deterministic plans for $m = 3$ and $d = 3$

Take $\rho_\alpha(r) = \alpha\rho_{Li}(r) + (1-\alpha)\rho_{exp}(r)$ and $\alpha \in [0,1]$ then...

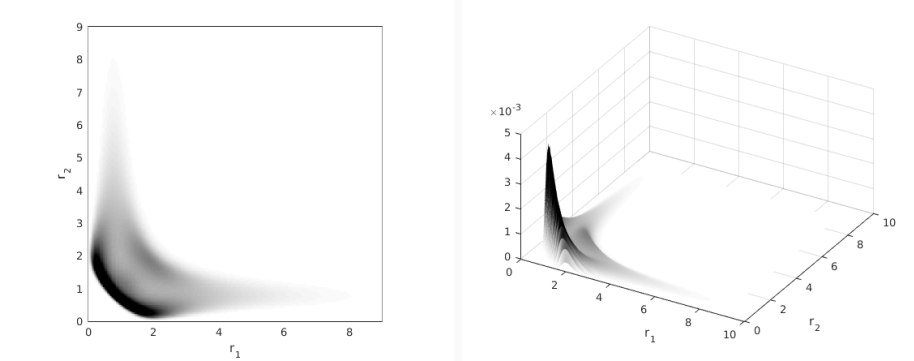


Figure 19: $\alpha = 0.2857$

The transition from spread to deterministic plans for $m = 3$ and $d = 3$

Take $\rho_\alpha(r) = \alpha\rho_{Li}(r) + (1-\alpha)\rho_{exp}(r)$ and $\alpha \in [0,1]$ then...

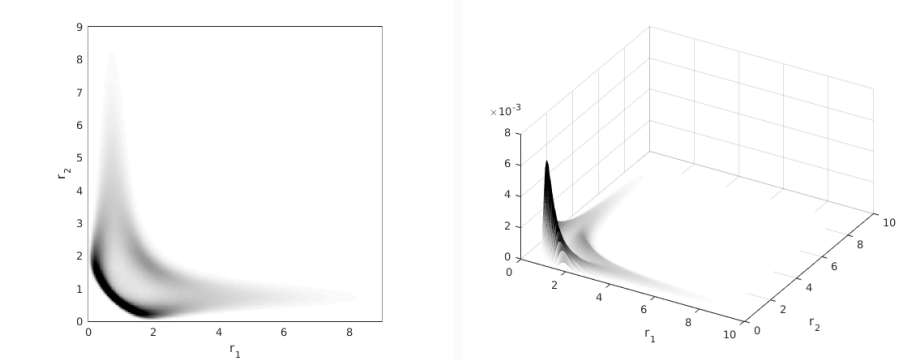


Figure 20: $\alpha = 0.4286$

The transition from spread to deterministic plans for $m = 3$ and $d = 3$

Take $\rho_\alpha(r) = \alpha\rho_{Li}(r) + (1-\alpha)\rho_{exp}(r)$ and $\alpha \in [0,1]$ then...

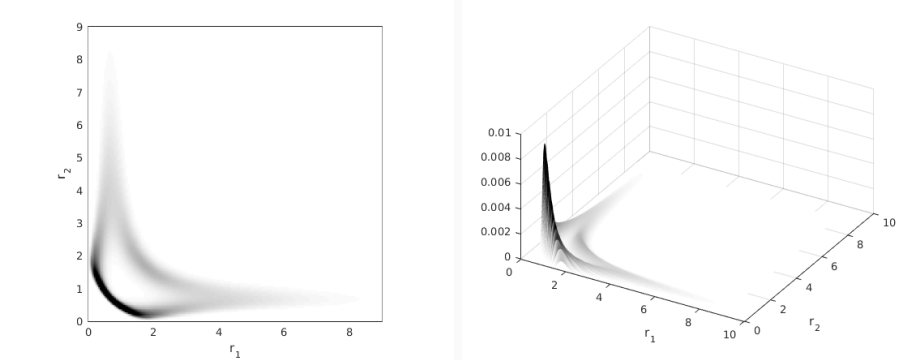
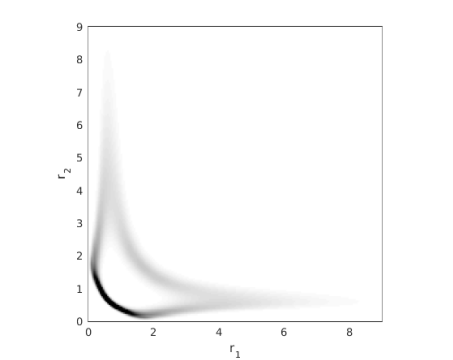


Figure 21: $\alpha = 0.5714$

The transition from spread to deterministic plans for $m = 3$ and $d = 3$

Take $\rho_\alpha(r) = \alpha\rho_{Li}(r) + (1-\alpha)\rho_{exp}(r)$ and $\alpha \in [0,1]$ then...



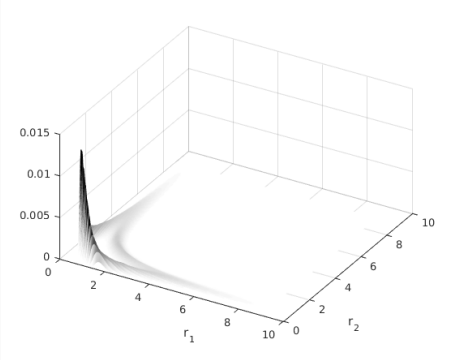
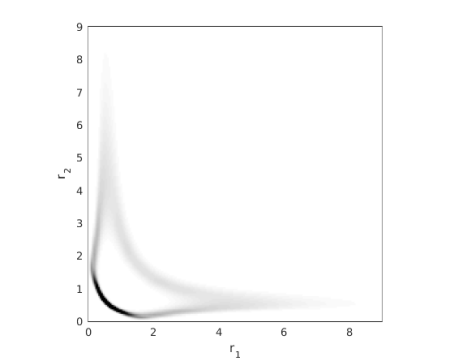


Figure 22: $\alpha = 0.7143$

The transition from spread to deterministic plans for $m = 3$ and $d = 3$

Take $\rho_\alpha(r) = \alpha\rho_{Li}(r) + (1-\alpha)\rho_{exp}(r)$ and $\alpha \in [0,1]$ then...



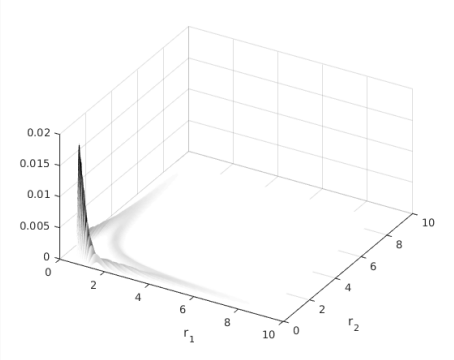
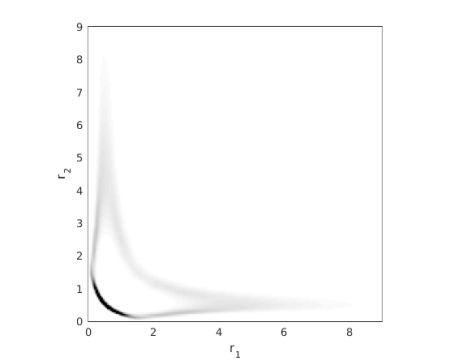


Figure 23: $\alpha = 0.8571$

The transition from spread to deterministic plans for $m = 3$ and $d = 3$

Take $\rho_\alpha(r) = \alpha\rho_{Li}(r) + (1-\alpha)\rho_{exp}(r)$ and $\alpha \in [0,1]$ then...



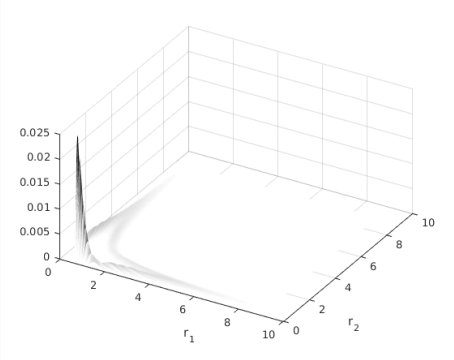


Figure 24: $\alpha = 1$

Entropic multi-marginal optimal transport

Definition of the problem

Consider

- $m \geq 2$ probability measures μ_i compactly supported on $\mathcal{C}^2$ submanifolds $X_i \subseteq \mathbb{R}^N$ of dim d_i ;
- a cost function $c : \mathrm{X} \to \mathbb{R}_+$ (e.g. continuous or lsc) where $\mathrm{X} := \times_i^m X_i$;

Definition of the problem

Consider

- $m \geq 2$ probability measures μ_i compactly supported on $\mathcal{C}^2$ submanifolds $X_i \subseteq \mathbb{R}^N$ of dim d_i ;
- a cost function $c : \mathsf{X} \to \mathbb{R}_+$ (e.g. continuous or lsc) where $\mathsf{X} := \times_i^m X_i$;

Entropic Multi-Marginal Optimal Transport problem

It reads as:

$$\mathsf{MOT}_\varepsilon := \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \left\{ \int_{\mathsf{X}} c(x_1, \ldots, x_m) \, \mathrm{d}\gamma(x_1, \ldots, x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\},$$

where

- $\Pi(\mu_1, \ldots, \mu_m)$ is the set of *couplings* $\gamma \in \mathcal{P}(\mathsf{X})$ having μ_i as marginals
- $\mathrm{Ent}(\gamma \mid \pi)$ is the Boltzmann-Shannon entropy, that is

$$\mathrm{Ent}(\gamma \mid \pi) = \int \rho \log \rho \mathrm{d}\pi, \text{ if } \gamma = \rho\pi.$$

Some useful remarks

- $\varepsilon = 0$ and $m = 2$. *Classical Optimal Transport problem*. Convex problem, but may have several solutions γ, with or without finite entropy!

Some useful remarks

- $\varepsilon = 0$ and $m = 2$. *Classical Optimal Transport problem*. Convex problem, but may have several solutions γ, with or without finite entropy!
- $\varepsilon > 0$. Strictly convex cost $\implies$ unique solution γ_ε with finite entropy.

Some useful remarks

- $\varepsilon = 0$ and $m = 2$. *Classical Optimal Transport problem*. Convex problem, but may have several solutions γ, with or without finite entropy!
- $\varepsilon > 0$. Strictly convex cost $\implies$ unique solution γ_ε with finite entropy.
- It admits a dual problem

$$\mathrm{MOT}_\varepsilon = \sup \left\{ \sum_{i=1}^{m} \int_{X_i} \varphi_i(x_i) \mathrm{d}\mu_i - \varepsilon \log \left(\int_{\boldsymbol{X}} e^{\frac{\sum_{i=1}^{m} \varphi_i(x_i) - c(\boldsymbol{x})}{\varepsilon}} \mathrm{d} \otimes_{i=1}^{m} \mu_i \right) \mid \varphi_i \in \mathcal{C}_b(X_i) \right\}.$$

Some useful remarks

- $\varepsilon = 0$ and $m = 2$. *Classical Optimal Transport problem*. Convex problem, but may have several solutions γ, with or without finite entropy!
- $\varepsilon > 0$. Strictly convex cost $\implies$ unique solution γ_ε with finite entropy.
- It admits a dual problem

$$\mathrm{MOT}_\varepsilon = \sup \left\{ \sum_{i=1}^m \int_{X_i} \varphi_i(x_i) \mathrm{d}\mu_i - \varepsilon \log \left(\int_{\boldsymbol{X}} e^{\frac{\sum_{i=1}^m \varphi_i(x_i) - c(\boldsymbol{x})}{\varepsilon}} \mathrm{d} \otimes_{i=1}^m \mu_i \right) \mid \varphi_i \in \mathcal{C}_b(X_i) \right\}.$$

- The solution γ_ε is "almost" explicit

$$\gamma_\varepsilon = \exp \left(\frac{\oplus_{i=1}^m \varphi_i^\varepsilon - c}{\varepsilon} \right) \otimes_{i=1}^m \mu_i.$$

Some useful remarks

- $\varepsilon = 0$ and $m = 2$. *Classical Optimal Transport problem*. Convex problem, but may have several solutions γ, with or without finite entropy!
- $\varepsilon > 0$. Strictly convex cost $\implies$ unique solution γ_ε with finite entropy.
- It admits a dual problem

$$\mathrm{MOT}_\varepsilon = \sup \left\{ \sum_{i=1}^m \int_{X_i} \varphi_i(x_i) \mathrm{d}\mu_i - \varepsilon \log \left(\int_{\mathbf{X}} e^{\frac{\sum_{i=1}^m \varphi_i(x_i) - c(\mathbf{x})}{\varepsilon}} \mathrm{d} \otimes_{i=1}^m \mu_i \right) \mid \varphi_i \in \mathcal{C}_b(X_i) \right\}.$$

- The solution γ_ε is "almost" explicit

$$\gamma_\varepsilon = \exp \left(\frac{\oplus_{i=1}^m \varphi_i^\varepsilon - c}{\varepsilon} \right) \otimes_{i=1}^m \mu_i.$$

- Easy to solve numerically via Sinkhorn (take $m = 2$ for simplicity)

$$\varphi_1^{k+1} = -\varepsilon \log \left(\int_{X_2} e^{\frac{\varphi_2^k - c}{\varepsilon}} \mathrm{d}\mu_2 \right), \quad \varphi_2^{k+1} = -\varepsilon \log \left(\int_{X_1} e^{\frac{\varphi_1^{k+1} - c}{\varepsilon}} \mathrm{d}\mu_1 \right).$$

The asymptotics for the MOT_ε (with P. Pegon)

What are we interested in?and what is known...

Asymptotics for $\varepsilon \to 0$ of

- the cost MOT_ε
- the optimal entropic plan γ_ε and optimal Schrödinger potentials $(\varphi_i^\varepsilon)_{0 \leq i \leq m}$

What are we interested in?and what is known...

Asymptotics for $\varepsilon \to 0$ of

- the cost MOT_ε
- the optimal entropic plan γ_ε and optimal Schrödinger potentials $(\varphi_i^\varepsilon)_{0 \leq i \leq m}$

What are we interested in?and what is known...

Asymptotics for $\varepsilon \to 0$ of

- the cost MOT_ε
- the optimal entropic plan γ_ε and optimal Schrödinger potentials $(\varphi_i^\varepsilon)_{0 \leq i \leq m}$

First remark: depends heavily on $c = c(\boldsymbol{x})$ and the marginals μ_i's.

What are we interested in?and what is known...

Asymptotics for $\varepsilon \to 0$ of

- the cost MOT_ε
- the optimal entropic plan γ_ε and optimal Schrödinger potentials $(\varphi_i^\varepsilon)_{0 \leq i \leq m}$

First remark: depends heavily on $c = c(\boldsymbol{x})$ and the marginals μ_i's.
What is known?

What are we interested in?and what is known...

Asymptotics for $\varepsilon \to 0$ of

- the cost MOT_ε
- the optimal entropic plan γ_ε and optimal Schrödinger potentials $(\varphi_i^\varepsilon)_{0 \leq i \leq m}$

First remark: depends heavily on $c = c(\boldsymbol{x})$ and the marginals μ_i's.
What is known? Mostly for the $m = 2$ case.

What are we interested in?and what is known...

Asymptotics for $\varepsilon \to 0$ of

- the cost MOT_ε
- the optimal entropic plan γ_ε and optimal Schrödinger potentials $(\varphi_i^\varepsilon)_{0 \leq i \leq m}$

First remark: depends heavily on $c = c(\boldsymbol{x})$ and the marginals μ_i's.
What is known? Mostly for the $m = 2$ case.

- Convergence rate under strong regularity assumptions **(Pal 2019)**, second-order expansion for dynamical quadratic OT **(Conforti and Tamanini 2021)**

What are we interested in?and what is known...

Asymptotics for $\varepsilon \to 0$ of

- the cost MOT_ε
- the optimal entropic plan γ_ε and optimal Schrödinger potentials $(\varphi_i^\varepsilon)_{0 \leq i \leq m}$

First remark: depends heavily on $c = c(\boldsymbol{x})$ and the marginals μ_i's.
What is known? Mostly for the $m = 2$ case.

- Convergence rate under strong regularity assumptions **(Pal 2019)**, second-order expansion for dynamical quadratic OT **(Conforti and Tamanini 2021)**
- Convergence rate for $2-$marginal and a general class of $\mathcal{C}^2$ non-degenerate costs **(Guillaume Carlier, Pegon, and Tamanini 2022)**

What are we interested in?and what is known...

Asymptotics for $\varepsilon \to 0$ of

- the cost MOT_ε
- the optimal entropic plan γ_ε and optimal Schrödinger potentials $(\varphi_i^\varepsilon)_{0 \leq i \leq m}$

First remark: depends heavily on $c = c(\boldsymbol{x})$ and the marginals μ_i's.
What is known? Mostly for the $m = 2$ case.

- Convergence rate under strong regularity assumptions **(Pal 2019)**, second-order expansion for dynamical quadratic OT **(Conforti and Tamanini 2021)**
- Convergence rate for $2-$marginal and a general class of $\mathcal{C}^2$ non-degenerate costs **(Guillaume Carlier, Pegon, and Tamanini 2022)**
- Upper bound for the multi-marginal **(Eckstein and Nutz 2022)** *with a condition on the optimal transport plans* in terms of quantization dimension ;

The upper bound

Assumptions

- μ_i are compactly supported measures in $L^\infty(X_i)$ where X_i are $\mathcal{C}^2$ submanifolds of dimension d_i ;
- $c \in \mathcal{C}^{1,1}_{\mathrm{loc}}(\mathrm{X})$ or more generally locally semi-concave (also, weaker upper bound $c \in \mathcal{C}^{0,1}(\mathrm{X})$);

Goal: get an upper bound of the form

$$\mathrm{MOT}_\varepsilon - \mathrm{MOT}_0 \leq \frac{1}{2}\Big(\sum_{1\leq i\leq m} d_i - \max_i d_i\Big)\varepsilon \log(1/\varepsilon) + O(\varepsilon).$$

The upper bound

Assumptions

- μ_i are compactly supported measures in $L^\infty(X_i)$ where X_i are $\mathcal{C}^2$ submanifolds of dimension d_i ;
- $c \in \mathcal{C}^{1,1}_{\mathrm{loc}}(\mathrm{X})$ or more generally locally semi-concave (also, weaker upper bound $c \in \mathcal{C}^{0,1}(\mathrm{X})$);

Goal: get an upper bound of the form

$$\mathrm{MOT}_\varepsilon - \mathrm{MOT}_0 \leq \frac{1}{2}\Big(\sum_{1\leq i\leq m} d_i - \max_i d_i\Big)\varepsilon \log(1/\varepsilon) + O(\varepsilon).$$

Strategy. Straightforward (almost) generalization of the upper bound in **(Guillaume Carlier, Pegon, and Tamanini 2022)** on $\mathcal{C}^2$ submanifolds:

- Build a suitable competitor for the entropic (primal) problem

$$\mathrm{MOT}_\varepsilon = \inf_{\gamma\in\Pi(\mu_1,\dots,\mu_m)} \left\{ \int_{\mathrm{X}} c(x_1,\dots,x_m)\,\mathrm{d}\gamma(x_1,\dots,x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\}.$$

  using an optimizer for (MOT_0) and a block-approximation of **(Guillaume Carlier, Duval, Peyré, and Schmitzer 2017)**.

The upper bound

Assumptions

- μ_i are compactly supported measures in $L^\infty(X_i)$ where X_i are $\mathcal{C}^2$ submanifolds of dimension d_i ;
- $c \in \mathcal{C}^{1,1}_{\mathrm{loc}}(\mathsf{X})$ or more generally locally semi-concave (also, weaker upper bound $c \in \mathcal{C}^{0,1}(\mathsf{X})$);

Goal: get an upper bound of the form

$$\mathsf{MOT}_\varepsilon - \mathsf{MOT}_0 \leq \frac{1}{2}\Big(\sum_{1\leq i\leq m} d_i - \max_i d_i\Big)\varepsilon \log(1/\varepsilon) + O(\varepsilon).$$

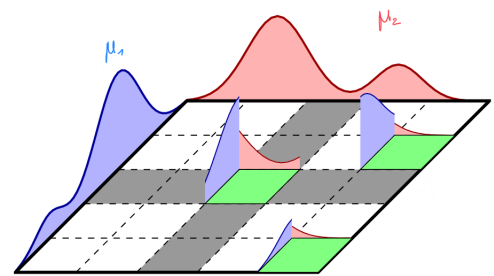
Strategy. Straightforward (almost) generalization of the upper bound in **(Guillaume Carlier, Pegon, and Tamanini 2022)** on $\mathcal{C}^2$ submanifolds:

- Build a suitable competitor for the entropic (primal) problem

$$\mathsf{MOT}_\varepsilon = \inf_{\gamma\in\Pi(\mu_1,\dots,\mu_m)} \left\{ \int_{\mathsf{X}} c(x_1,\dots,x_m)\,\mathrm{d}\gamma(x_1,\dots,x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\}.$$

  using an optimizer for (MOT_0) and a block-approximation of **(Guillaume Carlier, Duval, Peyré, and Schmitzer 2017)**.
- Show and use some integral variant of Alexandrov theorem on convex functions.

Upper bound: some details for $m = 2$, marginals μ_1, μ_2



For blocks $\bigsqcup_n A_n = \mathbb{R}^N$ of diameter $\leq \delta$, take as competitor

$$\gamma^\delta := \sum_{i,j \in \mathbb{N}} \gamma_0(A_i \times A_j) \frac{\mu_1 \llcorner A_i}{\mu_1(A_i)} \otimes \frac{\mu_2 \llcorner A_j}{\mu_2(A_j)}.$$

- Plug this competitor into the primal problem, write $E = c - \varphi \oplus \psi$ the duality gap, then:

$$\mathsf{MOT}_\varepsilon \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} c \mathrm{d}\gamma^\delta + \varepsilon \mathrm{Ent}(\gamma^\delta | \mu_1 \otimes \mu_2) = \mathsf{MOT}_0 + \int_{\mathbb{R}^d \times \mathbb{R}^d} E \mathrm{d}(\gamma^\delta - \gamma^0) + \varepsilon \mathrm{Ent}(\gamma^\delta | \mu_1 \otimes \mu_2)$$

- Plug this competitor into the primal problem, write $E = c - \varphi \oplus \psi$ the duality gap, then:

$$\mathrm{MOT}_\varepsilon \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} c \mathrm{d}\gamma^\delta + \varepsilon \mathrm{Ent}(\gamma^\delta | \mu_1 \otimes \mu_2) = \mathrm{MOT}_0 + \int_{\mathbb{R}^d \times \mathbb{R}^d} E \mathrm{d}(\gamma^\delta - \gamma^0) + \varepsilon \mathrm{Ent}(\gamma^\delta | \mu_1 \otimes \mu_2)$$

- Bound the entropy term, for well-chosen blocks:

$$\begin{aligned} \mathrm{Ent}(\gamma^\delta \mid \mu_1 \otimes \mu_2) &= \sum_{i,j \in \mathbb{N}} \gamma_0(A_i \times A_j) \log \left(\frac{\gamma_0(A_i \times A_j)}{\mu_1(A_i)\mu_2(A_j)} \right) \\ &\leq \sum_{j \in \mathbb{N}} \mu_2(A_j) \log(1/\mu_2(A_j)) = d_2 \log(1/\delta) + O(1). \end{aligned}$$

- Plug this competitor into the primal problem, write $E = c - \varphi \oplus \psi$ the duality gap, then:

$$\mathrm{MOT}_\varepsilon \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} c \mathrm{d}\gamma^\delta + \varepsilon \mathrm{Ent}(\gamma^\delta | \mu_1 \otimes \mu_2) = \mathrm{MOT}_0 + \int_{\mathbb{R}^d \times \mathbb{R}^d} E \mathrm{d}(\gamma^\delta - \gamma^0) + \varepsilon \mathrm{Ent}(\gamma^\delta | \mu_1 \otimes \mu_2)$$

- Bound the entropy term, for well-chosen blocks:

$$\begin{aligned} \mathrm{Ent}(\gamma^\delta \mid \mu_1 \otimes \mu_2) &= \sum_{i,j \in \mathbb{N}} \gamma_0(A_i \times A_j) \log \left(\frac{\gamma_0(A_i \times A_j)}{\mu_1(A_i)\mu_2(A_j)} \right) \\ &\leq \sum_{j \in \mathbb{N}} \mu_2(A_j) \log(1/\mu_2(A_j)) = d_2 \log(1/\delta) + O(1). \end{aligned}$$

- Show that $\int_{\mathbb{R}^d \times \mathbb{R}^d} E \mathrm{d}(\gamma^\delta - \gamma^0) = O(\delta^2)$ then take $\varepsilon = \delta^2$ (integral Alexandrov-type estimate):

$$\mathrm{MOT}_\varepsilon \leq \mathrm{MOT}_0 + O(\delta^2) + d_2 \varepsilon \log(1/\delta) + O(\varepsilon) = \mathrm{MOT}_0 + \frac{d^+}{2} \varepsilon \log(1/\varepsilon) + O(\varepsilon).$$

Reminder on multi-marginal optimal transport

Consider the unregularized problem

$$\mathrm{MOT}_0 = \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \left\{ \int_{\mathrm{X}} c(x_1, \ldots, x_m) \, \mathrm{d}\gamma(x_1, \ldots, x_m) \right\}.$$

Good news: we can estimate the dimension of the support of the optimal plans.

Reminder on multi-marginal optimal transport

Consider the unregularized problem

$$\mathrm{MOT}_0 = \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \left\{ \int_{\mathrm{X}} c(x_1, \ldots, x_m) \, \mathrm{d}\gamma(x_1, \ldots, x_m) \right\}.$$

Good news: we can estimate the dimension of the support of the optimal plans. • for each $p \in P$ we denote by g_p the bilinear form on X as

$$g_p = \sum_{i \in p_-, j \in p_+} D^2_{x_i, x_j} c + \sum_{i \in p_+, j \in p_-} D^2_{x_i, x_j} c.$$

Reminder on multi-marginal optimal transport

Consider the unregularized problem

$$\mathrm{MOT}_0 = \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \left\{ \int_{\mathrm{X}} c(x_1, \ldots, x_m) \, \mathrm{d}\gamma(x_1, \ldots, x_m) \right\}.$$

Good news: we can estimate the dimension of the support of the optimal plans. • for each $p \in P$ we denote by g_p the bilinear form on X as

$$g_p = \sum_{i \in p_-, j \in p_+} D^2_{x_i, x_j} c + \sum_{i \in p_+, j \in p_-} D^2_{x_i, x_j} c.$$

• Define $G := \{\sum_{p \in P} t_p g_p \mid (t_p)_{p \in P} \in \Delta_P\}$ to be the convex hull generated by the g_p.

Reminder on multi-marginal optimal transport

Consider the unregularized problem

$$\mathrm{MOT}_0 = \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \left\{ \int_{\mathrm{X}} c(x_1, \ldots, x_m) \, \mathrm{d}\gamma(x_1, \ldots, x_m) \right\}.$$

Good news: we can estimate the dimension of the support of the optimal plans. • for each $p \in P$ we denote by g_p the bilinear form on X as

$$g_p = \sum_{i \in p_-, j \in p_+} D^2_{x_i, x_j} c + \sum_{i \in p_+, j \in p_-} D^2_{x_i, x_j} c.$$

• Define $G := \{\sum_{p \in P} t_p g_p \mid (t_p)_{p \in P} \in \Delta_P\}$ to be the convex hull generated by the g_p.

Theorem (Upper bound on the dimension of the support of the optimal plan (Pass 2011))

Let γ_0 a solution to MOT_0 and suppose that at some point $\mathrm{x} \in \mathrm{X}$, the signature of some $g \in G$ is $(d^+(g), d^-(g), d^0(g))$. Then, there exists a neighbourhood N_{x} of x such that $N_{\mathrm{x}} \bigcap \mathrm{supp}(\gamma_0)$ is contained in a Lipschitz submanifold with dimension no greater than $\sum_i d_i - d^+(g)$.

Lower bound

- μ_i be compactly supported measures over X_i with L^∞ densities;
- $c \in \mathcal{C}^2(\mathrm{X})$;
- for every $\mathrm{x} \in \mathrm{X}$, $\quad d^+(g_{\mathrm{x}}) \geq d^\star$;

Goal: get a lower bound of the form

$$\mathrm{MOT}_\varepsilon - \mathrm{MOT}_0 \geq \frac{d^\star}{2}\varepsilon \log(1/\varepsilon) - L\varepsilon.$$

Lower bound

- μ_i be compactly supported measures over X_i with L^∞ densities;
- $c \in \mathcal{C}^2(\mathrm{X})$;
- for every $\mathrm{x} \in \mathrm{X}$, $\quad d^+(g_{\mathrm{x}}) \geq d^\star$;

Goal: get a lower bound of the form

$$\mathrm{MOT}_\varepsilon - \mathrm{MOT}_0 \geq \frac{d^\star}{2}\varepsilon \log(1/\varepsilon) - L\varepsilon.$$

Strategy

- Use the dual regularized problem (in log form):
- Take Kantorovich potentials (solution to un-regularized dual) as competitors and show that the duality gap $E \doteq c - \oplus_{i=1}^m \varphi_i$ grows enough near $\Sigma = \{E = 0\}$.

Lower bound: some details

Let $p = \{p_-, p_+\} \in P$ we identify $\boldsymbol{x} \in \boldsymbol{X}$ with (x_-, x_+) and write $\varphi_\pm(y) = \sum_{i \in p_\pm} \varphi_i(y_i)$.

Lower bound: some details

Let $p = \{p_-, p_+\} \in P$ we identify $\boldsymbol{x} \in \boldsymbol{X}$ with (x_-, x_+) and write $\varphi_\pm(y) = \sum_{i \in p_\pm} \varphi_i(y_i)$.

- If (φ_i) are c-conjugate, for $\boldsymbol{x}, \boldsymbol{x}' \in \mathrm{X}$, we have:

$$\begin{aligned} E(\boldsymbol{x}') &= c(x'_-, x'_+) - \varphi_-(x'_-) - \varphi_+(x'_+) \\ &\geq c(x'_-, x'_+) - (c(x'_-, x_+) - \varphi_+(x_+)) - (c(x_-, x'_+) - \varphi_-(x_-)) \\ &= c(x'_-, x'_+) - c(x'_-, x_+) - c(x_-, x'_+) + c(x_-, x_+) - E(\boldsymbol{x}). \end{aligned}$$

Lower bound: some details

Let $p = \{p_-, p_+\} \in P$ we identify $\boldsymbol{x} \in \boldsymbol{X}$ with (x_-, x_+) and write $\varphi_\pm(y) = \sum_{i \in p_\pm} \varphi_i(y_i)$.

- If (φ_i) are c-conjugate, for $\boldsymbol{x}, \boldsymbol{x}' \in \mathrm{X}$, we have:

$$\begin{aligned} E(\boldsymbol{x}') &= c(x'_-, x'_+) - \varphi_-(x'_-) - \varphi_+(x'_+) \\ &\geq c(x'_-, x'_+) - (c(x'_-, x_+) - \varphi_+(x_+)) - (c(x_-, x'_+) - \varphi_-(x_-)) \\ &= c(x'_-, x'_+) - c(x'_-, x_+) - c(x_-, x'_+) + c(x_-, x_+) - E(\boldsymbol{x}). \end{aligned}$$

- By Taylor's integral formula

$$E(\boldsymbol{x}') + E(\boldsymbol{x}) \geq \int_0^1 \int_0^1 D^2_{p_- p_+} c(\boldsymbol{x}_{s,t})(x'_- - x_-, x'_+ - x_+) = \frac{1}{2} g_p(\bar{\boldsymbol{x}})(\boldsymbol{x}' - \boldsymbol{x}) + O_{\bar{\boldsymbol{x}}}(\|\boldsymbol{x}' - \boldsymbol{x}\|^2)$$

Lower bound: some details

Let $p = \{p_-, p_+\} \in P$ we identify $\boldsymbol{x} \in \boldsymbol{X}$ with (x_-, x_+) and write $\varphi_\pm(y) = \sum_{i \in p_\pm} \varphi_i(y_i)$.

- If (φ_i) are c-conjugate, for $\boldsymbol{x}, \boldsymbol{x}' \in \mathrm{X}$, we have:

$$\begin{aligned} E(\boldsymbol{x}') &= c(x'_-, x'_+) - \varphi_-(x'_-) - \varphi_+(x'_+) \\ &\geq c(x'_-, x'_+) - (c(x'_-, x_+) - \varphi_+(x_+)) - (c(x_-, x'_+) - \varphi_-(x_-)) \\ &= c(x'_-, x'_+) - c(x'_-, x_+) - c(x_-, x'_+) + c(x_-, x_+) - E(\boldsymbol{x}). \end{aligned}$$

- By Taylor's integral formula

$$E(\boldsymbol{x}') + E(\boldsymbol{x}) \geq \int_0^1 \int_0^1 D^2_{p_- p_+} c(\boldsymbol{x}_{s,t})(x'_- - x_-, x'_+ - x_+) = \frac{1}{2} g_p(\bar{\boldsymbol{x}})(\boldsymbol{x}' - \boldsymbol{x}) + O_{\bar{\boldsymbol{x}}}(\|\boldsymbol{x}' - \boldsymbol{x}\|^2)$$

and taking a convex combination $g = \sum t_p g_p$, for diagonalizing coordinates (u^+, u^-, u^0)

$$E(\boldsymbol{x}') + E(\boldsymbol{x}) \geq |u^+(\boldsymbol{x}') - u^+(\boldsymbol{x})|^2 - |u^-(\boldsymbol{x}') - u^-(\boldsymbol{x})|^2 + O(|\boldsymbol{x}' - \boldsymbol{x}|^2)$$

$\implies$ quadratic detachment of the duality gap E in $d^+(g) \geq d^\star$ dimensions.

Lower bound: some details

Let $p = \{p_-, p_+\} \in P$ we identify $\boldsymbol{x} \in \boldsymbol{X}$ with (x_-, x_+) and write $\varphi_\pm(y) = \sum_{i \in p_\pm} \varphi_i(y_i)$.

- If (φ_i) are c-conjugate, for $\boldsymbol{x}, \boldsymbol{x}' \in \mathrm{X}$, we have:

$$\begin{aligned} E(\boldsymbol{x}') &= c(x'_-, x'_+) - \varphi_-(x'_-) - \varphi_+(x'_+) \\ &\geq c(x'_-, x'_+) - (c(x'_-, x_+) - \varphi_+(x_+)) - (c(x_-, x'_+) - \varphi_-(x_-)) \\ &= c(x'_-, x'_+) - c(x'_-, x_+) - c(x_-, x'_+) + c(x_-, x_+) - E(\boldsymbol{x}). \end{aligned}$$

- By Taylor's integral formula

$$E(\boldsymbol{x}') + E(\boldsymbol{x}) \geq \int_0^1 \int_0^1 D^2_{p_- p_+} c(\boldsymbol{x}_{s,t})(x'_- - x_-, x'_+ - x_+) = \frac{1}{2} g_p(\bar{\boldsymbol{x}})(\boldsymbol{x}' - \boldsymbol{x}) + O_{\bar{\boldsymbol{x}}}(\|\boldsymbol{x}' - \boldsymbol{x}\|^2)$$

  and taking a convex combination $g = \sum t_p g_p$, for diagonalizing coordinates (u^+, u^-, u^0)

$$E(\boldsymbol{x}') + E(\boldsymbol{x}) \geq |u^+(\boldsymbol{x}') - u^+(\boldsymbol{x})|^2 - |u^-(\boldsymbol{x}') - u^-(\boldsymbol{x})|^2 + O(|\boldsymbol{x}' - \boldsymbol{x}|^2)$$

  $\implies$ quadratic detachment of the duality gap E in $d^+(g) \geq d^\star$ dimensions.

- Taking $(\varphi_i)_{1 \leq i \leq m}$ as competitor in the dual of the entropic MOT:

$$\mathrm{MOT}_\varepsilon \geq \mathrm{MOT}_0 - \varepsilon \log \left(\int_{\Pi_{1 \leq i \leq m} X_i} e^{-\frac{E}{\varepsilon}} \, \mathrm{d} \otimes_{1 \leq i \leq m} \mu_i \right) \geq \mathrm{MOT}_0 + \frac{d^\star}{2} \varepsilon \log(1/\varepsilon) - O(\varepsilon).$$

Consequences and some examples

- $m = 2$ and **non-degenerate** cost function, then we retrieve the bounds in **(Guillaume Carlier, Pegon, and Tamanini 2022)**;

Consequences and some examples

- $m = 2$ and **non-degenerate** cost function, then we retrieve the bounds in **(Guillaume Carlier, Pegon, and Tamanini 2022)**;
- $m = 2$ and **degenerate** cost function such that $D^2_{x,y}c$ has rank r then the lower bound is such that $d^\star = r$;

Consequences and some examples

- $m = 2$ and **non-degenerate** cost function, then we retrieve the bounds in **(Guillaume Carlier, Pegon, and Tamanini 2022)**;
- $m = 2$ and **degenerate** cost function such that $D^2_{x,y}c$ has rank r then the lower bound is such that $d^\star = r$;
- $m = 2$, $d_2 < d_1$ (aka the **unequal dimensional case**) and $D^2_{x,y}c$ has full rank d_2 then

$$\mathrm{MOT}_\varepsilon = \mathrm{MOT}_0 + \frac{d_2}{2}\varepsilon \log(1/\varepsilon) + O(\varepsilon)$$

Consequences and some examples

- $m = 2$ and **non-degenerate** cost function, then we retrieve the bounds in **(Guillaume Carlier, Pegon, and Tamanini 2022)**;
- $m = 2$ and **degenerate** cost function such that $D^2_{x,y}c$ has rank r then the lower bound is such that $d^\star = r$;
- $m = 2$, $d_2 < d_1$ (aka the **unequal dimensional case**) and $D^2_{x,y}c$ has full rank d_2 then

$$\mathrm{MOT}_\varepsilon = \mathrm{MOT}_0 + \frac{d_2}{2}\varepsilon \log(1/\varepsilon) + O(\varepsilon)$$

- Consider $d_i = d$ for all i and the cost $c = h(\sum_{i=1}^m x_i)$ with $D^2 h < 0$ then $d^* = (m-1)d$ and

$$\mathrm{MOT}_\varepsilon = \mathrm{MOT}_0 + \frac{(m-1)d}{2}\varepsilon \log(1/\varepsilon) + O(\varepsilon).$$

This is the case of Gangbo-Święch cost, that is $\sum_{i<j} |x_i - x_j|^2$ which corresponds to the multi-marginal formulation of the Wasserstein barycenter problem.

Another way to characterise (and solve) MOT_ε: an ODE approach (with B. Pass)

What are we interested in and direction of our work

We are interested in solving the entropic multi-marginal optimal transport.

What are we interested in and direction of our work

We are interested in solving the entropic multi-marginal optimal transport.

Main steps of the work:

1. Introduce a suitable one parameter family of cost functions c_η, **interpolating** between the original multi-marginal problem and a simpler one whose complexity **scales linearly** in the number of marginals;

What are we interested in and direction of our work

We are interested in solving the entropic multi-marginal optimal transport.

Main steps of the work:

1. Introduce a suitable one parameter family of cost functions c_η, **interpolating** between the original multi-marginal problem and a simpler one whose complexity **scales linearly** in the number of marginals;
2. Differentiate the **optimality condition of the dual** $\mathrm{MOT}_\varepsilon := \sup_\varphi \Phi(\varphi, \eta)$ with respect to η (ε is now fixed);

What are we interested in and direction of our work

We are interested in solving the entropic multi-marginal optimal transport.

Main steps of the work:

1. Introduce a suitable one parameter family of cost functions c_η, **interpolating** between the original multi-marginal problem and a simpler one whose complexity **scales linearly** in the number of marginals;
2. Differentiate the **optimality condition of the dual** $\mathrm{MOT}_\varepsilon := \sup_\varphi \Phi(\varphi, \eta)$ with respect to η (ε is now fixed);
3. The solution of the original multi-marginal problem can be now recovered by solving an **ordinary differential equation** (ODE) whose initial condition is the solution to the simpler problem;

$$\boxed{\begin{cases} \dfrac{\mathrm{d}\varphi}{\mathrm{d}\eta}(\eta) = -[D^2_{\varphi,\varphi}\Phi(\varphi(\eta),\eta)]^{-1}\dfrac{\partial}{\partial\eta}\nabla_\varphi\Phi(\varphi(\eta),\eta), \\ \varphi(0) = \varphi_w, \end{cases}}$$

What are we interested in and direction of our work

We are interested in solving the entropic multi-marginal optimal transport.

Main steps of the work:

1. Introduce a suitable one parameter family of cost functions c_η, **interpolating** between the original multi-marginal problem and a simpler one whose complexity **scales linearly** in the number of marginals;
2. Differentiate the **optimality condition of the dual** $\mathrm{MOT}_\varepsilon := \sup_\varphi \Phi(\varphi, \eta)$ with respect to η (ε is now fixed);
3. The solution of the original multi-marginal problem can be now recovered by solving an **ordinary differential equation** (ODE) whose initial condition is the solution to the simpler problem;

$$\boxed{\begin{cases} \dfrac{\mathrm{d}\varphi}{\mathrm{d}\eta}(\eta) = -[D^2_{\varphi,\varphi}\Phi(\varphi(\eta),\eta)]^{-1}\dfrac{\partial}{\partial\eta}\nabla_\varphi\Phi(\varphi(\eta),\eta), \\ \varphi(0) = \varphi_w, \end{cases}}$$

Remark: This method is actually inspired by the one introduced in **(G. Carlier, A. Galichon, and F. Santambrogio 2009/10)** to compute the Monge solution of the two marginal problem, starting from the Knothe-Rosenblatt rearrangement.

How to derive the differential equation

Some assumptions to make it simple:

1. **(Equal marginals and discrete set)** All the marginals are equal $\mu_i = \rho = \sum_{x \in X} \rho_x \delta_x$, where X is a finite subset.
2. **(Pair-wise cost)** $c_\eta(x_1, \ldots, x_m) := \eta \sum_{i=2}^{m} \sum_{j=i+1}^{m} w(x_i, x_j) + \sum_{i=2}^{m} \boldsymbol{w(x_1, x_i)}$.
3. **(Symmetric cost)** The two body cost w is symmetric $w(x, y) = w(x, y)$.
4. **(Finite cost)** The two body cost function $w : X \times X \to \mathbb{R}$ is everywhere real-valued.

How to derive the differential equation

Some assumptions to make it simple:

1. **(Equal marginals and discrete set)** All the marginals are equal $\mu_i = \rho = \sum_{x \in X} \rho_x \delta_x$, where X is a finite subset.
2. **(Pair-wise cost)** $c_\eta(x_1, \ldots, x_m) := \eta \sum_{i=2}^m \sum_{j=i+1}^m w(x_i, x_j) + \boldsymbol{\sum_{i=2}^m w(x_1, x_i)}$.
3. **(Symmetric cost)** The two body cost w is symmetric $w(x, y) = w(x, y)$.
4. **(Finite cost)** The two body cost function $w : X \times X \to \mathbb{R}$ is everywhere real-valued.

Step 1: Consider the dual problem (it is convex!);

$$\inf_{\varphi} \{\Phi(\varphi, \eta)\}, \tag{5}$$

where

$$\Phi(\varphi, \eta) := -(m-1) \int_X \varphi \mathrm{d}\rho + \varepsilon \int_X \underbrace{\log\left(\int_{X^{m-1}} \exp\left(\frac{\sum_{i=2}^m \varphi - c_\eta}{\varepsilon}\right) \mathrm{d} \otimes^{m-1} \rho\right)}_{\text{Log-Sum-Exp}} \mathrm{d}\rho.$$

Step 2: Thanks to convexity we have that the minimizers are characterized by $\nabla_\varphi \Phi(\varphi, \eta) = 0$. Then, by differentiate w.r.t. η we obtain

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\eta}(\eta) = -[D^2_{\varphi,\varphi}\Phi(\varphi(\eta), \eta)]^{-1} \frac{\partial}{\partial \eta} \nabla_\varphi \Phi(\varphi(\eta), \eta).$$

Step 2: Thanks to convexity we have that the minimizers are characterized by $\nabla_\varphi \Phi(\varphi, \eta) = 0$. Then, by differentiate w.r.t. η we obtain

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\eta}(\eta) = -[D^2_{\varphi,\varphi}\Phi(\varphi(\eta), \eta)]^{-1} \frac{\partial}{\partial \eta} \nabla_\varphi \Phi(\varphi(\eta), \eta).$$

Step 3: The following well-posedness theorem then holds.

Theorem

Let $\varphi(\eta)$ be the solution to the dual problem above for all $\eta \in [0, 1]$. Then $\eta \mapsto \varphi(\eta)$ is $\mathcal{C}^1$ and is the unique solution to the Cauchy problem with $\varphi(0) = \varphi_w$.

Sketch of the proof:

Step 2: Thanks to convexity we have that the minimizers are characterized by $\nabla_\varphi \Phi(\varphi, \eta) = 0$. Then, by differentiate w.r.t. η we obtain

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\eta}(\eta) = -[D^2_{\varphi,\varphi}\Phi(\varphi(\eta), \eta)]^{-1} \frac{\partial}{\partial \eta} \nabla_\varphi \Phi(\varphi(\eta), \eta).$$

Step 3: The following well-posedness theorem then holds.

Theorem

Let $\varphi(\eta)$ be the solution to the dual problem above for all $\eta \in [0, 1]$. Then $\eta \mapsto \varphi(\eta)$ is $\mathcal{C}^1$ and is the unique solution to the Cauchy problem with $\varphi(0) = \varphi_w$.

Sketch of the proof:

- The pure second derivatives with respect to φ as well as the mixed second derivatives with respect to φ and η exist and are Lipschitz;

Step 2: Thanks to convexity we have that the minimizers are characterized by $\nabla_\varphi \Phi(\varphi, \eta) = 0$. Then, by differentiate w.r.t. η we obtain

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\eta}(\eta) = -[D^2_{\varphi,\varphi}\Phi(\varphi(\eta), \eta)]^{-1} \frac{\partial}{\partial \eta} \nabla_\varphi \Phi(\varphi(\eta), \eta).$$

Step 3: The following well-posedness theorem then holds.

Theorem

Let $\varphi(\eta)$ be the solution to the dual problem above for all $\eta \in [0, 1]$. Then $\eta \mapsto \varphi(\eta)$ is $\mathcal{C}^1$ and is the unique solution to the Cauchy problem with $\varphi(0) = \varphi_w$.

Sketch of the proof:

- The pure second derivatives with respect to φ as well as the mixed second derivatives with respect to φ and η exist and are Lipschitz;
- The Hessian with respect to φ is invertible: since the cost is bounded then the potentials are bounded too (**(carlier2021linear)**). So one can restrict the study of the well-posedness of the ODE on the set

$$U := \{\varphi \mid \varphi_{\mathbf{x_0}} = 0,\ ||\varphi||_\infty \leq C\}.$$

On this set the functional Φ is now **strongly convex**.

The algorithm to compute the ODE solution

- Algorithm to compute the φ via explicit Euler method takes the following form:

Require: $\varphi(0) = \varphi_w$

while $||\varphi^{(k+1)} - \varphi^{(k)}|| <$tol **do**

$\quad D^{(k)} := D^2_{\varphi,\varphi}\Phi(\varphi^{(k)}, kh)$

$\quad b^{(k)} := -\dfrac{\partial}{\partial\epsilon}\nabla_\varphi\Phi(\varphi^{(k)}, kh)$

$\quad$ Solve $D^{(k)}z = b^{(k)}$

$\quad \varphi^{(k+1)} = \varphi^{(k)} + hz$

end while

The algorithm to compute the ODE solution

- Algorithm to compute the φ via explicit Euler method takes the following form:

Require: $\varphi(0) = \varphi_w$

while $||\varphi^{(k+1)} - \varphi^{(k)}|| <$tol **do**

$\quad D^{(k)} := D^2_{\varphi,\varphi}\Phi(\varphi^{(k)}, kh)$

$\quad b^{(k)} := -\dfrac{\partial}{\partial\epsilon}\nabla_\varphi\Phi(\varphi^{(k)}, kh)$

$\quad$ Solve $D^{(k)}z = b^{(k)}$

$\quad \varphi^{(k+1)} = \varphi^{(k)} + hz$

end while

Remarks:

- The Euler scheme converges linearly and the uniform error between the discretized solution obtained via the scheme and the solution to the ODE is $O(h)$;
- Thanks to the regularity of the RHS of the ODE one can apply high order methods.

The algorithm to compute the ODE solution

- Algorithm to compute the φ via explicit Euler method takes the following form:

Require: $\varphi(0) = \varphi_w$

while $||\varphi^{(k+1)} - \varphi^{(k)}|| <$tol **do**

$\quad D^{(k)} := D^2_{\varphi,\varphi}\Phi(\varphi^{(k)}, kh)$

$\quad b^{(k)} := -\dfrac{\partial}{\partial\epsilon}\nabla_\varphi\Phi(\varphi^{(k)}, kh)$

$\quad$ Solve $D^{(k)}z = b^{(k)}$

$\quad \varphi^{(k+1)} = \varphi^{(k)} + hz$

end while

Remarks:

- The Euler scheme converges linearly and the uniform error between the discretized solution obtained via the scheme and the solution to the ODE is $O(h)$;
- Thanks to the regularity of the RHS of the ODE one can apply high order methods.
- At each step k we obtaine the solution of the entropic multi-marginal problem with cost c_{kh}!

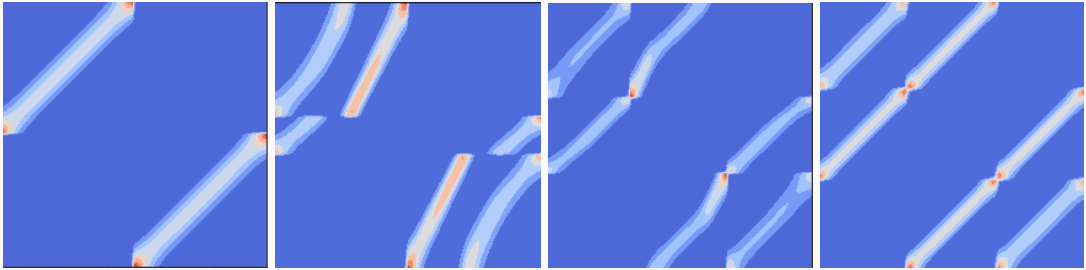
Comparison with Sinkhorn

Consider $\varepsilon = 0.006$, $m = 3$, the uniform measure on $[0, 1]$ uniformily discretized with 400 gridpoints, the pairwise interaction $w(x, y) = -\log(0.1 + |x - y|)$ and a reference solution φ_ε computed via a gradient descent algorithm. Then we have the following comparison between the ODE approach and Sinkhorn in terms of performances

	3rd RK	5th RK	8th RK	Sinkhorn
relative error	1.47×10^{-5}	7.8×10^{-6}	7.62×10^{-6}	5.46×10^{-6}
iterations	87	87	87	820
CPU time (sec)	72.39	158.9	385.1	102.8

Some numerical results

- Log cost and support of the coupling $\gamma_{1,2}^{\eta}$.



$\eta = 0$ $\eta = 0.25$ $\eta = 0.75$ $\eta = 1$

Generalized solutions to incompressible Euler Equations

- Brenier's relaxed formulation consists in finding a probability measure over absolutely continuous paths which minimizes the average kinetic energy.

Generalized solutions to incompressible Euler Equations

- Brenier's relaxed formulation consists in finding a probability measure over absolutely continuous paths which minimizes the average kinetic energy.
- The **incompressibility** at each time t, the distribution of position need be uniform.

Generalized solutions to incompressible Euler Equations

- Brenier's relaxed formulation consists in finding a probability measure over absolutely continuous paths which minimizes the average kinetic energy.
- The **incompressibility** at each time t, the distribution of position need be uniform.
- If we consider a uniform discretization of $[0, T]$ (where T is the final time) with m steps in time, we recover a multi-marginal formulation of the Brenier principle with the specific cost function

$$c(x_1, \dots, x_m) = \frac{m^2}{2T^2} \sum_{i=1}^{m-1} |x_{i+1} - x_i|^2 + \beta |F(x_1) - x_m|^2,$$

where $\beta > 0$ is a penalization parameter in order to enforce the initial-final constraint.

Generalized solutions to incompressible Euler Equations

- Brenier's relaxed formulation consists in finding a probability measure over absolutely continuous paths which minimizes the average kinetic energy.
- The **incompressibility** at each time t, the distribution of position need be uniform.
- If we consider a uniform discretization of $[0, T]$ (where T is the final time) with m steps in time, we recover a multi-marginal formulation of the Brenier principle with the specific cost function

$$c(x_1, \dots, x_m) = \frac{m^2}{2T^2} \sum_{i=1}^{m-1} |x_{i+1} - x_i|^2 + \beta |F(x_1) - x_m|^2,$$

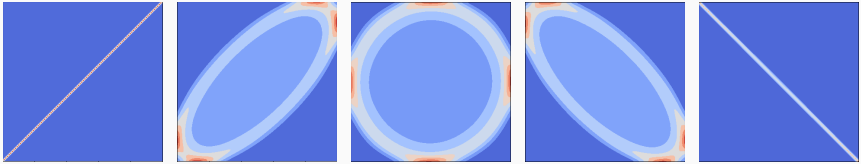
  where $\beta > 0$ is a penalization parameter in order to enforce the initial-final constraint.
- If we consider now the ODE setting, we have now to deal with a non symmetric case and so to solve a system, still well posed, of ODEs. In particular we consider the following c_η cost

$$\boxed{c_\eta(x_1, \dots, x_m) = \frac{m^2}{2T^2} |x_2 - x_1|^2 + \eta \left(\frac{m^2}{2T^2} \sum_{i=2}^{m-1} |x_{i+1} - x_i|^2 \right) + \beta |F(x_1) - x_m|^2.}$$

At $\eta = 1$ we plot the coupling $\gamma_{1,i}$ giving the probability of finding a generalized particle initially at x_1 to be at x_i at time i.

At $\eta = 1$ we plot the coupling $\gamma_{1,i}$ giving the probability of finding a generalized particle initially at x_1 to be at x_i at time i.

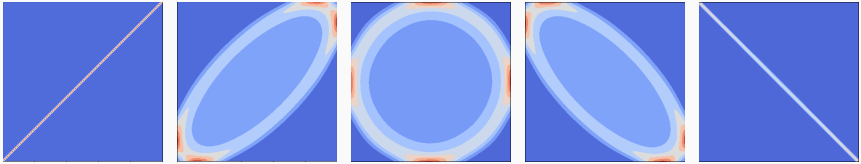
- $F(x) = 1 - x$



$t = 0$ $t = 1/4$ $t = 1/2$ $t = 3/4$ $t = 1$

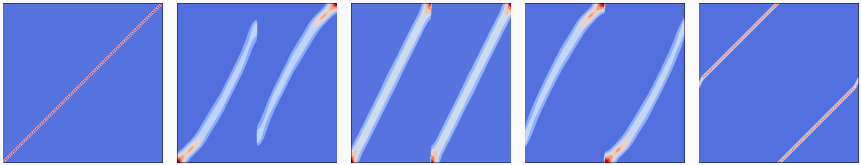
At $\eta = 1$ we plot the coupling $\gamma_{1,i}$ giving the probability of finding a generalized particle initially at x_1 to be at x_i at time i.

- $F(x) = 1 - x$



$t = 0$ $t = 1/4$ $t = 1/2$ $t = 3/4$ $t = 1$

- $F(x) = (x + 1/2) \bmod 1$



$t = 0$ $t = 1/4$ $t = 1/2$ $t = 3/4$ $t = 1$

Extension to general multi-marginal problems (joint work with B. Pass and J. Zoen-Git Hiew)

Consider the following "1st" generalization

$$\boxed{\mathrm{MOT}_\varepsilon := \inf_{\gamma \in \Pi(\mu_1, \dots, \mu_m)} \left\{ \int_X c(\boldsymbol{\eta}, x_1, \dots, x_m) \, \mathrm{d}\gamma(x_1, \dots, x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\},}$$

where the cost function is not anymore symmetric but such that $c(0, x_1, \dots, x_m)$ give a MOT easy to solve:

1. $c(0, x_1, \dots, x_m) = 0$;

Extension to general multi-marginal problems (joint work with B. Pass and J. Zoen-Git Hiew)

Consider the following "1st" generalization

$$\boxed{\mathrm{MOT}_\varepsilon := \inf_{\gamma \in \Pi(\mu_1, \dots, \mu_m)} \left\{ \int_X c(\boldsymbol{\eta}, x_1, \dots, x_m)\, \mathrm{d}\gamma(x_1, \dots, x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\},}$$

where the cost function is not anymore symmetric but such that $c(0, x_1, \dots, x_m)$ give a MOT easy to solve:

1. $c(0, x_1, \dots, x_m) = 0$;
2. $c(\eta, x_1, \dots, x_m)$ is the Euler cost;

Extension to general multi-marginal problems (joint work with B. Pass and J. Zoen-Git Hiew)

Consider the following "1st" generalization

$$\boxed{\mathrm{MOT}_\varepsilon := \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \left\{ \int_X c(\boldsymbol{\eta}, x_1, \ldots, x_m)\, \mathrm{d}\gamma(x_1, \ldots, x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\},}$$

where the cost function is not anymore symmetric but such that $c(0, x_1, \ldots, x_m)$ give a MOT easy to solve:

1. $c(0, x_1, \ldots, x_m) = 0$;
2. $c(\eta, x_1, \ldots, x_m)$ is the Euler cost;
3. $c(\eta, x_1, z, x_2) = (1-\eta)|x_1 - z|^2 + \eta|z - x_3|^2$, γ is a 3 marginals coupling with only two fixed marginals, μ_1 and μ_2. Then the $z-$marginal of γ gives the **Wasserstein geodesic** at time η.

Extension to general multi-marginal problems (joint work with B. Pass and J. Zoen-Git Hiew)

Consider the following "1st" generalization

$$\boxed{\mathrm{MOT}_\varepsilon := \inf_{\gamma \in \Pi(\mu_1,\dots,\mu_m)} \left\{ \int_X c(\boldsymbol{\eta}, x_1, \dots, x_m)\, \mathrm{d}\gamma(x_1, \dots, x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\},}$$

where the cost function is not anymore symmetric but such that $c(0, x_1, \dots, x_m)$ give a MOT easy to solve:

1. $c(0, x_1, \dots, x_m) = 0$;
2. $c(\eta, x_1, \dots, x_m)$ is the Euler cost;
3. $c(\eta, x_1, z, x_2) = (1-\eta)|x_1 - z|^2 + \eta|z - x_3|^2$, γ is a 3 marginals coupling with only two fixed marginals, μ_1 and μ_2. Then the $z-$marginal of γ gives the **Wasserstein geodesic** at time η.
4. $c(\eta, x_1, \dots, x_m, z) = \sum_{i=1}^m \lambda_i(\eta)|x_i - z|^2$ such that $\sum_{i=1}^m \lambda_i(\eta) = 1$ for every η and γ is an $m+1$ coupling with m fixed marginals. Then at for every η the $z-$marginal of γ is the **Wasserstein barycenter** with weights $\lambda_i(\eta)$.

Extension to general multi-marginal problems (joint work with B. Pass and J. Zoen-Git Hiew)

Consider the following "2nd" generalization

$$\boxed{\mathrm{MOT}_\varepsilon := \inf_{\gamma \in \boldsymbol{\Pi^Q(\mu_1,\ldots,\mu_m)}} \left\{ \int_{\mathrm{X}} c(\eta, x_1, \ldots, x_m)\, \mathrm{d}\gamma(x_1, \ldots, x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\},}$$

where $\boldsymbol{\Pi^Q(\mu_1, \ldots, \mu_m)}$ is the set of coupling having $\mu_1, \ldots, \mu_m$ as marginals and satisfying an additional constraint $\int q \mathrm{d}\gamma = 0$ for all $q \in Q$ where Q be a set of bounded continuous function on $\boldsymbol{X}$.

- **Classical case:** $\boldsymbol{Q} = \{0\}$;

Extension to general multi-marginal problems (joint work with B. Pass and J. Zoen-Git Hiew)

Consider the following "2nd" generalization

$$\boxed{\mathrm{MOT}_\varepsilon := \inf_{\gamma \in \boldsymbol{\Pi^Q(\mu_1,\ldots,\mu_m)}} \left\{ \int_{\mathrm{X}} c(\eta, x_1, \ldots, x_m)\, \mathrm{d}\gamma(x_1, \ldots, x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\},}$$

where $\boldsymbol{\Pi^Q(\mu_1, \ldots, \mu_m)}$ is the set of coupling having $\mu_1, \ldots, \mu_m$ as marginals and satisfying an additional constraint $\int q \mathrm{d}\gamma = 0$ for all $q \in Q$ where Q be a set of bounded continuous function on $\boldsymbol{X}$.

- **Classical case:** $Q = \{0\}$;
- **Generalized Euler solution:** force $\gamma_{1,m} = (Id, F)_\sharp \mathcal{L}$;

Extension to general multi-marginal problems (joint work with B. Pass and J. Zoen-Git Hiew)

Consider the following "2nd" generalization

$$\boxed{\mathrm{MOT}_\varepsilon := \inf_{\gamma \in \boldsymbol{\Pi^Q(\mu_1,\ldots,\mu_m)}} \left\{ \int_X c(\eta, x_1, \ldots, x_m)\, \mathrm{d}\gamma(x_1, \ldots, x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\},}$$

where $\boldsymbol{\Pi^Q(\mu_1, \ldots, \mu_m)}$ is the set of coupling having $\mu_1, \ldots, \mu_m$ as marginals and satisfying an additional constraint $\int q \mathrm{d}\gamma = 0$ for all $q \in Q$ where Q be a set of bounded continuous function on $\boldsymbol{X}$.

- **Classical case:** $Q = \{0\}$;
- **Generalized Euler solution:** force $\gamma_{1,m} = (Id, F)_\sharp \mathcal{L}$;
- **Martingale OT:** $\Pi^Q(\mu_1, \mu_2)$ with extra constraint

$$\int q(x_1)(x_2 - x_1) \mathrm{d}\gamma = 0, \quad \forall q \in \mathcal{C}_b(X_1).$$

Extension to general multi-marginal problems (joint work with B. Pass and J. Zoen-Git Hiew)

Consider the following "2nd" generalization

$$\boxed{\mathrm{MOT}_\varepsilon := \inf_{\gamma \in \boldsymbol{\Pi^Q(\mu_1,\ldots,\mu_m)}} \left\{ \int_X c(\eta, x_1, \ldots, x_m)\, \mathrm{d}\gamma(x_1, \ldots, x_m) + \varepsilon \mathrm{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \right\},}$$

where $\boldsymbol{\Pi^Q(\mu_1,\ldots,\mu_m)}$ is the set of coupling having $\mu_1, \ldots, \mu_m$ as marginals and satisfying an additional constraint $\int q \mathrm{d}\gamma = 0$ for all $q \in Q$ where Q be a set of bounded continuous function on $\boldsymbol{X}$.

- **Classical case:** $\boldsymbol{Q} = \{0\}$;
- **Generalized Euler solution:** force $\gamma_{1,m} = (Id, F)_\sharp \mathcal{L}$;
- **Martingale OT:** $\Pi^Q(\mu_1, \mu_2)$ with extra constraint

$$\int q(x_1)(x_2 - x_1) \mathrm{d}\gamma = 0, \quad \forall q \in \mathcal{C}_b(X_1).$$

- **Multi-period martingale OT:** e.g. 3−period $\Pi^Q(\mu_1, \mu_2, \mu_3)$ with extra constraint

$$\int [q(x_1)(x_2 - x_1) + h(x_1, x_2)(x_3 - x_2)] \mathrm{d}\gamma = 0, \quad \forall q \in \mathcal{C}_b(X_1), \forall h \in \mathcal{C}_b(X_1 \times X_2).$$

• Si vous voulez goûter des gourmandises italiennes : soutenance de HDR le 6 mars à 14h30 (à Orsay)! :)

• Si vous voulez goûter des gourmandises italiennes : soutenance de HDR le 6 mars à 14h30 (à Orsay)! :)

Thank You!!

• Si vous voulez goûter des gourmandises italiennes : soutenance de HDR le 6 mars à 14h30 (à Orsay)! :)

Thank You!!

There are some more slides.

Spectral risk measures and MOT (with H. Ennaji, Q. Mérigot and B. Pass)

Spectral risk measures

We will consider *spectral risk measures* to quantify the risk associated with μ. Given an integrable, non-negative and nondecreasing function $\alpha : [0,1] \to \mathbb{R}_+$ with $\int_0^1 \alpha(t)\mathrm{d}t = 1$, the α-risk, is defined as

$$R_\alpha(\mu) = \int_0^1 F_\mu^{-1}(m)\alpha(m)\mathrm{d}m.$$

Spectral risk measures

We will consider *spectral risk measures* to quantify the risk associated with μ. Given an integrable, non-negative and nondecreasing function $\alpha : [0,1] \to \mathbb{R}_+$ with $\int_0^1 \alpha(t)\mathrm{d}t = 1$, the α-risk, is defined as

$$R_\alpha(\mu) = \int_0^1 F_\mu^{-1}(m)\alpha(m)\mathrm{d}m.$$

- Lemma (variational representation): If the function α is non-decreasing, then

$$R_\alpha(\mu) = \max_{\gamma \in \Pi(\alpha_\sharp \mathrm{Leb}_{[0,1]}, \mu)} \int_{\mathbb{R}\times\mathbb{R}} xy \mathrm{d}\gamma(x,y),$$

the problem is indeed an optimal transport problem with 2 marginals.

Spectral risk measures

We will consider *spectral risk measures* to quantify the risk associated with μ. Given an integrable, non-negative and nondecreasing function $\alpha : [0,1] \to \mathbb{R}_+$ with $\int_0^1 \alpha(t)\mathrm{d}t = 1$, the α-risk, is defined as

$$R_\alpha(\mu) = \int_0^1 F_\mu^{-1}(m)\alpha(m)\mathrm{d}m.$$

- Lemma (variational representation): If the function α is non-decreasing, then

$$R_\alpha(\mu) = \max_{\gamma \in \Pi(\alpha_\sharp \mathrm{Leb}_{[0,1]}, \mu)} \int_{\mathbb{R}\times\mathbb{R}} xy\mathrm{d}\gamma(x,y),$$

the problem is indeed an optimal transport problem with 2 marginals.

- Lemma: $\mu \mapsto R_\alpha(\mu)$ is concave on $\mathcal{P}(\mathbb{R})$.

The m parameters case

Let denote by $b : X_1 \times \cdots \times X_m \to \mathbb{R}$ the output function which describes the level of the risk depending on the parameters of the systems and $\mu_1, \cdots, \mu_N$ the probability measures, associated to each parameter. Then, the problem of determining the worst case is then to maximize the $\alpha-$risk of $b_\sharp \eta$ over all η

$$\max_{\eta \in \Pi(\mu_1, \mu_2, \ldots, \mu_m)} R_\alpha(b_\# \eta).$$

The m parameters case

Let denote by $b : X_1 \times \cdots \times X_m \to \mathbb{R}$ the output function which describes the level of the risk depending on the parameters of the systems and $\mu_1, \cdots, \mu_N$ the probability measures, associated to each parameter. Then, the problem of determining the worst case is then to maximize the $\alpha-$risk of $b_\sharp \eta$ over all η

$$\max_{\eta \in \Pi(\mu_1, \mu_2, \ldots, \mu_m)} R_\alpha(b_\# \eta).$$

- $b_\# \gamma$ represents the distribution of outputs.

The m parameters case

Let denote by $b : X_1 \times \cdots \times X_m \to \mathbb{R}$ the output function which describes the level of the risk depending on the parameters of the systems and $\mu_1, \cdots, \mu_N$ the probability measures, associated to each parameter. Then, the problem of determining the worst case is then to maximize the $\alpha-$risk of $b_\sharp \eta$ over all η

$$\max_{\eta \in \Pi(\mu_1, \mu_2, \ldots, \mu_m)} R_\alpha(b_\# \eta).$$

- $b_\# \gamma$ represents the distribution of outputs.
- Proposition: under some mild assumption on b there exists a solution to the above problem. Moreover, $\eta \mapsto R_\alpha(b_\# \eta)$ is concave on $\mathcal{P}(\mathbb{R}^N)$.

The m parameters case

Let denote by $b : X_1 \times \cdots \times X_m \to \mathbb{R}$ the output function which describes the level of the risk depending on the parameters of the systems and $\mu_1, \cdots, \mu_N$ the probability measures, associated to each parameter. Then, the problem of determining the worst case is then to maximize the $\alpha-$risk of $b_\sharp \eta$ over all η

$$\max_{\eta \in \Pi(\mu_1, \mu_2, \ldots, \mu_m)} R_\alpha(b_\# \eta).$$

- $b_\# \gamma$ represents the distribution of outputs.
- Proposition: under some mild assumption on b there exists a solution to the above problem. Moreover, $\eta \mapsto R_\alpha(b_\# \eta)$ is concave on $\mathcal{P}(\mathbb{R}^N)$.
- A double optimization problem:

$$R_\alpha(b_\# \eta) = \max_{\sigma \in \Pi(\alpha_\sharp \mathrm{Leb}_{[0,1]}, b_\# \eta)} \int x_0 y \, \mathrm{d}\sigma(x_0, y).$$

Can we re-formulate it as a multi-marginal problem?

Link between spectral measures and MMOT

Let $\mu_0 = \alpha_\sharp \mathrm{Leb}_{[0,1]}$, the other X_i and μ_i representing the domains and distributions of the underlying variables, respectively, and

$$s(x_0, x_1, \ldots, x_m) = x_0 b(x_1, \ldots, x_m).$$

Then the following result holds

Theorem

A probability measure γ in $\Pi(\mu_0, \mu_1, \ldots, \mu_m)$ is optimal for the MMOT problem with the cost function defined above if and only if its $(1, \ldots, N)$-marginal is optimal in $\max_{\eta \in \Pi(\mu_1, \mu_2, \ldots, \mu_m)} R_\alpha(b_\# \eta)$, and $\tau_\gamma = \Big((x_0, x_1, x_2, \ldots, x_m) \mapsto (x_0, b(x_1, x_2, \ldots, x_m))\Big)_\# \gamma$ has monotone increasing support.

Stability

We can also establish some stability results with respect to the marginals. Indeed we have

Lemma

If $\alpha \leq M$, then $\mu \in \mathcal{P}(\mathbb{R}) \mapsto R_\alpha(\mu)$ is M-Lipschitz for the 1-Wasserstein distance,

$$|R_\alpha(\mu) - R_\alpha(\nu)| \leq M\mathcal{W}_1(\mu, \nu).$$

and for the multi-marginal case

Proposition

Assume that the cost function b is k-Lipschitz with respect to $||\cdot||_p$ on $\mathbb{R}^d$ and that α is non-decreasing and bounded by M. Then,

$$|\sup_{\eta \in \Pi(\mu_{\mathbf{1}}, \ldots, \mu_m)} R_\alpha(b_\# \eta) - \sup_{\eta \in \Pi(\nu_{\mathbf{1}}, \ldots, \nu_m)} R_\alpha(b_\# \eta)| \leq M \left(k \sum_i \mathcal{W}_p(\mu_i, \nu_i) \right)^{1/p}$$

Solutions for one-dimensional assets and compatible outputs

Suppose that the output function b satisfies the following assumptions

- b is weak compatible;
- b monotone increasing in each $x_i \in S_+$ and monotone decreasing for each $x_i \in S_-$.

Then one can prove that the s-comonotone $(Id, G_1, \ldots, G_m)_{\#}\mathrm{Leb}_{[0,1]}$ maximizes $\sup_{\eta \in \Pi(\mu_1, \ldots, \mu_m)} R_\alpha(b_{\#}\eta)$ and the maximal value is given by

$$\int_0^1 \alpha(t) b(G_1(t), \cdots, G_m(t)) \mathrm{d}t.$$

Solutions for one-dimensional assets and compatible outputs

Suppose that the output function b satisfies the following assumptions

- b is weak compatible;
- b monotone increasing in each $x_i \in S_+$ and monotone decreasing for each $x_i \in S_-$.

Then one can prove that the s-comonotone $(Id, G_1, \ldots, G_m)_{\#}\mathrm{Leb}_{[0,1]}$ maximizes $\sup_{\eta \in \Pi(\mu_1, \ldots, \mu_m)} R_\alpha(b_{\#}\eta)$ and the maximal value is given by

$$\int_0^1 \alpha(t) b(G_1(t), \cdots, G_m(t)) \mathrm{d}t.$$

- Example : consider again the output function from **(Iooss and Lemaître 2015)**

$$S = Z_\nu + \left(\frac{Q}{BK_s \sqrt{\frac{Z_m - Z_\nu}{L}}} \right)^{0.6} - H_d - C_b,$$

up to a change of variable, it satisfies the assumption above. This implies that we have an explicit solution for this model!

Some extensions

Higher dimensional assets

- Suppose now that $X_i \subset \mathbb{R}^d$ with $d > 1$, $b : X_1 \times \cdots \times X_m \to \mathbb{R}$ and α as before.

Higher dimensional assets

- Suppose now that $X_i \subset \mathbb{R}^d$ with $d > 1$, $b : X_1 \times \cdots \times X_m \to \mathbb{R}$ and α as before.
- We have to deal with MMOT where when the underlying variables lie in more general spaces.

Higher dimensional assets

- Suppose now that $X_i \subset \mathbb{R}^d$ with $d > 1$, $b : X_1 \times \cdots \times X_m \to \mathbb{R}$ and α as before.
- We have to deal with MMOT where when the underlying variables lie in more general spaces.
- Warning: it is generally not possible to derive explicit solutions! But we can still prove that in some cases the solutions are of Monge type.

Higher dimensional assets

- Suppose now that $X_i \subset \mathbb{R}^d$ with $d > 1$, $b : X_1 \times \cdots \times X_m \to \mathbb{R}$ and α as before.
- We have to deal with MMOT where when the underlying variables lie in more general spaces.
- Warning: it is generally not possible to derive explicit solutions! But we can still prove that in some cases the solutions are of Monge type.

Proposition

Suppose that $m = 2$, μ_1 absolutely continuous with respect to Lebesgue measures and that $x_2 \mapsto D_{x_1} b(x_1, x_2)$ is injective for each fixed x_2, and that for each $(x_1, x_2) \in X_1 \times X_2$ we have

$$D_{x_2} b(x_1, x_2)[D^2_{x_1 x_2} b(x_1, x_2)]^{-1} D_{x_1} b(x_1, x_2) > 0.$$

Then the solution of $\sup_{\eta \in \Pi(\mu_1, \ldots, \mu_m)} R_\alpha(b_\# \eta)$ is concentrated on a graph, aka is of Monge type.

Multidimensional measures of risk

- We consider the framework in **(Ekeland, Alfred Galichon, and Henry 2012)** in which risk is measured in a multi-dimensional way. We have now a vector valued output function $b : X_1 \times \cdots \times X_m \to \mathbb{R}^d$.

Multidimensional measures of risk

- We consider the framework in **(Ekeland, Alfred Galichon, and Henry 2012)** in which risk is measured in a multi-dimensional way. We have now a vector valued output function $b : X_1 \times \cdots \times X_m \to \mathbb{R}^d$.
- Define a multi-variate risk measure on the distribution $b_{\#}\eta$ of output variables as in **(ibid.)** by

$$R_\alpha(b_{\#}\eta) = \max_{\sigma \in \Pi(\mu,\nu)} \int z \cdot y \, \mathrm{d}\sigma$$

where $\mu = b_\sharp \eta$ and $\nu = \alpha_\sharp \mathrm{Leb}$ with $\alpha : [0,1]^d \to \mathbb{R}^d$, we consider the problem of maximizing $R_\alpha(b_{\#}\eta)$ over all $\eta \in \Pi(\mu_1, \ldots, \mu_N)$.

Multidimensional measures of risk

• We consider the framework in **(Ekeland, Alfred Galichon, and Henry 2012)** in which risk is measured in a multi-dimensional way. We have now a vector valued output function $b : X_1 \times \cdots \times X_m \to \mathbb{R}^d$.

• Define a multi-variate risk measure on the distribution $b_{\#}\eta$ of output variables as in **(ibid.)** by

$$R_\alpha(b_{\#}\eta) = \max_{\sigma \in \Pi(\mu,\nu)} \int z \cdot y \, \mathrm{d}\sigma$$

where $\mu = b_\sharp \eta$ and $\nu = \alpha_\sharp \mathrm{Leb}$ with $\alpha : [0,1]^d \to \mathbb{R}^d$, we consider the problem of maximizing $R_\alpha(b_{\#}\eta)$ over all $\eta \in \Pi(\mu_1, \ldots, \mu_N)$.

Some results:

• If the underlying variables are one dimensional and b is supermodular that the solution is still of Monge type.

Multidimensional measures of risk

• We consider the framework in **(Ekeland, Alfred Galichon, and Henry 2012)** in which risk is measured in a multi-dimensional way. We have now a vector valued output function $b : X_1 \times \cdots \times X_m \to \mathbb{R}^d$.

• Define a multi-variate risk measure on the distribution $b_{\#}\eta$ of output variables as in **(ibid.)** by

$$R_\alpha(b_{\#}\eta) = \max_{\sigma \in \Pi(\mu,\nu)} \int z \cdot y \, \mathrm{d}\sigma$$

where $\mu = b_\sharp \eta$ and $\nu = \alpha_\sharp \mathrm{Leb}$ with $\alpha : [0,1]^d \to \mathbb{R}^d$, we consider the problem of maximizing $R_\alpha(b_{\#}\eta)$ over all $\eta \in \Pi(\mu_1, \ldots, \mu_N)$.

Some results:

• If the underlying variables are one dimensional and b is supermodular that the solution is still of Monge type.

• If $\nu \ll \mathrm{Leb}$, $m \leq d$ and b is invertible then there exists a unique solution concentrated on a graph of y.