Numerical methods for Multi-Marginal Optimal Transport

From geodesics in Wasserstein space to variational Mean Field Games

Luca Nenna

(LMO) Université Paris-Saclay

Lecture 4, 09/03/2022, Orsay

Overview

Overview

Overview

1. Introduction: Classical vs Multi-Marginal Optimal Transport
   - The three universes of Numerical Optimal Transportation
   - The discretized problem
2. Entropic Optimal Transport
   - The numerical method
   - How the regularization works
   - Sinkhornizing the world!!
3. Application I: MMOT for computing geodesics in the Wasserstein space
4. Application II: MMOT and variational Mean Field Games
   - Eulerian and Lagrangian formulation for MFG with quadratic Hamiltonian

Overview

Introduction: Classical vs Multi-Marginal Optimal Transport

Classical Optimal Transportation Theory

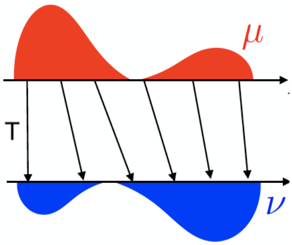
Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ ($X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^n$), the Optimal Transport (OT) problem is defined as follows

$$(\mathcal{MK}) \quad E_c(\mu,\nu) = \inf \{\mathcal{E}_c(\gamma) \mid \gamma \in \Pi(\mu,\nu)\} \tag{1}$$

where $\Pi(\mu,\nu) := \{\gamma \in \mathcal{P}(X \times Y) | \quad \pi_{1,\sharp}\gamma = \mu, \ \pi_{2,\sharp}\gamma = \nu\}$ and

$$\mathcal{E}_c(\gamma) := \int c(x_1,x_2)d\gamma(x_1,x_2).$$

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



The Multi-Marginal Optimal Transportation

Let us take N probability measures $\mu_i \in \mathcal{P}(X)$ with $i = 1, \cdots, N$ and $c : X^N \to [0, +\infty]$ a continuous cost function. Then the multi-marginal OT problem reads as:

$$(\mathcal{MK}_N) \quad E_c(\mu_1, \cdots, \mu_N) = \inf \{\mathcal{E}_c(\gamma) \mid \gamma \in \Pi_N(\mu_1, \cdots, \mu_N)\} \qquad (2)$$

where $\Pi_N(\mu_1, \cdots, \mu_N)$ denotes the set of couplings $\gamma(x_1, \cdots, x_N)$ having μ_i as marginals and

$$\mathcal{E}_c(\gamma) := \int c(x_1, \cdots, x_N) d\gamma(x_1, \cdots, x_N)$$

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

Why is it a difficult problem to treat?

Example : $N = 3$, $d = 1$, $\mu_i = \mathcal{L}_{[0,1]}$ $\forall i$ 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**

The Multi-Marginal Optimal Transportation

Let us take N probability measures $\mu_i \in \mathcal{P}(X)$ with $i = 1, \cdots, N$ and $c : X^N \to [0, +\infty]$ a continuous cost function. Then the multi-marginal OT problem reads as:

$$(\mathcal{MK}_N) \quad E_c(\mu_1, \cdots, \mu_N) = \inf \{\mathcal{E}_c(\gamma) \mid \gamma \in \Pi_N(\mu_1, \cdots, \mu_N)\} \tag{2}$$

where $\Pi_N(\mu_1, \cdots, \mu_N)$ denotes the set of couplings $\gamma(x_1, \cdots, x_N)$ having μ_i as marginals and

$$\mathcal{E}_c(\gamma) := \int c(x_1, \cdots, x_N) d\gamma(x_1, \cdots, x_N)$$

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

Why is it a difficult problem to treat?

Example : $N = 3$, $d = 1$, $\mu_i = \mathcal{L}_{[0,1]}$ $\forall i$ 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**

The Multi-Marginal Optimal Transportation

Let us take N probability measures $\mu_i \in \mathcal{P}(X)$ with $i = 1, \cdots, N$ and $c : X^N \to [0, +\infty]$ a continuous cost function. Then the multi-marginal OT problem reads as:

$$(\mathcal{MK}_N) \quad E_c(\mu_1, \cdots, \mu_N) = \inf \{\mathcal{E}_c(\gamma) \mid \gamma \in \Pi_N(\mu_1, \cdots, \mu_N)\} \tag{2}$$

where $\Pi_N(\mu_1, \cdots, \mu_N)$ denotes the set of couplings $\gamma(x_1, \cdots, x_N)$ having μ_i as marginals and

$$\mathcal{E}_c(\gamma) := \int c(x_1, \cdots, x_N) d\gamma(x_1, \cdots, x_N)$$

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

Why is it a difficult problem to treat?

Example : $N = 3$, $d = 1$, $\mu_i = \mathcal{L}_{[0,1]}$ $\forall i$ 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**

The dual formulation of $(\mathcal{MK})$

We consider the 2 marginals case for simplicity. The $(\mathcal{MK})$ problem admits a dual formulation:

$$\sup \{\mathcal{J}(\phi, \psi) \mid (\phi, \psi) \in \mathcal{K}\}. \tag{3}$$

where

$$\mathcal{J}(\phi, \psi) := \int_X \phi d\mu(x) + \int_Y \psi d\nu(y)$$

and $\mathcal{K}$ is the set of bounded and continuous functions ϕ, ψ such that $\phi(x) + \psi(y) \leq c(x, y)$.

Remark

Notice that the constraint on a couple (ϕ, ψ) may be rewritten as

$$\psi(y) \leq \inf_x c(x, y) - \phi(x) := \phi^c(y).$$

So for an admissible couple (ϕ, ψ) one has $\mathcal{J}(\phi, \phi^c) \geq \mathcal{J}(\phi, \psi)$

The dual formulation of $(\mathcal{MK})$

We consider the 2 marginals case for simplicity. The $(\mathcal{MK})$ problem admits a dual formulation:

$$\sup \{\mathcal{J}(\phi, \psi) \mid (\phi, \psi) \in \mathcal{K}\}. \tag{3}$$

where

$$\mathcal{J}(\phi, \psi) := \int_X \phi d\mu(x) + \int_Y \psi d\nu(y)$$

and $\mathcal{K}$ is the set of bounded and continuous functions ϕ, ψ such that $\phi(x) + \psi(y) \leq c(x, y)$.

Remark

Notice that the constraint on a couple (ϕ, ψ) may be rewritten as

$$\psi(y) \leq \inf_x c(x, y) - \phi(x) := \phi^c(y).$$

So for an admissible couple (ϕ, ψ) one has $\mathcal{J}(\phi, \phi^c) \geq \mathcal{J}(\phi, \psi)$

Some applications

- 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. The transport plan γ matches individuals from each team μ_i minimizing a given cost;
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \cdots, x_N)$ returns the probability of finding electrons at position $x_1, \cdots, x_N$;
- Incompressible Euler Equations **(Yann 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)**.
- Variational Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;
- etc...

Some applications

- 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. The transport plan γ matches individuals from each team μ_i minimizing a given cost;
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \cdots, x_N)$ returns the probability of finding electrons at position $x_1, \cdots, x_N$;
- Incompressible Euler Equations **(Yann 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)**.
- Variational Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;
- etc...

Some applications

- 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. The transport plan γ matches individuals from each team μ_i minimizing a given cost;
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \cdots, x_N)$ returns the probability of finding electrons at position $x_1, \cdots, x_N$;
- Incompressible Euler Equations **(Yann 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)**.
- Variational Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;
- etc...

Some applications

- 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. The transport plan γ matches individuals from each team μ_i minimizing a given cost;
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \cdots, x_N)$ returns the probability of finding electrons at position $x_1, \cdots, x_N$;
- Incompressible Euler Equations **(Yann 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)**.
- Variational Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;
- etc...

Some applications

- 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. The transport plan γ matches individuals from each team μ_i minimizing a given cost;
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \cdots, x_N)$ returns the probability of finding electrons at position $x_1, \cdots, x_N$;
- Incompressible Euler Equations **(Yann 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)**.
- Variational Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;
- etc...

Some applications

- 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. The transport plan γ matches individuals from each team μ_i minimizing a given cost;
- In Density Functional Theory: the electron-electron repulsion (see **(Buttazzo, De Pascale, and Gori-Giorgi 2012; Cotar, Friesecke, and Klüppelberg 2013)**). The plan $\gamma(x_1, \cdots, x_N)$ returns the probability of finding electrons at position $x_1, \cdots, x_N$;
- Incompressible Euler Equations **(Yann 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)**.
- Variational Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;
- etc...

The three universes of Numerical Optimal Transportation

Let's consider the two marginal case then we can have the three following numerical approach to Optimal Transport

- Discrete-2-Discrete: the marginals μ have an atomic form, i.e. $\mu(x) = \sum_i \mu_i \delta_{x_i}$ (and ν as well). **Remarks:**
  - The problem becomes a standard linear programming problem.
  - Works for any kind of cost function.
  - Can be easily generalized to the multi-marginal case.
- Continous-2-Discrete: $\mu = \bar{\mu} dx$ and $\nu(y) = \sum_i \nu_i \delta_{y_i}$. **Remarks:**
  - The semi-discrete approach **(Mérigot 2011)**.
  - Used for generalized euler equations (kind of mmot problem) à la Brenier **(Mérigot and Mirebeau 2016)**.
- Continous-2-Continous: $\mu = \bar{\mu} dx$ (and ν too). **Remarks:**
  - The Benamou-Brenier formulation for Optimal Transport (J.-D. Benamou and Y. Brenier 2000)

The three universes of Numerical Optimal Transportation

Let's consider the two marginal case then we can have the three following numerical approach to Optimal Transport

- Discrete-2-Discrete: the marginals μ have an atomic form, i.e. $\mu(x) = \sum_i \mu_i \delta_{x_i}$ (and ν as well). **Remarks:**
  - The problem becomes a standard linear programming problem.
  - Works for any kind of cost function.
  - Can be easily generalized to the multi-marginal case.
- Continous-2-Discrete: $\mu = \bar{\mu} dx$ and $\nu(y) = \sum_i \nu_i \delta_{y_i}$. **Remarks:**
  - The semi-discrete approach **(Mérigot 2011)**.
  - Used for generalized euler equations (kind of mmot problem) à la Brenier **(Mérigot and Mirebeau 2016)**.
- Continous-2-Continous $\mu = \bar{\mu} dx$ (and ν too). **Remarks**
  - The Benamou-Brenier formulation for Optimal Transport! **(J.-D. Benamou and Y. Brenier 2000)**

The three universes of Numerical Optimal Transportation

Let's consider the two marginal case then we can have the three following numerical approach to Optimal Transport

- Discrete-2-Discrete: the marginals μ have an atomic form, i.e. $\mu(x) = \sum_i \mu_i \delta_{x_i}$ (and ν as well). **Remarks:**
  - The problem becomes a standard linear programming problem.
  - Works for any kind of cost function.
  - Can be easily generalized to the multi-marginal case.
- Continous-2-Discrete: $\mu = \bar{\mu} dx$ and $\nu(y) = \sum_i \nu_i \delta_{y_i}$. **Remarks:**
  - The semi-discrete approach **(Mérigot 2011)**.
  - Used for generalized euler equations (kind of mmot problem) à la Brenier **(Mérigot and Mirebeau 2016)**.
- Continous-2-Continous $\mu = \bar{\mu} dx$ (and ν too). **Remarks**
  - The Benamou-Brenier formulation for Optimal Transport! **(J.-D. Benamou and Y. Brenier 2000)**

The three universes of Numerical Optimal Transportation

Let's consider the two marginal case then we can have the three following numerical approach to Optimal Transport

- Discrete-2-Discrete: the marginals μ have an atomic form, i.e. $\mu(x) = \sum_i \mu_i \delta_{x_i}$ (and ν as well). **Remarks:**
  - The problem becomes a standard linear programming problem.
  - Works for any kind of cost function.
  - Can be easily generalized to the multi-marginal case.
- Continous-2-Discrete: $\mu = \bar{\mu} dx$ and $\nu(y) = \sum_i \nu_i \delta_{y_i}$. **Remarks:**
  - The semi-discrete approach **(Mérigot 2011)**.
  - Used for generalized euler equations (kind of mmot problem) à la Brenier **(Mérigot and Mirebeau 2016)**.
- Continous-2-Continous $\mu = \bar{\mu} dx$ (and ν too). **Remarks**
  - The Benamou-Brenier formulation for Optimal Transport! **(J.-D. Benamou and Y. Brenier 2000)**

The discretized Monge-Kantorovich problem

Let's take $c_{ij} = c(x_i, y_j) \in \mathbb{R}^{M\times M}$ (M are the gridpoints used to discretize X) then the discretized $(\mathcal{MK})$, reads as

$$\min\{\sum_{i,j=1}^{M} c_{ij}\gamma_{ij} \mid \sum_{j=1}^{M}\gamma_{ij} = \mu_i \ \forall i, \ \sum_{i=1}^{M}\gamma_{ij} = \nu_j \ \forall j\} \tag{4}$$

and the dual problem

$$\max\{\sum_{i=1}^{M}\phi_i\mu_i + \sum_{j=1}^{M}\psi_j\nu_j \mid \phi_i + \psi_j \leq c_{ij} \ \forall (i,j) \in \{1,\cdots,M\}^2\}. \tag{5}$$

Remarks

- The primal has M^2 unknowns and $M \times 2$ linear constraints.
- The dual has $M \times 2$ unknowns, but M^2 constraints.

The discretized Monge-Kantorovich problem

Let's take $c_{ij} = c(x_i, y_j) \in \mathbb{R}^{M \times M}$ (M are the gridpoints used to discretize X) then the discretized $(\mathcal{MK})$, reads as

$$\min\{\sum_{i,j=1}^{M} c_{ij}\gamma_{ij} \mid \sum_{j=1}^{M} \gamma_{ij} = \mu_i \ \forall i, \ \sum_{i=1}^{M} \gamma_{ij} = \nu_j \ \forall j\} \tag{4}$$

and the dual problem

$$\max\{\sum_{i=1}^{M} \phi_i \mu_i + \sum_{j=1}^{M} \psi_j \nu_j \mid \phi_i + \psi_j \leq c_{ij} \ \forall (i,j) \in \{1, \cdots, M\}^2\}. \tag{5}$$

Remarks

- The primal has M^2 unknowns and $M \times 2$ linear constraints.
- The dual has $M \times 2$ unknowns, but M^2 constraints.

The importance of being sparse

A multi-scale approach to reduce M **(J.-D. Benamou, G. Carlier, and L. Nenna 2016)**

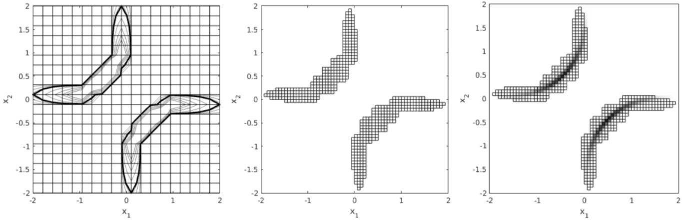


Figure: Support of the optimal γ for 2 marginals and the Coulomb cost

Some references:

Schmitzer, Bernhard (2019). "Stabilized sparse scaling algorithms for entropy regularized transport problems". In: *SIAM J. Sci. Comput.* 41.3, A1443–A1481. ISSN: 1064-8275. DOI: 10.1137/16M1106018. URL: https://mathscinet.ams.org/mathscinet-getitem?mr=3947294.

Mérigot, Quentin (2011). "A multiscale approach to optimal transport". In: *Computer Graphics Forum*. Vol. 30. 5. Wiley Online Library, pp. 1583–1592.

The discretized Monge-Kantorovich problem

Let's take $c_{j_1,\cdots,j_N} = c(x_{j_1},\cdots,x_{j_N}) \in \otimes_1^N \mathbb{R}^M$ (M are the gridpoints used to discretize $\mathbb{R}^d$) then the discretized $(\mathcal{MK}_N)$, reads as

$$\min\{\sum_{(j_1,\cdots,j_N)=1}^{M} c_{j_1,\cdots,j_N}\gamma_{j_1,\cdots,j_N} \quad | \quad \sum_{j_k,k\neq i}\gamma_{j_1,\cdots,j_{i-1},j_{i+1},\cdots,j_N} = \mu^i_{j_i}\} \tag{6}$$

and the dual problem

$$\max\{\sum_{i=1}^{N}\sum_{j_i=1}^{M} u^i_{j_i}\mu^i_{j_i} \quad | \quad \sum_{k=1}^{N} u^k_{j_k} \leq c_{j_1,\dots,j_N} \quad \forall (j_1,\cdots,j_N)\in\{1,\cdots,M\}^N\}. \tag{7}$$

Drawbacks

- The primal has M^N unknowns and $M\times N$ linear constraints.
- The dual has $M\times N$ unknowns, but M^N constraints.

université PARIS-SACLAY

The discretized Monge-Kantorovich problem

Let's take $c_{j_1,\cdots,j_N} = c(x_{j_1},\cdots,x_{j_N}) \in \otimes_1^N \mathbb{R}^M$ (M are the gridpoints used to discretize $\mathbb{R}^d$) then the discretized $(\mathcal{MK}_N)$, reads as

$$\min\{ \sum_{(j_1,\cdots,j_N)=1}^{M} c_{j_1,\cdots,j_N}\gamma_{j_1,\cdots,j_N} \quad | \quad \sum_{j_k,k\neq i} \gamma_{j_1,\cdots,j_{i-1},j_{i+1},\cdots,j_N} = \mu^i_{j_i} \} \tag{6}$$

and the dual problem

$$\max\{\sum_{i=1}^{N}\sum_{j_i=1}^{M} u^i_{j_i}\mu^i_{j_i} \quad | \quad \sum_{k=1}^{N} u^k_{j_k} \leq c_{j_1,\dots,j_N} \quad \forall (j_1,\cdots,j_N) \in \{1,\cdots,M\}^N\}. \tag{7}$$

Drawbacks

- The primal has M^N unknowns and $M \times N$ linear constraints.
- The dual has $M \times N$ unknowns, but M^N constraints.

Entropic Optimal Transport

The entropic OT problem

We present a numerical method to solve the regularized (**(Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Peyré 2015; M. Cuturi 2013; Galichon and Salanié 2009)**) optimal transport problem (let us consider, for simplicity, 2 marginals)

$$\min_{\gamma \in \mathcal{C}} \sum_{i,j} c_{ij}\gamma_{ij} + \begin{cases} \epsilon \sum_{ij} \gamma_{ij} \log\left(\frac{\gamma_{ij}}{\mu_i \nu_j}\right) & \gamma \geq 0 \\ +\infty & otherwise \end{cases} . \tag{8}$$

where C is the matrix associated to the cost, γ is the discrete transport plan and $\mathcal{C}$ is the intersection between $\mathcal{C}_1 = \{\gamma \mid \sum_j \gamma_{ij} = \mu_i\}$ and $\mathcal{C}_2 = \{\gamma \mid \sum_i \gamma_{ij} = \nu_j\}$.

Remark: Think at ϵ as the temperature, then entropic OT is just OT at positive temperature.

The problem (8) can be re-written as

$$\min_{\gamma \in \mathcal{C}} \mathcal{H}(\gamma|\bar{\gamma}) \tag{9}$$

where $\mathcal{H}(\gamma|\bar{\gamma}) = \sum_{ij} \gamma_{ij} \left(\log \frac{\gamma_{ij}}{\bar{\gamma}_{ij}} \right)$ $(= \mathrm{KL}(\gamma|\bar{\gamma})$ aka the Kullback-Leibler divergence) and $\bar{\gamma}_{ij} = e^{-\frac{c_{ij}}{\epsilon}} \mu_i \nu_j$.

Remarks:

- **Unique and semi-explicit solution** (we will see it in 2/3 minutes!)
- Problem (9) dates back to Schrodinger, (see Christian Léonard's web page).
- $\mathcal{H} \to \mathcal{MK}$ as $\epsilon \to 0$. (see **(Guillaume Carlier, Duval, Peyré, and Bernhard Schmitzer 2017; Léonard 2012)**).
- The dual problem is an unconstrained optimization problem

The problem (8) can be re-written as

$$\min_{\gamma \in \mathcal{C}} \mathcal{H}(\gamma|\bar{\gamma}) \tag{9}$$

where $\mathcal{H}(\gamma|\bar{\gamma}) = \sum_{ij} \gamma_{ij} \left(\log \frac{\gamma_{ij}}{\bar{\gamma}_{ij}} \right)$ ($= \mathrm{KL}(\gamma|\bar{\gamma})$ aka the Kullback-Leibler divergence) and $\bar{\gamma}_{ij} = e^{-\frac{c_{ij}}{\epsilon}} \mu_i \nu_j$.

Remarks:

- **Unique and semi-explicit solution** (we will see it in 2/3 minutes!)
- Problem (9) dates back to Schrödinger, (see Christian Léonard's web page).
- $\mathcal{H} \to \mathcal{MK}$ as $\epsilon \to 0$. (see **(Guillaume Carlier, Duval, Peyré, and Bernhard Schmitzer 2017; Léonard 2012)**).
- The dual problem is an unconstrained optimization problem.

The problem (8) can be re-written as

$$\min_{\gamma \in \mathcal{C}} \mathcal{H}(\gamma|\bar{\gamma}) \tag{9}$$

where $\mathcal{H}(\gamma|\bar{\gamma}) = \sum_{ij} \gamma_{ij} \left(\log \frac{\gamma_{ij}}{\bar{\gamma}_{ij}} \right)$ $(= \mathrm{KL}(\gamma|\bar{\gamma})$ aka the Kullback-Leibler divergence) and $\bar{\gamma}_{ij} = e^{-\frac{c_{ij}}{\epsilon}} \mu_i \nu_j$.

Remarks:

- **Unique and semi-explicit solution** (we will see it in 2/3 minutes!)
- Problem (9) dates back to Schrödinger, (see Christian Léonard's web page).
- $\mathcal{H} \to \mathcal{MK}$ as $\epsilon \to 0$. (see **(Guillaume Carlier, Duval, Peyré, and Bernhard Schmitzer 2017; Léonard 2012)**).
- The dual problem is an unconstrained optimization problem.

The problem (8) can be re-written as

$$\min_{\gamma \in \mathcal{C}} \mathcal{H}(\gamma|\bar{\gamma}) \tag{9}$$

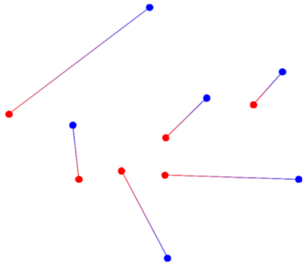
where $\mathcal{H}(\gamma|\bar{\gamma}) = \sum_{ij} \gamma_{ij} \left(\log \frac{\gamma_{ij}}{\bar{\gamma}_{ij}} \right)$ $(= \mathrm{KL}(\gamma|\bar{\gamma})$ aka the Kullback-Leibler divergence) and $\bar{\gamma}_{ij} = e^{-\frac{c_{ij}}{\epsilon}} \mu_i \nu_j$.

Remarks:

- **Unique and semi-explicit solution** (we will see it in 2/3 minutes!)
- Problem (9) dates back to Schrödinger, (see Christian Léonard's web page).
- $\mathcal{H} \to \mathcal{MK}$ as $\epsilon \to 0$. (see **(Guillaume Carlier, Duval, Peyré, and Bernhard Schmitzer 2017; Léonard 2012)**).
- The dual problem is an unconstrained optimization problem.

The “bridge” between quadratic Monge-Kantorovich and Schrödinger

From deterministic to stochastic matching **(Léonard 2012)**

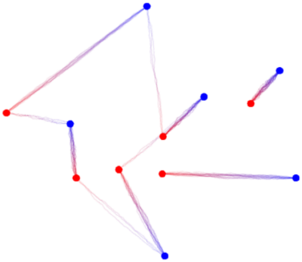


$\varepsilon = 0$

Figure: G. Peyre's twitter account

The "bridge" between quadratic Monge-Kantorovich and Schrödinger

From deterministic to stochastic matching **(Léonard 2012)**

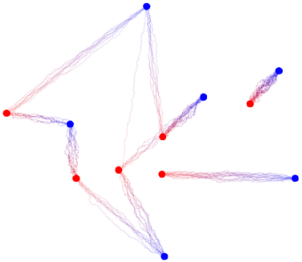


$$\varepsilon = .05$$

Figure: G. Peyre's twitter account

The "bridge" between quadratic Monge-Kantorovich and Schrödinger

From deterministic to stochastic matching **(Léonard 2012)**

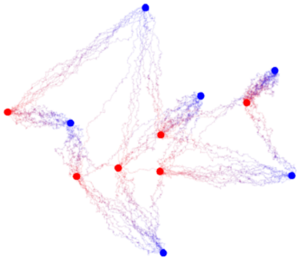


$$\varepsilon = 0.2$$

Figure: G. Peyre's twitter account

The "bridge" between quadratic Monge-Kantorovich and Schrödinger

From deterministic to stochastic matching **(Léonard 2012)**



$$\varepsilon = 1$$

Figure: G. Peyre's twitter account

The Sinkhorn algorithm

Theorem ((Franklin and Lorenz 1989))

The optimal plan $\gamma^\star$ has the form $\gamma^\star_{ij} = a^\star_i b^\star_j \bar{\gamma}_{ij}$. Moreover $a^\star_i$ and $b^\star_j$ can be uniquely determined (up to a multiplicative constant) as follows

$$b^\star_j = \frac{\nu_j}{\sum_i a^\star_i \bar{\gamma}_{ij}}, \; a^\star_i = \frac{\mu_i}{\sum_j b^\star_j \bar{\gamma}_{ij}}$$

The Sinkhorn algorithm (aka IPFP)

$$b^{n+1}_j = \frac{\nu_j}{\sum_i a^n_i \bar{\gamma}_{ij}}, \; a^{n+1}_i = \frac{\mu_i}{\sum_j b^{n+1}_j \bar{\gamma}_{ij}}$$

Theorem ((ibid.))

a^n and b^n converge to $a^\star$ and $b^\star$

Remark: $\phi_i = \epsilon \log(a_i)$ and $\psi_j = \epsilon \log(b_j)$ are the (regularized) Kantorovich potentials.

The Sinkhorn algorithm

Theorem ((Franklin and Lorenz 1989))

The optimal plan $\gamma^\star$ has the form $\gamma^\star_{ij} = a^\star_i b^\star_j \bar{\gamma}_{ij}$. Moreover $a^\star_i$ and $b^\star_j$ can be uniquely determined (up to a multiplicative constant) as follows

$$b^\star_j = \frac{\nu_j}{\sum_i a^\star_i \bar{\gamma}_{ij}}, \; a^\star_i = \frac{\mu_i}{\sum_j b^\star_j \bar{\gamma}_{ij}}$$

The Sinkhorn algorithm (aka IPFP)

$$b^{n+1}_j = \frac{\nu_j}{\sum_i a^n_i \bar{\gamma}_{ij}}, \; a^{n+1}_i = \frac{\mu_i}{\sum_j b^{n+1}_j \bar{\gamma}_{ij}}$$

Theorem ((ibid.))

a^n and b^n converge to $a^\star$ and $b^\star$

Remark: $\phi_i = \epsilon \log(a_i)$ and $\psi_j = \epsilon \log(b_j)$ are the (regularized) Kantorovich potentials.

The Sinkhorn algorithm

Theorem ((Franklin and Lorenz 1989))

The optimal plan $\gamma^\star$ has the form $\gamma^\star_{ij} = a^\star_i b^\star_j \bar{\gamma}_{ij}$. Moreover $a^\star_i$ and $b^\star_j$ can be uniquely determined (up to a multiplicative constant) as follows

$$b^\star_j = \frac{\nu_j}{\sum_i a^\star_i \bar{\gamma}_{ij}}, \; a^\star_i = \frac{\mu_i}{\sum_j b^\star_j \bar{\gamma}_{ij}}$$

The Sinkhorn algorithm (aka IPFP)

$$b^{n+1}_j = \frac{\nu_j}{\sum_i a^n_i \bar{\gamma}_{ij}}, \; a^{n+1}_i = \frac{\mu_i}{\sum_j b^{n+1}_j \bar{\gamma}_{ij}}$$

Theorem ((ibid.))

a^n and b^n converge to $a^\star$ and $b^\star$

Remark: $\phi_i = \epsilon \log(a_i)$ and $\psi_j = \epsilon \log(b_j)$ are the (regularized) Kantorovich potentials.

The Sinkhorn algorithm

Theorem ((Franklin and Lorenz 1989))

The optimal plan $\gamma^\star$ has the form $\gamma^\star_{ij} = a^\star_i b^\star_j \bar{\gamma}_{ij}$. Moreover $a^\star_i$ and $b^\star_j$ can be uniquely determined (up to a multiplicative constant) as follows

$$b^\star_j = \frac{\nu_j}{\sum_i a^\star_i \bar{\gamma}_{ij}},\ a^\star_i = \frac{\mu_i}{\sum_j b^\star_j \bar{\gamma}_{ij}}$$

The Sinkhorn algorithm (aka IPFP)

$$b^{n+1}_j = \frac{\nu_j}{\sum_i a^n_i \bar{\gamma}_{ij}},\ a^{n+1}_i = \frac{\mu_i}{\sum_j b^{n+1}_j \bar{\gamma}_{ij}}$$

Theorem ((ibid.))

a^n and b^n converge to $a^\star$ and $b^\star$

Remark:$\phi_i = \epsilon \log(a_i)$ and $\psi_j = \epsilon \log(b_j)$ are the (regularized) Kantorovich potentials.

Some Remarks

- In **(Franklin and Lorenz 1989)** proved the convergence of Sinkhorn by using the Hilbert metric.
- The entropic regularization spreads the support and this helps to stabilize: it defines a strongly convex program with a unique solution.
- The solution can be obtained through elementary operations (trivially parallelizable).
- The regularized solution γ^ϵ converges to the solution γ^{ot} of $\mathcal{MK}$ pb. with minimal entropy as $\epsilon \to 0$ (in **(Cominetti and San Martin 1994)** the authors proved that the convergence is exponential).
- The complexity depends on the cost function: with Euler's cost $\mathcal{O}((N-1)M^{2.37})$...still exponential in N for the Coulomb cost **:(** .

Some Remarks

- In **(Franklin and Lorenz 1989)** proved the convergence of Sinkhorn by using the Hilbert metric.
- The entropic regularization spreads the support and this helps to stabilize: it defines a strongly convex program with a unique solution.
- The solution can be obtained through elementary operations (trivially parallelizable).
- The regularized solution γ^{ϵ} converges to the solution γ^{ot} of $\mathcal{MK}$ pb. with minimal entropy as $\epsilon \to 0$ (in **(Cominetti and San Martin 1994)** the authors proved that the convergence is exponential).
- The complexity depends on the cost function: with Euler's cost $\mathcal{O}((N-1)M^{2.37})$...still exponential in N for the Coulomb cost **:(** .

Some Remarks

- In **(Franklin and Lorenz 1989)** proved the convergence of Sinkhorn by using the Hilbert metric.
- The entropic regularization spreads the support and this helps to stabilize: it defines a strongly convex program with a unique solution.
- The solution can be obtained through elementary operations (trivially parallelizable).
- The regularized solution γ^{ϵ} converges to the solution γ^{ot} of $\mathcal{MK}$ pb. with minimal entropy as $\epsilon \to 0$ (in **(Cominetti and San Martin 1994)** the authors proved that the convergence is exponential).
- The complexity depends on the cost function: with Euler's cost $\mathcal{O}((N-1)M^{2.37})$...still exponential in N for the Coulomb cost **:(** .

Some Remarks

- In **(Franklin and Lorenz 1989)** proved the convergence of Sinkhorn by using the Hilbert metric.
- The entropic regularization spreads the support and this helps to stabilize: it defines a strongly convex program with a unique solution.
- The solution can be obtained through elementary operations (trivially parallelizable).
- The regularized solution γ^ϵ converges to the solution γ^{ot} of $\mathcal{MK}$ pb. with minimal entropy as $\epsilon \to 0$ (in **(Cominetti and San Martin 1994)** the authors proved that the convergence is exponential).
- The complexity depends on the cost function: with Euler's cost $\mathcal{O}((N-1)M^{2.37})$...still exponential in N for the Coulomb cost **:(** .

Some Remarks

- In **(Franklin and Lorenz 1989)** proved the convergence of Sinkhorn by using the Hilbert metric.
- The entropic regularization spreads the support and this helps to stabilize: it defines a strongly convex program with a unique solution.
- The solution can be obtained through elementary operations (trivially parallelizable).
- The regularized solution γ^{ϵ} converges to the solution γ^{ot} of $\mathcal{MK}$ pb. with minimal entropy as $\epsilon \to 0$ (in **(Cominetti and San Martin 1994)** the authors proved that the convergence is exponential).
- The complexity depends on the cost function: with Euler's cost $\mathcal{O}((N-1)M^{2.37})$...still exponential in N for the Coulomb cost **:(** .

How the regularization works: from spread to deterministic plan (quadratic cost)

Take the quadratic cost and solve the regularized problem. Then as $\epsilon \to 0$ ($N = 512$), we have

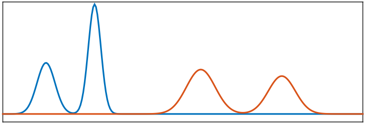


Figure: Marginals μ and ν

Figure: $\epsilon = 60/N$

How the regularization works: from spread to deterministic plan (quadratic cost)

Take the quadratic cost and solve the regularized problem. Then as $\epsilon \to 0$ ($N = 512$), we have

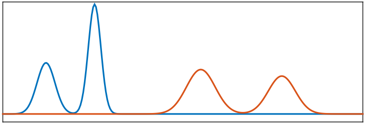


Figure: Marginals μ and ν

Figure: $\epsilon = 40/N$

How the regularization works: from spread to deterministic plan (quadratic cost)

Take the quadratic cost and solve the regularized problem. Then as $\epsilon \to 0$ ($N = 512$), we have

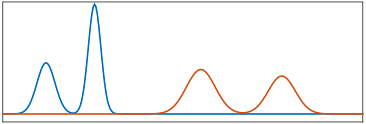


Figure: Marginals μ and ν

Figure: $\epsilon = 20/N$

How the regularization works: from spread to deterministic plan (quadratic cost)

Take the quadratic cost and solve the regularized problem. Then as $\epsilon \to 0$ ($N = 512$), we have

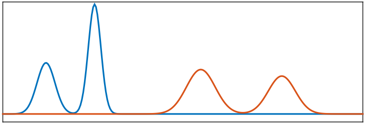


Figure: Marginals μ and ν

Figure: $\epsilon = 10/N$

How the regularization works: from spread to deterministic plan (quadratic cost)

Take the quadratic cost and solve the regularized problem. Then as $\epsilon \to 0$ ($N = 512$), we have

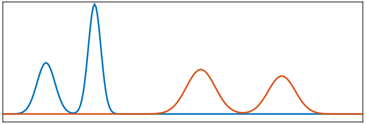


Figure: Marginals μ and ν

Figure: $\epsilon = 6/N$

How the regularization works: from spread to deterministic plan (quadratic cost)

Take the quadratic cost and solve the regularized problem. Then as $\epsilon \to 0$ ($N = 512$), we have

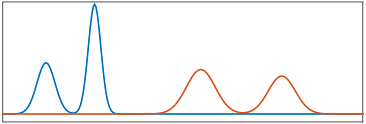


Figure: Marginals μ and ν

Figure: $\epsilon = 4/N$

The extension to the Multi-Marginal problem

The entropic multi-marginal problem becomes

$$\min_{\gamma\in\mathcal{C}} \mathcal{H}(\gamma|\bar{\gamma}) \tag{10}$$

where $\mathcal{H}(\gamma|\bar{\gamma}) = \sum_{i,j,k}\gamma_{ijk}(\log\frac{\gamma_{ijk}}{\bar{\gamma}_{ijk}} - 1)$ is the relative entropy, and $\mathcal{C} = \bigcap_{i=1}^{3}\mathcal{C}_i$ (i.e. $\mathcal{C}_1 = \{\gamma \mid \sum_{j,k}\gamma_{ijk} = \mu^1_i\}$).

The optimal plan $\gamma^\star$ becomes $\gamma^\star_{ijk} = a^\star_i b^\star_j c^\star_k \bar{\gamma}_{ijk}$ $a^\star_i$, $b^\star_j$ and $c^\star_k$ can be determined by the marginal constraints.

$$b^\star_j = \frac{\mu^2_j}{\sum_{ik} a^\star_i c^\star_k \bar{\gamma}_{ijk}} \quad\Rightarrow\quad b^{n+1}_j = \frac{\mu^2_j}{\sum_{ik} a^n_i c^n_k \bar{\gamma}_{ijk}}$$

$$c^\star_k = \frac{\mu^3_k}{\sum_{ij} a^\star_i b^\star_j \bar{\gamma}_{ijk}} \quad\Rightarrow\quad c^{n+1}_k = \frac{\mu^3_k}{\sum_{ij} a^n_i b^{n+1}_j \bar{\gamma}_{ijk}}$$

$$a^\star_i = \frac{\mu^1_i}{\sum_{jk} b^\star_j c^\star_k \bar{\gamma}_{ijk}} \quad\Rightarrow\quad a^{n+1}_i = \frac{\mu^1_i}{\sum_{jk} b^{n+1}_j c^{n+1}_k \bar{\gamma}_{ijk}}$$

The extension to the Multi-Marginal problem

The entropic multi-marginal problem becomes

$$\min_{\gamma\in\mathcal{C}} \mathcal{H}(\gamma|\bar\gamma) \tag{10}$$

where $\mathcal{H}(\gamma|\bar\gamma) = \sum_{i,j,k} \gamma_{ijk}(\log \frac{\gamma_{ijk}}{\bar\gamma_{ijk}} - 1)$ is the relative entropy, and $\mathcal{C} = \bigcap_{i=1}^{3} \mathcal{C}_i$ (i.e. $\mathcal{C}_1 = \{\gamma \;\;|\;\; \sum_{j,k}\gamma_{ijk} = \mu_i^1\}$).

The optimal plan $\gamma^\star$ becomes $\gamma^\star_{ijk} = a_i^\star b_j^\star c_k^\star \bar\gamma_{ijk}$ $a_i^\star$, $b_j^\star$ and $c_k^\star$ can be determined by the marginal constraints.

$$b_j^\star = \frac{\mu_j^2}{\sum_{ik} a_i^\star c_k^\star \bar\gamma_{ijk}} \qquad \Rightarrow \qquad b_j^{n+1} = \frac{\mu_j^2}{\sum_{ik} a_i^n c_k^n \bar\gamma_{ijk}}$$

$$c_k^\star = \frac{\mu_k^3}{\sum_{ij} a_i^\star b_j^\star \bar\gamma_{ijk}} \qquad \Rightarrow \qquad c_k^{n+1} = \frac{\mu_k^3}{\sum_{ij} a_i^n b_j^{n+1} \bar\gamma_{ijk}}$$

$$a_i^\star = \frac{\mu_i^1}{\sum_{jk} b_j^\star c_k^\star \bar\gamma_{ijk}} \qquad \Rightarrow \qquad a_i^{n+1} = \frac{\mu_i^1}{\sum_{jk} b_j^{n+1} c_k^{n+1} \bar\gamma_{ijk}}$$

The extension to the Multi-Marginal problem

The entropic multi-marginal problem becomes

$$\min_{\gamma\in\mathcal{C}} \mathcal{H}(\gamma|\bar\gamma) \tag{10}$$

where $\mathcal{H}(\gamma|\bar\gamma) = \sum_{i,j,k}\gamma_{ijk}(\log\frac{\gamma_{ijk}}{\bar\gamma_{ijk}} - 1)$ is the relative entropy, and $\mathcal{C} = \bigcap_{i=1}^3 \mathcal{C}_i$ (i.e. $\mathcal{C}_1 = \{\gamma \quad | \quad \sum_{j,k}\gamma_{ijk} = \mu_i^1\}$).

The optimal plan $\gamma^\star$ becomes $\gamma^\star_{ijk} = a_i^\star b_j^\star c_k^\star \bar\gamma_{ijk}$ $a_i^\star$, $b_j^\star$ and $c_k^\star$ can be determined by the marginal constraints.

$$b_j^\star = \frac{\mu_j^2}{\sum_{ik} a_i^\star c_k^\star \bar\gamma_{ijk}} \quad\Rightarrow\quad b_j^{n+1} = \frac{\mu_j^2}{\sum_{ik} a_i^n c_k^n \bar\gamma_{ijk}}$$

$$c_k^\star = \frac{\mu_k^3}{\sum_{ij} a_i^\star b_j^\star \bar\gamma_{ijk}} \quad\Rightarrow\quad c_k^{n+1} = \frac{\mu_k^3}{\sum_{ij} a_i^n b_j^{n+1} \bar\gamma_{ijk}}$$

$$a_i^\star = \frac{\mu_i^1}{\sum_{jk} b_j^\star c_k^\star \bar\gamma_{ijk}} \quad\Rightarrow\quad a_i^{n+1} = \frac{\mu_i^1}{\sum_{jk} b_j^{n+1} c_k^{n+1} \bar\gamma_{ijk}}$$

Sinkhornizing the world!!

- Wasserstein Barycenter **(Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Peyré 2015)**;
- Matching for teams **(Luca Nenna 2016)**;
- Optimal transport with capacity constraint **(Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Peyré 2015)**;
- Partial Optimal Transport **(Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Peyré 2015; Chizat, Peyré, B. Schmitzer, and Vialard 2016)**;
- Multi-Marginal Optimal Transport **(Luca Nenna 2016; J.-D. Benamou, G. Carlier, and L. Nenna 2016; Jean-David Benamou, Guillaume Carlier, and Luca Nenna 2018; Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Peyré 2015)**;
- Wasserstein Gradient Flows (JKO) **(Peyré 2015)**;
- Unbalanced Optimal Transport **(Chizat, Peyré, B. Schmitzer, and Vialard 2016)**;
- Cournot-Nash equilibria **(Blanchet, Guillaume Carlier, and Luca Nenna 2017)**
- Mean Field Games **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)**;
- And more is coming...

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

$$\inf \left\{ \int_{X\times X} \frac{1}{2}|x-y|^2 d\gamma \mid \gamma \in \Pi(\mu,\nu) \right\}$$

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

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

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

$$\inf \int_0^1 \int_X \frac{1}{2}|v_t|^2 \rho_t dx dt \quad s.t.\ \partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$$

$$\rho(0,\cdot) = \mu,\ \rho(1,\cdot) = \nu$$

Some remarks and a MMOT formulation

Remarks:

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

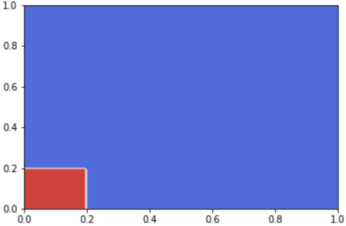
$$\pi_t(x,y)_\sharp \gamma = (e_t)_\sharp Q = \rho_t^\star,$$

  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

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

The geodesic in 2D



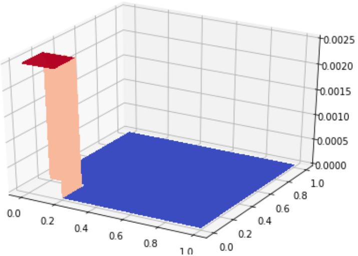
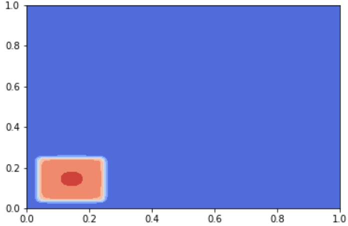


Figure: $t = 0$

The geodesic in 2D



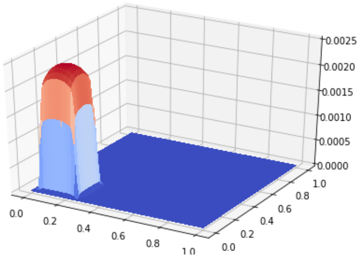
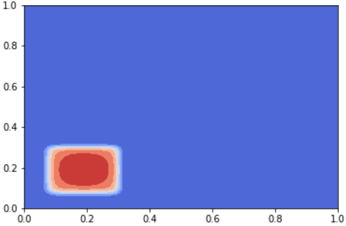


Figure: $t = \dfrac{1}{14}$

The geodesic in 2D



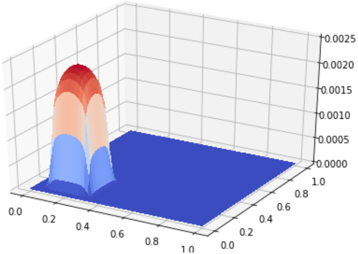
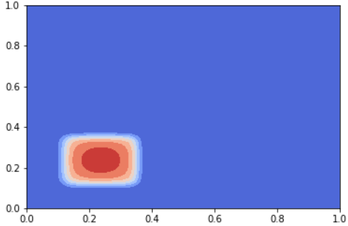


Figure: $t = \frac{2}{14}$

The geodesic in 2D



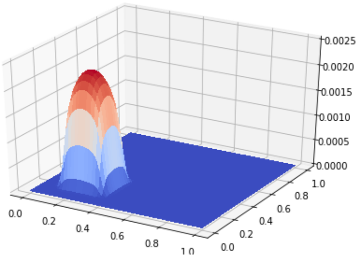
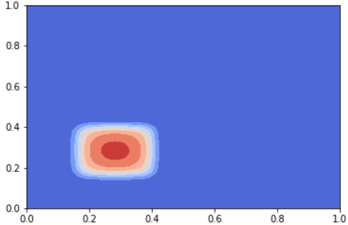


Figure: $t = \frac{3}{14}$

The geodesic in 2D



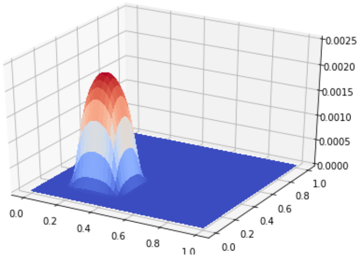


Figure: $t = \frac{4}{14}$

The geodesic in 2D

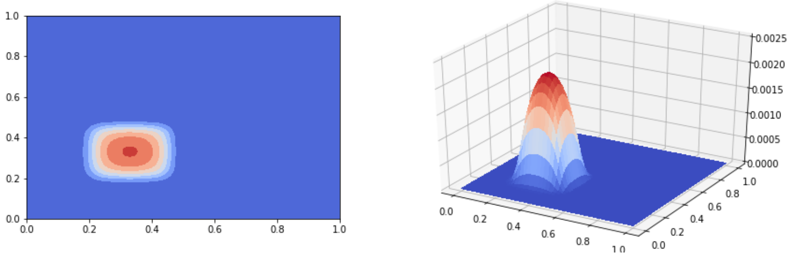


Figure: $t = \frac{5}{14}$

The geodesic in 2D

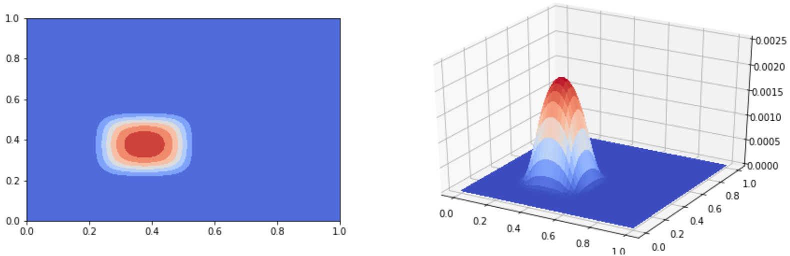


Figure: $t = \frac{6}{14}$

The geodesic in 2D

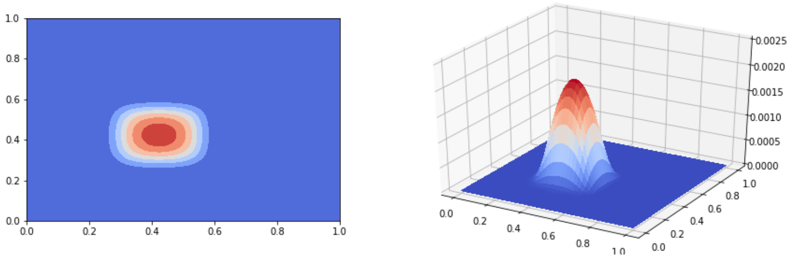


Figure: $t = \frac{7}{14}$

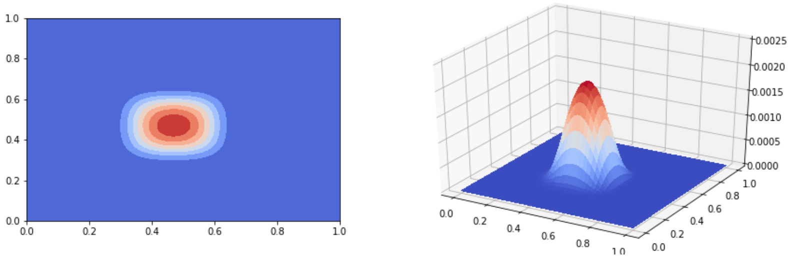


Figure: $t = \frac{8}{14}$

The geodesic in 2D

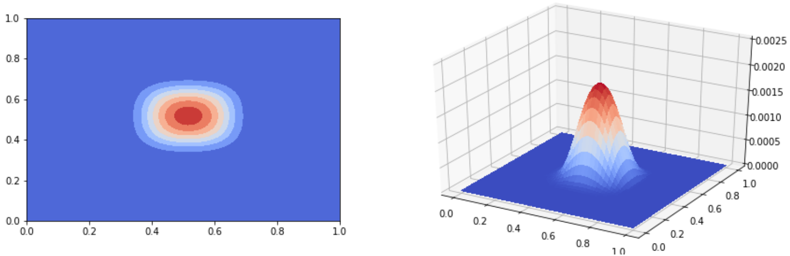


Figure: $t = \frac{9}{14}$

The geodesic in 2D

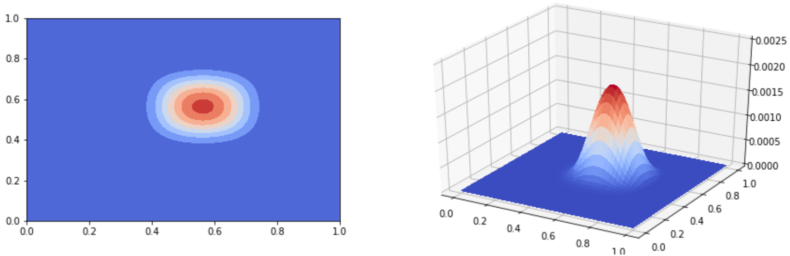


Figure: $t = \dfrac{10}{14}$

The geodesic in 2D

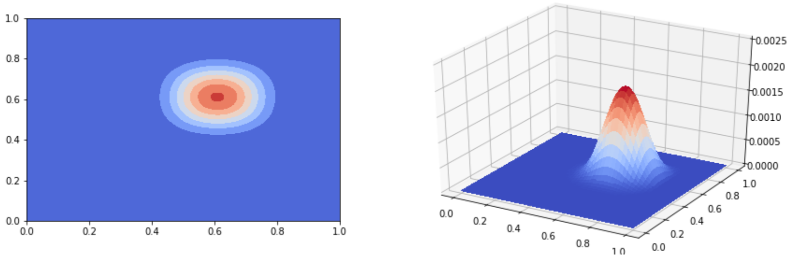


Figure: $t = \frac{11}{14}$

The geodesic in 2D

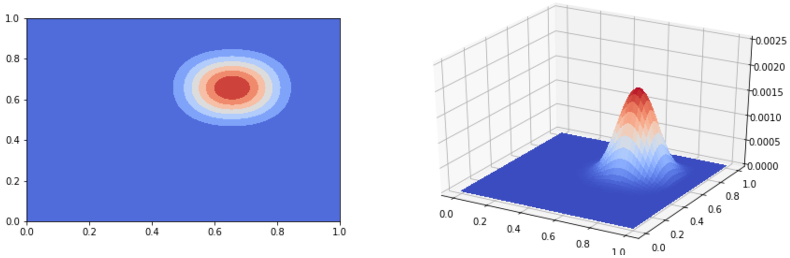
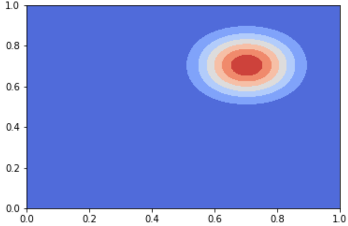


Figure: $t = \frac{12}{14}$

The geodesic in 2D



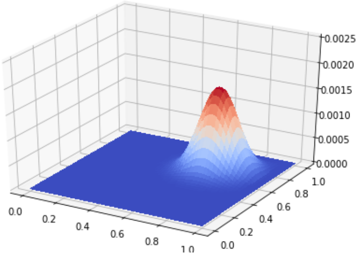


Figure: $t = \frac{13}{14}$

The geodesic in 2D

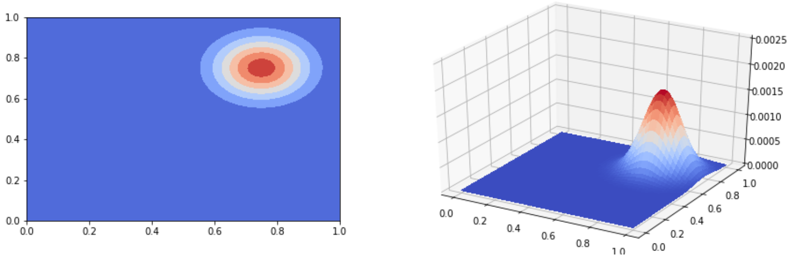


Figure: $t = 1$

The geodesic between images

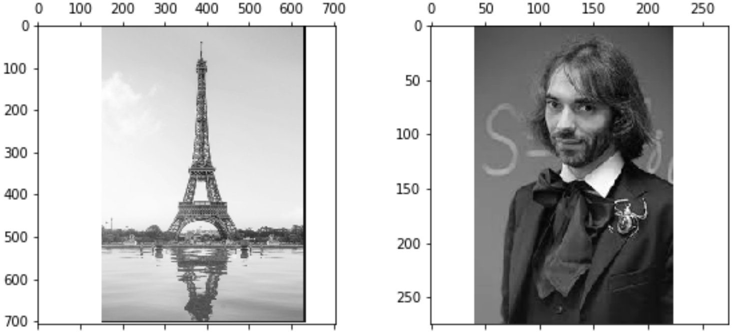


Figure: $t = 0$

The geodesic between images

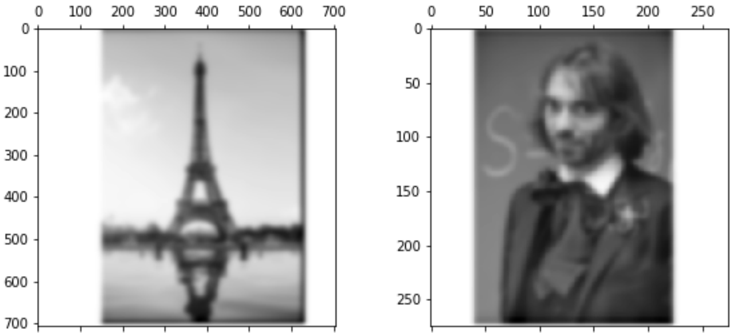


Figure: $t = \frac{1}{14}$

The geodesic between images

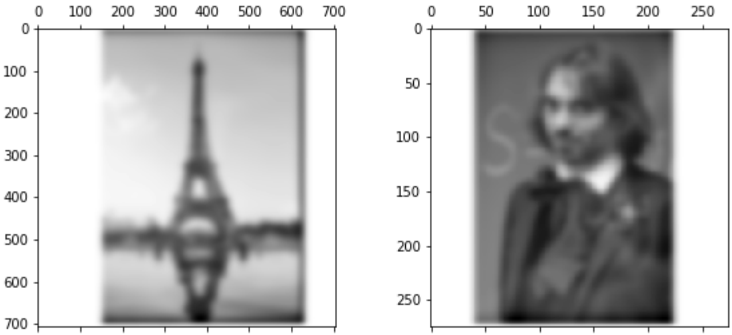


Figure: $t = \frac{2}{14}$

The geodesic between images

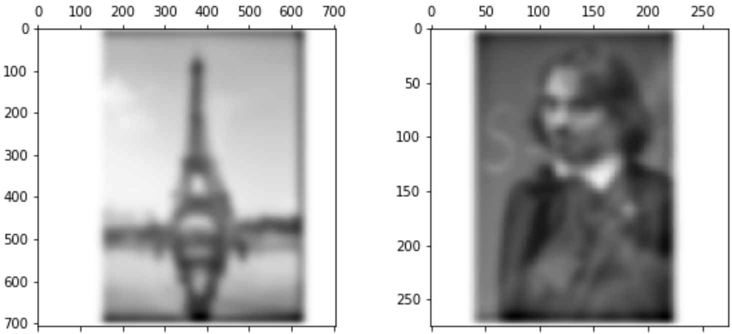


Figure: $t = \frac{3}{14}$

The geodesic between images

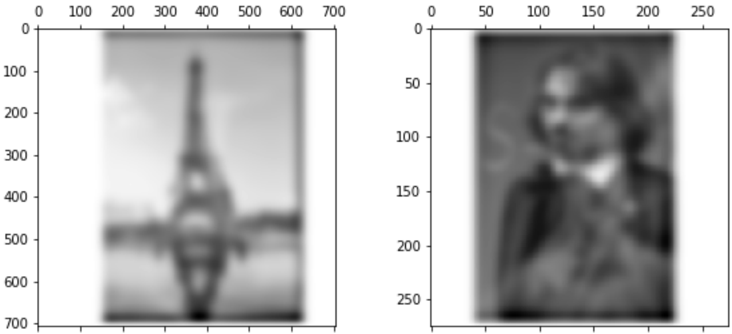


Figure: $t = \frac{4}{14}$

The geodesic between images

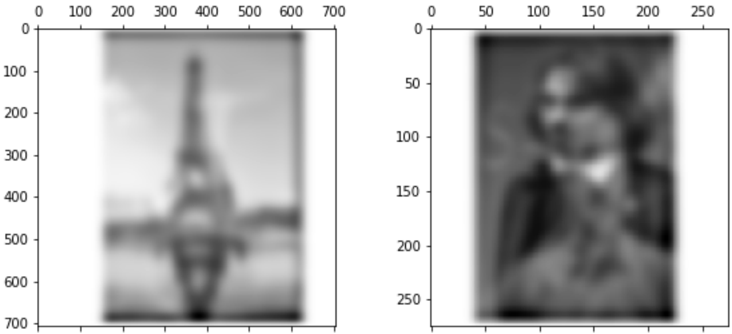
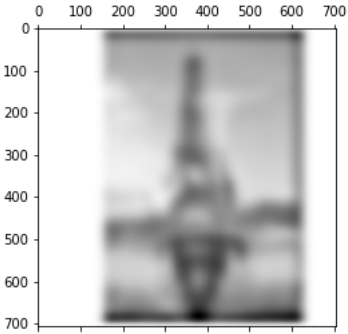


Figure: $t = \dfrac{5}{14}$

The geodesic between images



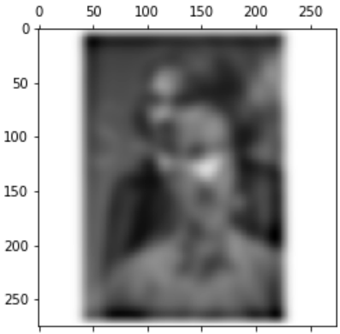


Figure: $t = \frac{6}{14}$

The geodesic between images

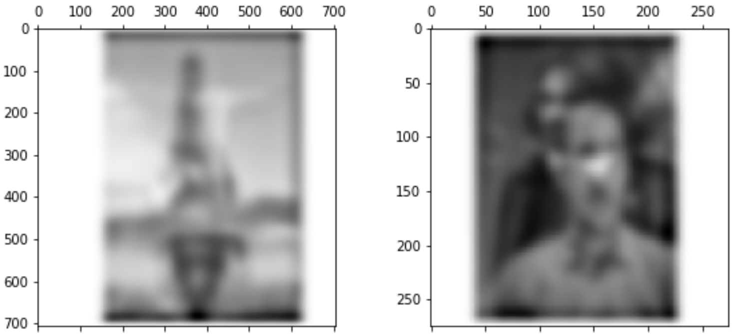


Figure: $t = \dfrac{7}{14}$

The geodesic between images

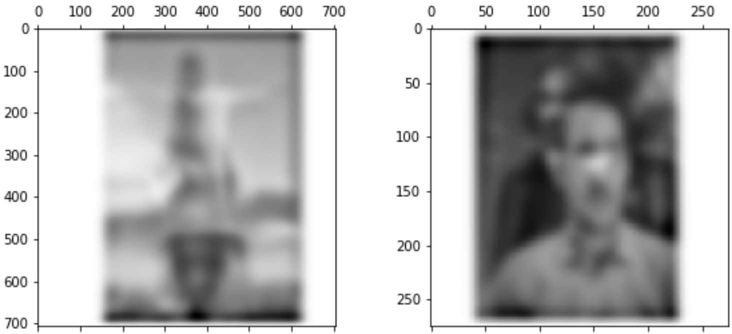


Figure: $t = \frac{8}{14}$

The geodesic between images

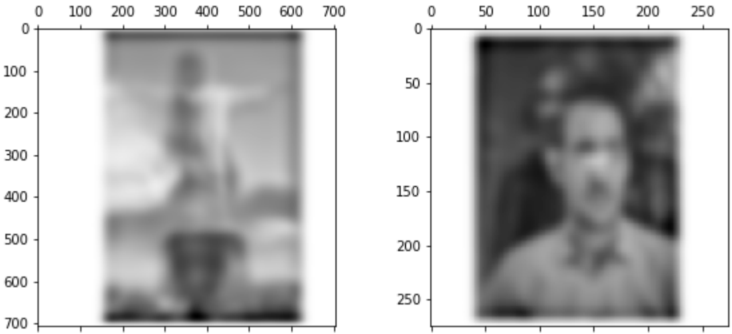


Figure: $t = \frac{9}{14}$

The geodesic between images

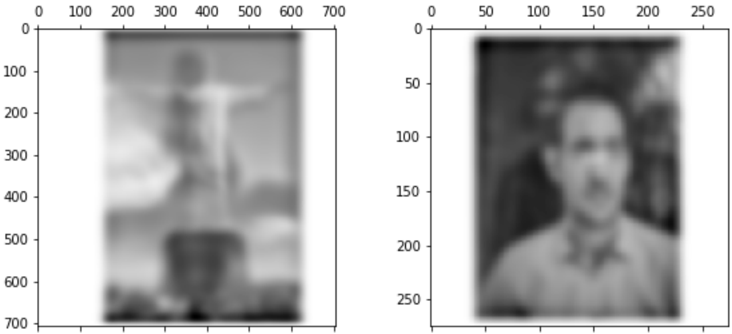


Figure: $t = \frac{10}{14}$

The geodesic between images

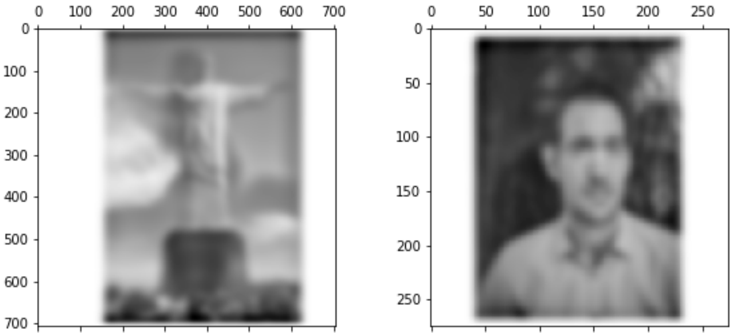


Figure: $t = \frac{11}{14}$

The geodesic between images

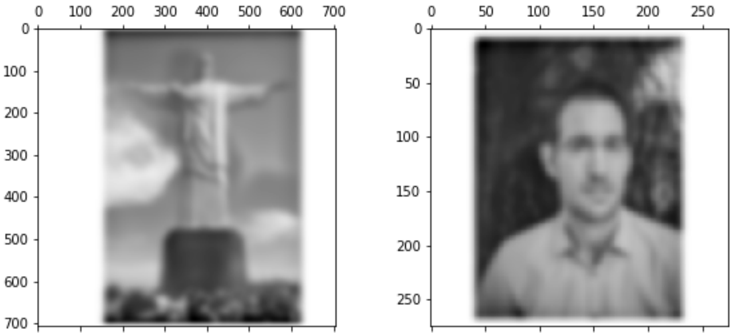


Figure: $t = \frac{12}{14}$

The geodesic between images

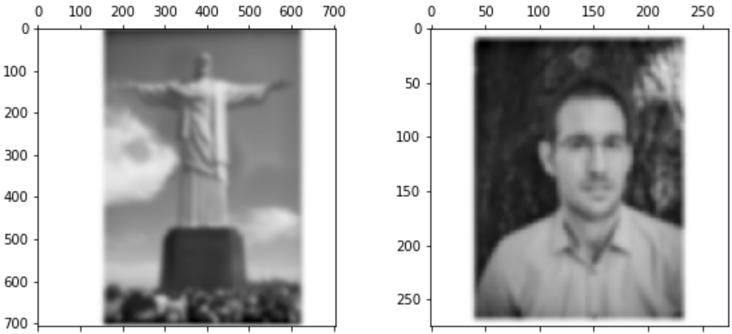


Figure: $t = \frac{13}{14}$

The geodesic between images

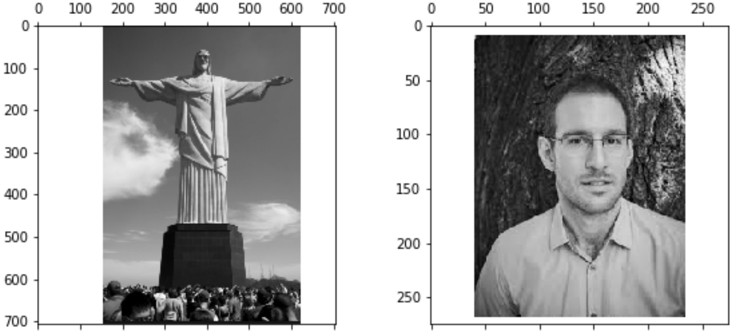


Figure: $t = 1$

Application II: MMOT and variational Mean Field Games

Lagrangian formulation for 1st order MFG

Consider a first order MFG system then we have the following "equivalence" (see **(Lasry and Lions 2007)**)

MFG system	(Eulerian) Variational Formulation
$\partial_t \rho - \mathrm{div}(\rho \nabla \phi) = 0$	$\inf \int_0^1 \int_\Omega (\frac{1}{2}\|v_t\|^2 \rho_t dxdt + G(x, \rho_t)) + F(\rho_1)$
$\rho(0, \cdot) = \rho_0$	$s.t.\ \partial_t \rho_t + \mathrm{div}(\rho_t v_t) = 0,\ \rho(0, \cdot) = \rho_0,$ $\Updownarrow$ (Lagrangian) Formulation **(J.-D. Benamou, G. Carlier, and Santambrogio 2017)**
$-\partial_t \phi + \frac{1}{2}\|\nabla \phi\|^2 = g(x, \rho)$ $\phi(1, \cdot) = \Psi$	$\min \int_C K(\omega) dQ(\omega) + \int_0^1 \int_\Omega G(x, e_{t,\sharp} Q) dxdt + F(e_{1,\sharp} Q)$ $s.t.\ e_{0,\sharp} Q = \rho_0.$

where G is the anti-derivative of g w.r.t its second variable, $C = H^1([0, 1]; \Omega)$, $F(\rho_1) = \int_\Omega \Psi d\rho_1$ is a final cost and $K(\omega) \stackrel{def}{=} \frac{1}{2} \int_0^1 |\dot{\omega}|^2 dt$

A Lagrangian formulation via Entropy minimization

What about a second order MFG system?

MFG system	(Eulerian) Variational Formulation
$\partial_t \rho - \mathrm{div}(\rho \nabla \phi) - \frac{\epsilon}{2}\Delta\rho = 0$	**(Cardaliaguet, Graber, Porretta, and Tonon 2015)**
$\rho(0,\cdot) = \rho_0$	$\inf \int_0^1 \int_\Omega (\frac{1}{2}\|v_t\|^2 \rho_t dxdt + G(x,\rho_t)) + F(\rho_1)$ $s.t.\ \partial_t \rho_t + \mathrm{div}(\rho_t v_t) - \frac{\epsilon}{2}\Delta\rho = 0,\ \rho(0,\cdot) = \rho_0,$ $\Updownarrow$
$-\partial_t \phi + \frac{1}{2}\|\nabla\phi\|^2 - \frac{\epsilon}{2}\Delta\phi = g(x,\rho)$ $\phi(1,\cdot) = \Psi$	(Lagrangian) Formulation **(J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna 2018)** $\min \boldsymbol{\mathcal{H}(Q\|R^\epsilon)} + \int_0^1 \int_\Omega G(x, e_{t,\sharp}Q)dxdt + F(e_{1,\sharp}Q),$ $s.t.\ e_{0,\sharp}Q = \rho_0.$

where $\mathcal{H}(q|r) = \int \log(\frac{dq}{dr})dq$ is the relative entropy ($q \ll r$) and R^ϵ is the Wiener measure $R^\epsilon := \int \delta_{x+B^\epsilon} dx$ of variance ϵ.

The discretised (in time) problems

The discrete Lagrangian formulations read

- 1st order MFG

$$\inf\left\{\int_{\Omega^{T+1}} K_T dQ_T(x_0,\cdots,x_T)+\sum_{i=1}^{T-1}\int_\Omega G(x,\pi_{i,\sharp}Q_T)dx_i+F(\pi_{T,\sharp}Q_T)\mid \pi_{0,\sharp}Q_T=\rho_0\right\},$$

where $K_T=\frac{1}{2T}\sum_{i=0}^{T-1}|x_{i+1}-x_i|^2$, $Q_T\in\mathcal{P}(\Omega^{T+1})$.

- 2nd order MFG

$$\inf\left\{\mathcal{H}(Q_T|R_T^\epsilon)+\sum_{i=1}^{T-1}\int_\Omega G(x,\pi_{i,\sharp}Q_T)dx_i+F(\pi_{N,\sharp}Q_T)\mid \pi_{0,\sharp}Q_T=\rho_0\right\},$$

where $R_T^\epsilon\stackrel{def}{=}\prod_{n=0}^{T}\xi_{n,n+1}$ and $\xi_{ij}=\exp^{-\frac{|x_i-x_j|^2}{2T\epsilon}}$.

IDEA: a generalised Sinkhorn to compute the solution of both problems

The discretised (in time) problems

The discrete Lagrangian formulations read

- 1st order MFG

$$\inf\left\{\int_{\Omega^{T+1}} K_T dQ_T(x_0,\cdots,x_T) + \sum_{i=1}^{T-1}\int_\Omega G(x,\pi_{i,\sharp}Q_T)dx_i + F(\pi_{T,\sharp}Q_T) \mid \pi_{0,\sharp}Q_T = \rho_0\right\},$$

where $K_T = \frac{1}{2T}\sum_{i=0}^{T-1}|x_{i+1}-x_i|^2$, $Q_T \in \mathcal{P}(\Omega^{T+1})$.

- 2nd order MFG

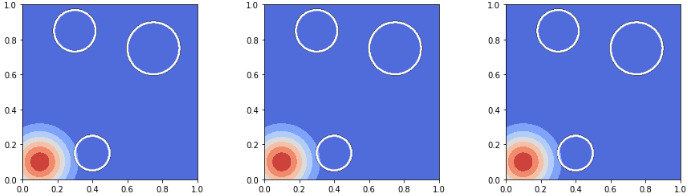
$$\inf\left\{\mathcal{H}(Q_T|R_T^\epsilon) + \sum_{i=1}^{T-1}\int_\Omega G(x,\pi_{i,\sharp}Q_T)dx_i + F(\pi_{N,\sharp}Q_T) \mid \pi_{0,\sharp}Q_T = \rho_0\right\},$$

where $R_T^\epsilon \overset{def}{=} \prod_{n=0}^{T}\xi_{n,n+1}$ and $\xi_{ij} = \exp^{-\dfrac{|x_i-x_j|^2}{2T\epsilon}}$.

IDEA: a generalised Sinkhorn to compute the solution of both problems

Hard congestion with obstacle, behaviour as $\epsilon \to 0$

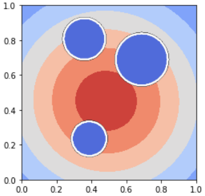
$T = 32$ time steps; grid: uniform discretization of $[0, 1]^2$ with $N \times N$ points
$N = 250$



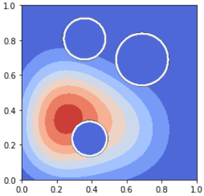
$\epsilon = 1$ $\qquad$ $\epsilon = 0.01$ $\qquad$ $\epsilon = 0.001$

Hard congestion with obstacle, behaviour as $\epsilon \to 0$

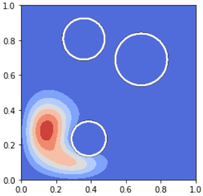
$T = 32$ time steps; grid: uniform discretization of $[0,1]^2$ with $N \times N$ points
$N = 250$



$\epsilon = 1$



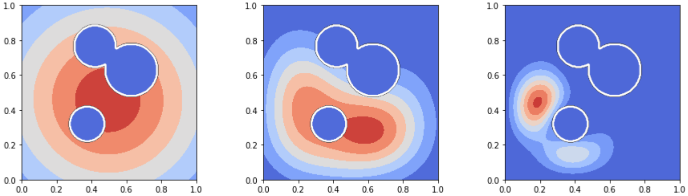
$\epsilon = 0.01$



$\epsilon = 0.001$

Hard congestion with obstacle, behaviour as $\epsilon \to 0$

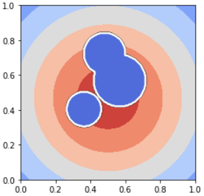
$T = 32$ time steps; grid: uniform discretization of $[0,1]^2$ with $N \times N$ points
$N = 250$



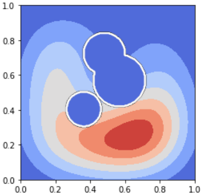
$\epsilon = 1$ $\epsilon = 0.01$ $\epsilon = 0.001$

Hard congestion with obstacle, behaviour as $\epsilon \to 0$

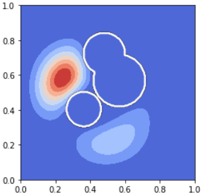
$T = 32$ time steps; grid: uniform discretization of $[0,1]^2$ with $N \times N$ points
$N = 250$



$\epsilon = 1$



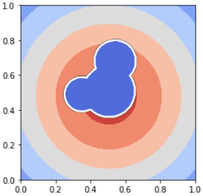
$\epsilon = 0.01$



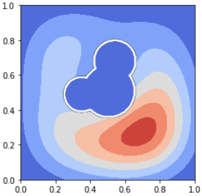
$\epsilon = 0.001$

Hard congestion with obstacle, behaviour as $\epsilon \to 0$

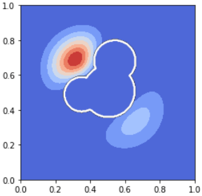
$T = 32$ time steps; grid: uniform discretization of $[0,1]^2$ with $N \times N$ points
$N = 250$



$\epsilon = 1$



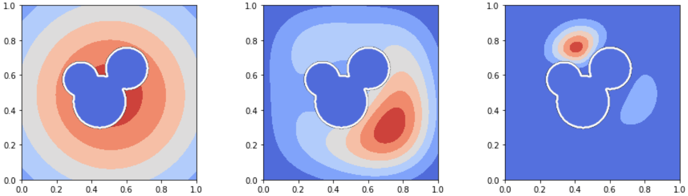
$\epsilon = 0.01$



$\epsilon = 0.001$

Hard congestion with obstacle, behaviour as $\epsilon \to 0$

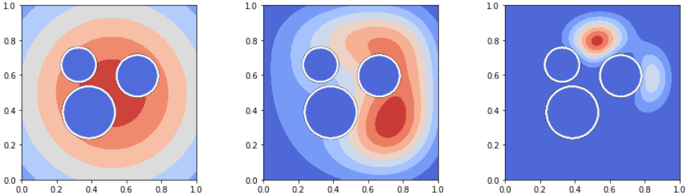
$T = 32$ time steps; grid: uniform discretization of $[0,1]^2$ with $N \times N$ points
$N = 250$



$\epsilon = 1$ $\epsilon = 0.01$ $\epsilon = 0.001$

Hard congestion with obstacle, behaviour as $\epsilon \to 0$

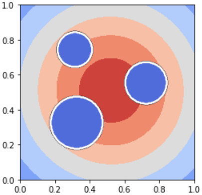
$T = 32$ time steps; grid: uniform discretization of $[0,1]^2$ with $N \times N$ points
$N = 250$



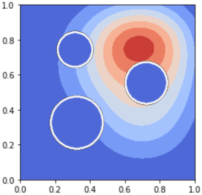
$\epsilon = 1$ $\quad$ $\epsilon = 0.01$ $\quad$ $\epsilon = 0.001$

Hard congestion with obstacle, behaviour as $\epsilon \to 0$

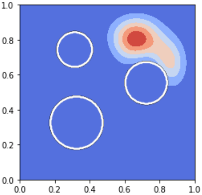
$T = 32$ time steps; grid: uniform discretization of $[0,1]^2$ with $N \times N$ points
$N = 250$



$\epsilon = 1$



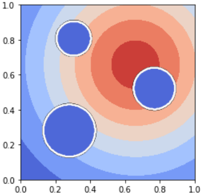
$\epsilon = 0.01$



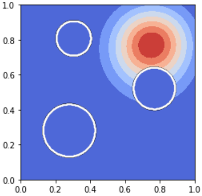
$\epsilon = 0.001$

Hard congestion with obstacle, behaviour as $\epsilon \to 0$

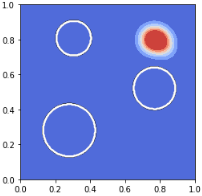
$T = 32$ time steps; grid: uniform discretization of $[0,1]^2$ with $N \times N$ points
$N = 250$



$\epsilon = 1$



$\epsilon = 0.01$



$\epsilon = 0.001$

References

Benamou, J.-D., G. Carlier, & L. Nenna (2016). "A Numerical Method to solve Multi-Marginal Optimal Transport Problems with Coulomb Cost". In: *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer International Publishing, pp. 577–601.

Benamou, Jean-David, Guillaume Carlier, Marco Cuturi, Luca Nenna, & Gabriel Peyré (2015). "Iterative Bregman projections for regularized transportation problems". In: *SIAM J. Sci. Comput.* 37.2, A1111–A1138. ISSN: 1064-8275. DOI: 10.1137/141000439. URL: http://dx.doi.org/10.1137/141000439.

Nenna, Luca (2016). "Numerical methods for multi-marginal optimal transportation". PhD thesis. PSL Research University.

Peyré, Gabriel & Marco Cuturi (2017). *Computational optimal transport*. Tech. rep.

Chizat, L., G. Peyré, B. Schmitzer, & F.-X. Vialard (2016). *Scaling Algorithms for Unbalanced Transport Problems*. Tech. rep. http://arxiv.org/abs/1607.05816.

Léonard, C. (2012). "From the Schrödinger problem to the Monge-Kantorovich problem". In: *Journal of Functional Analysis* 262.4, pp. 1879–1920.

Mérigot, Quentin (2011). "A multiscale approach to optimal transport". In: *Computer Graphics Forum*. Vol. 30. 5. Wiley Online Library, pp. 1583–1592.

Cuturi, M. (2013). "Sinkhorn Distances: Lightspeed Computation of Optimal Transport.". In: *Advances in Neural Information Processing Systems (NIPS) 26*, pp. 2292–2300.

Galichon, A. & B. Salanié (2009). *Matching with Trade-offs: Revealed Preferences over Competing Characteristics*. Tech. rep. Preprint SSRN-1487307.