An ODE characterisation of entropic (multi-marginal) optimal transport

Luca Nenna

joint work with B. Pass

Canum, 28/05/2024, Île de Ré

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

Overview

1. A crash introduction to (multi-marginal) Optimal Transport

   Classical Optimal transport

   Multi-marginal optimal transport

2. Connecting 2−marginal OT and MOT

   Characterisations of $\mathcal{W}^2_\nu(\mu_0, \mu_1)$

3. Entropic Multi-Marginal Optimal Transport

4. The algorithm and some numerical results

A crash introduction to (multi-marginal) Optimal Transport

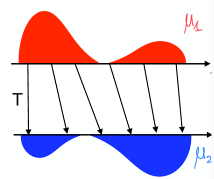
Classical Optimal Transportation Theory

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 $\gamma = (Id, T)_\sharp \mu$ where $T_\sharp \mu_1 = \mu_2$.



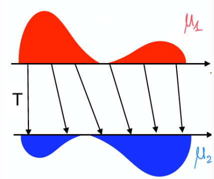
Classical Optimal Transportation Theory

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 $\gamma = (Id, T)_\sharp \mu$ where $T_\sharp \mu_1 = \mu_2$.



- Duality:

$$\sup\left\{ \mathcal{J}(\phi_1, \phi_2) \mid (\phi_1, \phi_2) \in \mathcal{K} \right\}. \tag{2}$$

where

$$\mathcal{J}(\phi_1, \phi_2) := \int_{X_1} \phi_1 \mathrm{d}\mu_1 + \int_{X_2} \phi_2 \mathrm{d}\mu_2$$

and $\mathcal{K}$ is the set of bounded and continuous functions (ϕ_1, ϕ_2) such that $\phi_1(x_1) + \phi(x_2) \leq c(x_1, x_2)$.

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:

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

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

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:

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

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 $T_{i\sharp}\mu_1 = \mu_i$.

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:

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

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? (the discrete case)

Let $c_{j_1, \cdots, j_m} = c(x_{j_1}, \cdots, x_{j_m}) \in \otimes_1^N \mathbb{R}^M$ (M is the number of gridpoints to discretize $\mathbb{R}^{d_i}$)

$$\inf_{\gamma \in \Pi(\mu_1, \dots, \mu_m)} \sum_{(j_1, \cdots, j_m)=1}^{M} c_{j_1, \cdots, j_m} \gamma_{j_1, \cdots, j_m}$$

M^m **unknowns!**

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **[Agueh, Carlier '11]**): statistics, machine learning, image processing;

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **[Agueh, Carlier '11]**): statistics, machine learning, image processing;
- Matching for teams problem (see **[Carlier,Ekeland '10]**): economics.

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **[Agueh, Carlier '11]**): statistics, machine learning, image processing;
- Matching for teams problem (see **[Carlier,Ekeland '10]**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **[Buttazzo, De Pascale, Gori-Giorgi '12]** and **[Cotar, Friesecke, Klüppelberg '13]**). 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, Carlier '11]**): statistics, machine learning, image processing;
- Matching for teams problem (see **[Carlier,Ekeland '10]**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **[Buttazzo, De Pascale, Gori-Giorgi '12]** and **[Cotar, Friesecke, Klüppelberg '13]**). The plan $\gamma(x_1, \dots, x_m)$ returns the probability of finding electrons at position $x_1, \dots, x_m$;
- Incompressible Euler Equations (see **[Brenier '89]**) : $\gamma(\omega)$ gives "the mass of fluid" which follows a path ω. See also **[Benamou, Carlier, N. '18]**.

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **[Agueh, Carlier '11]**): statistics, machine learning, image processing;
- Matching for teams problem (see **[Carlier,Ekeland '10]**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **[Buttazzo, De Pascale, Gori-Giorgi '12]** and **[Cotar, Friesecke, Klüppelberg '13]**). The plan $\gamma(x_1, \dots, x_m)$ returns the probability of finding electrons at position $x_1, \dots, x_m$;
- Incompressible Euler Equations (see **[Brenier '89]**) : $\gamma(\omega)$ gives "the mass of fluid" which follows a path ω. See also **[Benamou, Carlier, N. '18]**.
- Mean Field Games **[Benamou, Carlier, Di Marino, N. '18]**;

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **[Agueh, Carlier '11]**): statistics, machine learning, image processing;
- Matching for teams problem (see **[Carlier,Ekeland '10]**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **[Buttazzo, De Pascale, Gori-Giorgi '12]** and **[Cotar, Friesecke, Klüppelberg '13]**). The plan $\gamma(x_1, \dots, x_m)$ returns the probability of finding electrons at position $x_1, \dots, x_m$;
- Incompressible Euler Equations (see **[Brenier '89]**) : $\gamma(\omega)$ gives "the mass of fluid" which follows a path ω. See also **[Benamou, Carlier, N. '18]**.
- Mean Field Games **[Benamou, Carlier, Di Marino, N. '18]**;
- Risk measures **[Ennaji, Mérigot, N., Pass '23]**;

Why are we interested in MOT?

- The Wasserstein barycenter problem can be rewritten as a MMOT problem (see **[Agueh, Carlier '11]**): statistics, machine learning, image processing;
- Matching for teams problem (see **[Carlier,Ekeland '10]**): economics.
- In Density Functional Theory: the electron-electron repulsion (see **[Buttazzo, De Pascale, Gori-Giorgi '12]** and **[Cotar, Friesecke, Klüppelberg '13]**). The plan $\gamma(x_1, \dots, x_m)$ returns the probability of finding electrons at position $x_1, \dots, x_m$;
- Incompressible Euler Equations (see **[Brenier '89]**) : $\gamma(\omega)$ gives "the mass of fluid" which follows a path ω. See also **[Benamou, Carlier, N. '18]**.
- Mean Field Games **[Benamou, Carlier, Di Marino, N. '18]**;
- Risk measures **[Ennaji, Mérigot, N., Pass '23]**;
- Multi-period martingale transport, etc **[Hiew, N., Pass '24]**

Connecting $2-$marginal OT and MOT

The $\nu-$based Wasserstein distance

$\nu-$based Wasserstein distance

$$\mathcal{W}_\nu^2(\mu_1, \mu_2) = \inf \left\{ \int |x_1 - x_2|^2 d\gamma(y, x_1, x_2); \gamma \in \mathcal{P}(X^2 \times Y),\ \gamma_{y,x_i} \in \Pi_O(\mu_i, \nu) \right\},$$

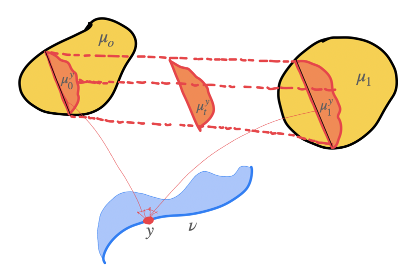
where $\Pi_O(\mu_i, \nu)$ is the set of optimal plans between μ_i and ν.

The $\nu-$based Wasserstein distance

$\nu-$based Wasserstein distance

$$\mathcal{W}_\nu^2(\mu_1,\mu_2) = \inf\left\{\int |x_1 - x_2|^2 d\gamma(y,x_1,x_2); \gamma \in \mathcal{P}(X^2 \times Y),\ \gamma_{y,x_i} \in \Pi_O(\mu_i,\nu)\right\},$$

where $\Pi_O(\mu_i,\nu)$ is the set of optimal plans between μ_i and ν.

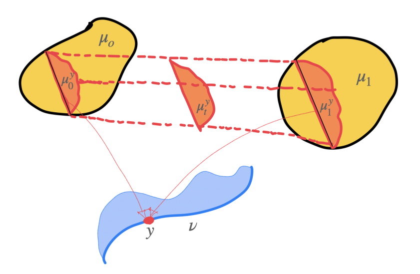


The $\nu-$based Wasserstein distance

$\nu-$based Wasserstein distance

$$\mathcal{W}_\nu^2(\mu_1,\mu_2) = \inf\left\{\int |x_1 - x_2|^2 d\gamma(y,x_1,x_2); \gamma \in \mathcal{P}(X^2 \times Y),\ \gamma_{y,x_i} \in \Pi_O(\mu_i,\nu)\right\},$$

where $\Pi_O(\mu_i,\nu)$ is the set of optimal plans between μ_i and ν.



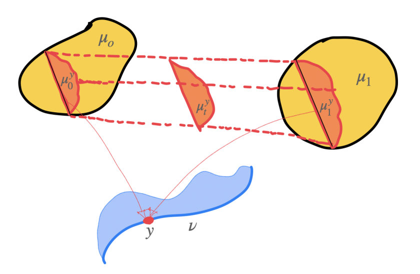
- This is only a **semi-metric** in general, that is the triangle inequality fails.

The $\nu-$based Wasserstein distance

$\nu-$based Wasserstein distance

$$\mathcal{W}_\nu^2(\mu_1,\mu_2) = \inf\left\{\int |x_1 - x_2|^2 d\gamma(y,x_1,x_2); \gamma \in \mathcal{P}(X^2 \times Y),\ \gamma_{y,x_i} \in \Pi_O(\mu_i,\nu)\right\},$$

where $\Pi_O(\mu_i,\nu)$ is the set of optimal plans between μ_i and ν.



- This is only a **semi-metric** in general, that is the triangle inequality fails.
- If we restrict to the class $\mathcal{P}_\nu^u(X) \subseteq \mathcal{P}(X)$ of probability measures on X for which optimal transport is unique, then it is a metric. If $\boldsymbol{m} = \boldsymbol{n}$ it coincides with the Linear Optimal Transport distance introduced in **[Wang,Slepčev, Basu, Ozolek, and Rohde '13]**.

Characterisations of $\mathcal{W}_\nu^2(\mu_1, \mu_2)$

- Via **conditional probabilities**

$$\mathcal{W}_\nu^2(\mu_1, \mu_2) = \inf \left\{ \int_X \mathcal{W}_2^2(\mu_1^y, \mu_2^y) \mathrm{d}\nu(y) \mid \pi_i \in \Pi_O(\mu_i, \nu),\ i = 1, 2 \right\},$$

where μ_i^y is the conditional probability given y of the optimal coupling $\pi_i(x, y) = \nu(y) \otimes \mu_i^y(x)$ between ν and μ_i.

- Via **conditional probabilities**

$$\mathcal{W}^2_\nu(\mu_1, \mu_2) = \inf \left\{ \int_X \mathcal{W}^2_2(\mu_1^y, \mu_2^y)\mathrm{d}\nu(y) \mid \pi_i \in \Pi_O(\mu_i, \nu),\ i = 1, 2 \right\},$$

where μ_i^y is the conditional probability given y of the optimal coupling $\pi_i(x, y) = \nu(y) \otimes \mu_i^y(x)$ between ν and μ_i.

Via Multi-Marginal OT

any weak limit point $\bar{\gamma}$ as $\eta \to 0$ of minimizers γ_η of the multi-marginal problem

$$\inf_{\gamma \in \Pi(\nu, \mu_1, \mu_2)} \int_{X^3} (\eta|x_1 - x_2|^2 + |x_1 - y|^2 + |x_2 - y|^2)\mathrm{d}\gamma(y, x_1, x_2)$$

is an optimal coupling between μ_1 and μ_2 for the problem defining $\mathcal{W}^2_\nu(\mu_1, \mu_2)$.

Entropic Multi-Marginal Optimal Transport

Entropic Multi-Marginal Optimal Transport

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 $\mathsf{X} := \times_i^m X_i$;

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:

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

where

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

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

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

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

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

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

- The solution γ_ε is "almost" explicit

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

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

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

- The solution γ_ε is "almost" explicit

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

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

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

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

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

- The solution γ_ε is "almost" explicit

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

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

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

- **More on entropic transport on Friday!**

What are we interested in and direction of our work

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

What are we interested in and direction of our work

We are interested in solving the **entropic discrete** 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 discrete** 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. Some assumptions to make it simple:
   2.1 **(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.
   2.2 **(Pair-wise cost)** $c_\eta(x_1, \dots, x_m) := \eta \sum_{i=2}^{m} \sum_{j=i+1}^{m} w(x_i, x_j) + \boldsymbol{\sum_{i=2}^{m} w(x_1, x_i)}$.
   2.3 **(Symmetric cost)** The two body cost w is symmetric $w(x, y) = w(x, y)$.
   2.4 **(Finite cost)** The two body cost function $w : X \times X \to \mathbb{R}$ is everywhere real-valued.

What are we interested in and direction of our work

We are interested in solving the **entropic discrete** 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. Some assumptions to make it simple:
   2.1 **(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.
   2.2 **(Pair-wise cost)** $c_\eta(x_1, \dots, x_m) := \eta \sum_{i=2}^{m} \sum_{j=i+1}^{m} w(x_i, x_j) + \boldsymbol{\sum_{i=2}^{m} w(x_1, x_i)}$.
   2.3 **(Symmetric cost)** The two body cost w is symmetric $w(x, y) = w(x, y)$.
   2.4 **(Finite cost)** The two body cost function $w : X \times X \to \mathbb{R}$ is everywhere real-valued.

Rmk: We can drop hypothesis 2.1→2.3.

How to derive the differential equation

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

$$\inf_{\varphi} \left\{ \Psi(\varphi, \eta) \right\}. \tag{4}$$

How to derive the differential equation

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

$$\inf_{\varphi} \left\{ \Psi(\varphi, \eta) \right\}. \tag{4}$$

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

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

How to derive the differential equation

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

$$\inf_{\varphi} \{\Psi(\varphi, \eta)\} . \tag{4}$$

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

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

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 $\varphi(0) = \varphi_0$.

Idea of the proof: fix the value of the potential in one point and then show that Ψ is strongly convex.

The algorithm and some numerical results

The algorithm to compute the ODE solution

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

Require: $\phi(0) = \phi_w$

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

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

$\quad b^{(k)} := -\frac{\partial}{\partial \epsilon}\nabla_\phi \tilde{\Phi}(\phi^{(k)}, kh)$

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

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

end while

The algorithm to compute the ODE solution

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

Require: $\phi(0) = \phi_w$

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

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

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

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

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

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: $\phi(0) = \phi_w$

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

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

$\quad b^{(k)} := -\frac{\partial}{\partial \epsilon}\nabla_{\phi}\tilde{\Phi}(\phi^{(k)}, kh)$

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

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

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

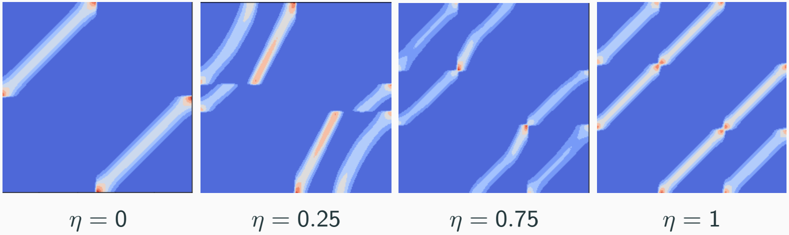


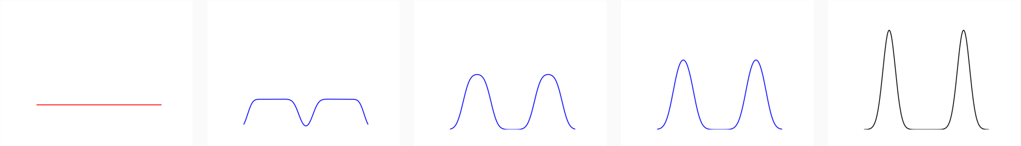
Figure 1: Support of the coupling $\gamma_{1,2}^{\eta}$.

Some remarks

1. By taking $c_\eta = \eta c$ we have an interpolation **(1)** between different costs and **(2)** between the solution $\otimes_{i=1}^{m} \mu_i$ (when entropy dominates) and the one to optimal transport;

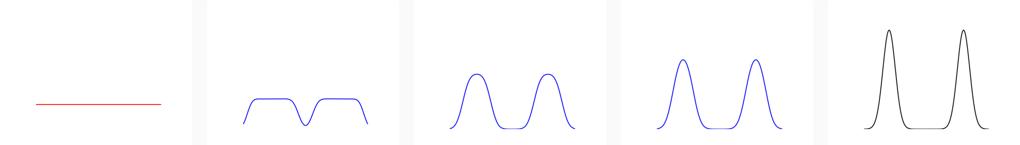
Some remarks

1. By taking $c_\eta = \eta c$ we have an interpolation **(1)** between different costs and **(2)** between the solution $\otimes_{i=1}^m \mu_i$ (when entropy dominates) and the one to optimal transport;
2. $c_\eta(x_1, z, x_2) = (1-\eta)|x_1 - z|^2 + \eta|z - x_3|^2$, 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.



Some remarks

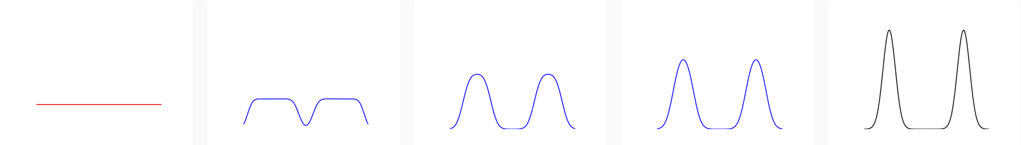
1. By taking $c_\eta = \eta c$ we have an interpolation **(1)** between different costs and **(2)** between the solution $\otimes_{i=1}^{m}\mu_i$ (when entropy dominates) and the one to optimal transport;
2. $c_\eta(x_1, z, x_2) = (1-\eta)|x_1 - z|^2 + \eta|z - x_3|^2$, γ is a 3 marginals coupling with only two fixed marginals, μ_1 and μ_2. Then the $z-$marginal of γ gives the Wasserstein geodesic at time η.



3. One can also compute the **Wasserstein barycenter**.

Some remarks

1. By taking $c_\eta = \eta c$ we have an interpolation **(1)** between different costs and **(2)** between the solution $\otimes_{i=1}^m \mu_i$ (when entropy dominates) and the one to optimal transport;
2. $c_\eta(x_1, z, x_2) = (1-\eta)|x_1 - z|^2 + \eta|z - x_3|^2$, 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.



3. One can also compute the **Wasserstein barycenter**.
4. We can add extra linear constraints to treat other generalization of optimal transport: e.g. martingale OT, etc

Conclusion

Take-home messages:

- An ODE to characterize entropic multi-marginal optimal transport;

Conclusion

Take-home messages:

- An ODE to characterize entropic multi-marginal optimal transport;
- It works for symmetric and non symmetric cost;

Conclusion

Take-home messages:

- An ODE to characterize entropic multi-marginal optimal transport;
- It works for symmetric and non symmetric cost;
- Regularity allows to use high order methods;

Conclusion

Take-home messages:

- An ODE to characterize entropic multi-marginal optimal transport;
- It works for symmetric and non symmetric cost;
- Regularity allows to use high order methods;
- It allows to interpolate between different costs and for each $\eta \in [0, 1]$ it returns the solution to the corresponding multi-marginal problems.

Conclusion

Take-home messages:

- An ODE to characterize entropic multi-marginal optimal transport;
- It works for symmetric and non symmetric cost;
- Regularity allows to use high order methods;
- It allows to interpolate between different costs and for each $\eta \in [0, 1]$ it returns the solution to the corresponding multi-marginal problems.
- Wasserstein geodesics, Barycenter problem and Martingale transport;

Conclusion

Take-home messages:

- An ODE to characterize entropic multi-marginal optimal transport;
- It works for symmetric and non symmetric cost;
- Regularity allows to use high order methods;
- It allows to interpolate between different costs and for each $\eta \in [0,1]$ it returns the solution to the corresponding multi-marginal problems.
- Wasserstein geodesics, Barycenter problem and Martingale transport;

Thank You!!