Dynamic formulations of Optimal Transportation and variational relaxation of Euler equations

Jean-David Benamou
EPC MOKAPLAN

Barcelona FOCM July 2017

Inria informatics mathematics
DAUPHINE UNIVERSITÉ PARIS | CEREMADE UMR CNRS 7534
cnrs

Summary

Modern setting of Monge Problem (1781)

- Source/Target Data : $\boxed{d\rho_i(x)(= \rho_i(x)\, dx), \;\; i = 0, 1}$

$$\rho_i \geq 0, \int_D \rho_0(x)dx = \int_D \rho_1(x)dx = 1, \; D \subset \mathbb{R}^n$$

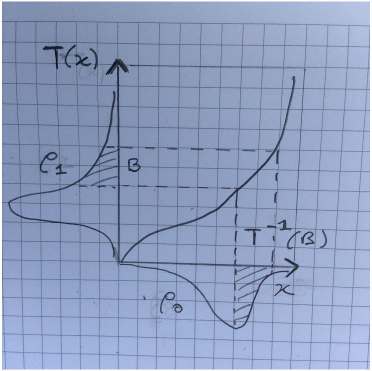
- Measure preserving Transport Maps :

$$\boxed{\mathcal{M} = \{T : D \to D, \; T_\# \rho_0 = \rho_1\}}$$

$$\forall B \subset D$$

$$T_\# \rho_0\,(B) = \rho_0(T^{-1}(B))$$

$$det(DT)\rho_1(T(x)) = \rho_0(x)$$



Modern setting of Monge Problem (1781)

- Cost Function : $\boxed{\mathcal{I}(T) = \int_D c(x, T(x))\, \rho_0(x)\, dx}$
- Monge Problem : $\boxed{\text{(MP)} \inf_{T \in \mathcal{M}} \mathcal{I}(T)}$
- Costs : typically $c(x, y) = \frac{1}{p} \|y - x\|^p$ (Monge $\rightarrow p = 1$).
- Th. Brenier (1991) $(p = 2)$: $\exists! \, \nabla\varphi$, φ convex such that $\mathcal{I}(\nabla\varphi(x)) = \min_{T \in \mathcal{M}} \mathcal{I}(T)$
- Measure preserving property yields : $\boxed{(MABV2) \quad det(D^2\varphi)\rho_1(\nabla\varphi) = \rho_0, \quad \nabla\varphi(X_0) \subset X_1}$
- Extensive Sobolev regularity theory develloped since by Cafarelli and Ambrosio schools ...

  O(N) Numerical methods : Monotone FD scheme B. Froese Oberman (2014) B. Collino Mirebeau (2016) and B. Duval (2017) and Semi-Discrete approaches Mérigot (2011) Lévy (2015).

Adding the dynamics

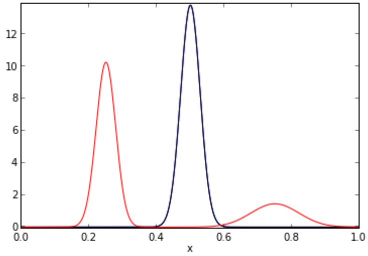
- Displacement Interpolation - McCann (1997). Def :

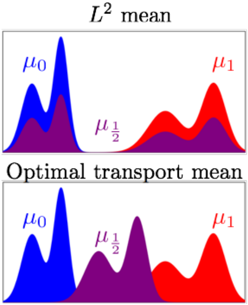
$$\boxed{x \mapsto \Theta(t,x) = x + t\,(\nabla\varphi(x) - x), \quad t \in]0,1[}$$

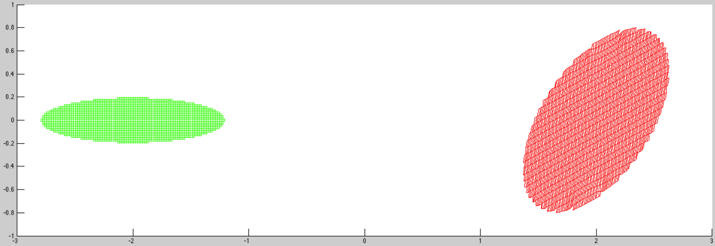
$$\boxed{\rho^*(t,.) = (\Theta(t,.))_{\#}\rho_0}$$

$$(t \mapsto \rho^*(t, \Theta(t,x)) = \frac{\rho_0(x)}{det(D_x\Theta(t,x))} \;)$$

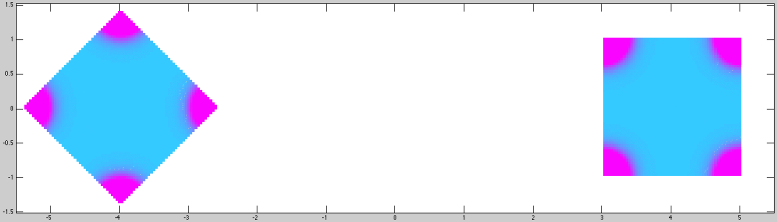
- for all t, $\Theta(t,.)$ solves (MP) from ρ_0 to $\rho^*(t,.)$.
- $W_2(\rho_0, \rho^*(t,.)) = \sqrt{\mathcal{I}(\Theta(t,.))}$ is a geodesic distance on $\mathcal{P}(D)$.







1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-3
-2
-1
0
1
2
3



1.5
1
0.5
0
-0.5
-1
-1.5
-5
-4
-3
-2
-1
0
1
2
3
4
5

The CFD Formulation

- Particles move in straigth line at constant speed

$$\dot{\Theta}(t,x_0) = (\nabla\varphi(x_0) - x_0) \stackrel{\text{def.}}{=} v^*(t,\Theta(t,x_0))$$

- B. Brenier (2000) : (ρ^*, v^*) is the unique minimum of

$$\inf_{(\rho,v)\,satisfies\,(CE)} \int_0^1 \int_D \frac{1}{2}\rho(t,x)\,\|v(t,x)\|^2\,dx\,dt$$

$$(CE) \quad \partial_t\rho + div(\rho\, v) = 0,\ \partial_\nu v = 0\ on\ \partial D,\ \rho(i,.) = \rho_i(.)$$

- This is a non-smooth convex relaxation (under $(\rho, v) \to (\rho, \sigma \stackrel{\text{def.}}{=} \rho\, v)$) proximal spliting methods achieve $O(N^3)$ heuristically.

- A variational deterministic MFG - Lions Lasry (2007) : (CE) and $v = \text{grad}\,\psi$ and $(HJ) \quad \partial_t\psi + \frac{1}{2}\|\nabla\psi\|^2 = 0$ see B. Carlier (2015).

Kantorovich Relaxation (1942)

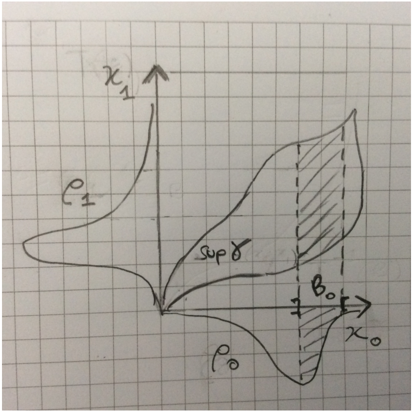
- Transport Plans :

$$\boxed{\Pi(\rho_0, \rho_1) = \{\gamma \in \mathcal{P}(D_0 \times D_1),\ (\Pi_{D_i})_{\#}\gamma = \rho_i,\ i = 0, 1\}}$$

$\forall B_0 \subset D_0$

$(\Pi_{D_0})_{\#}\gamma\,(B_0) = \gamma(B_0, D_1)$

$\gamma(B_0, D_1) = \rho_0(B_0)$



- $\Pi(\rho_0, \rho_1)$ is non empty : $\rho_0 \otimes \rho_1(x_0, x_1) = \rho_0(x_0)\,\rho_1(x_1)$

Kantorovich Relaxation (1942)

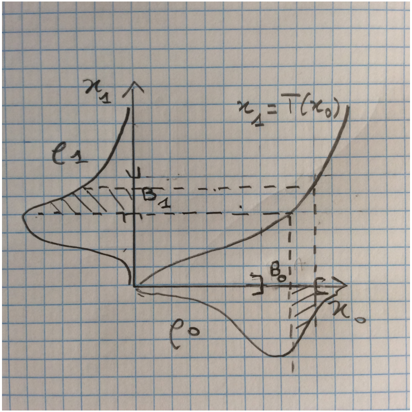
Deterministic Transport plan :

$$\boxed{\gamma_T \overset{\text{def.}}{=} (Id, T)_{\#}\rho_0}$$

$\gamma_T(B_0, B_1) =$

$\rho_0(\{x \in B_0,\ s.t.\ T(x) \in B_1\})$

$T \in \mathcal{M} \Leftrightarrow \gamma_T \in \Pi(\rho_0, \rho_1)$



- $\gamma_{\nabla\varphi}$ solves $\boxed{(MK) \qquad \inf_{\gamma \in \Pi(\rho_0, \rho_1)} \int_{D_0 \times D_1} c(x_0, x_1)\, d\gamma(x_0, x_1)}$
- Linear program but N^2 unknowns Simplex or Interior point methods stuck to $N \simeq 100$.

Dynamic Kantorovich relaxation

- Defs : $\Omega(D) = C([0,1]; D)$ the set of abs. cont. path $\omega : t \in [0,1] \mapsto \omega(t) \in D$.

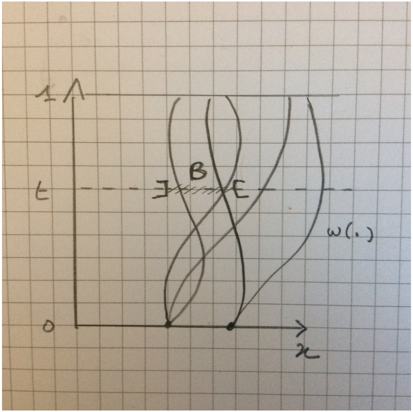
  $Q \in \mathcal{P}(\Omega(D))$ a probability measure on $\Omega(D)$.

  $e_t : \Omega(D) \mapsto D$ the t-evaluation function - $e_t(\omega) = \omega(t)$.

$\forall B \subset D$

$(e_t)_\# Q\,(B) =$

$Q(\{\omega \in \Omega(D),\ \omega(t) \in B\})$



Dynamic Kantorovich relaxation

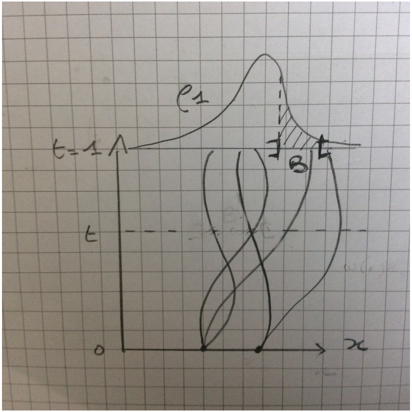
- Defs : $\Omega(D) = C([0,1]; D)$ the set of abs. cont. path $\omega : t \in [0,1] \mapsto \omega(t) \in D$.

  $Q \in \mathcal{P}(\Omega(D))$ a probability measure on $\Omega(D)$.

  $e_t : \Omega(D) \mapsto D$ the t-evaluation function - $e_t(\omega) = \omega(t)$.

$\forall B \subset D$

$(e_1)_{\#} Q\,(B) = \rho_1(B)$



Dynamic Kantorovich relaxation

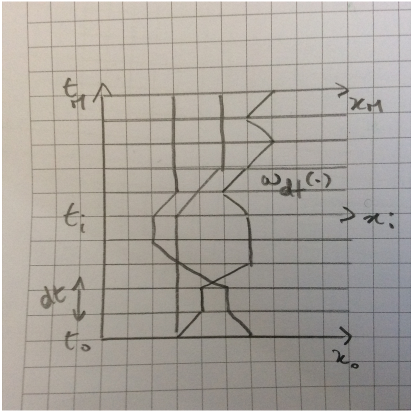
$$(DMK) \qquad \inf_{\{Q\in\mathcal{P}(\Omega(D)),\,(e_i)_\# Q=\rho_i,\,i=0,1\}} \int_{\Omega(D)} \int_0^1 \|\dot{\omega}(t)\|^2 \, dt \, dQ(\omega)$$

- $\Phi_\Theta : D \rightarrow \Omega(D)$, $\Phi_\Theta(x_0) = \Theta(., x_0)$.
- The solution $Q^* = (\Phi_\Theta)_\# \rho_0$ is deterministic.

  $\forall O \subset \Omega(D),\, Q^*(O) = \rho_0(\{x_0 \in D_0,\ s.t.\, \Theta(x_0, .) \in O\})$
- $\rho(t, .) = (e_t)_\# Q^*$ is the CFD geodesic.
  Analysis by Ambrosio school, see Santambrogio book (2015)
- $\mathcal{P}(\Omega(D))$ is a BIG space : next section present an efficient numerical method.

Time Discretization

- Discretize time : Set $dt = \frac{1}{M}$ $t_i = i\, dt,\ i = 0..M$
- Restrict to piecewise linear path $\omega_{dt} = \{x_0, x_1, .., x_M\}$ $(\omega_{dt}(t_i) = x_i)$.



Time Discretization

- Minimize w.r.t. $Q_{dt}(x_0, x_1, ..., x_M) \in \mathcal{P}(\otimes_{i=0,M} D_i)$
- $(e_{t_i})_\# Q_{dt} = \rho_i$ becomes a margin condition :

$$\int_{\otimes_{j \neq i} D_j} dQ_{dt}(x_0, x_1, .., x_M) = \rho_i(x_i)$$

- Time integration of linear path in (DMK) :

$$\inf_{Q \in \mathcal{E}} \int_{\otimes_{i=0,M} D_i} \left(\sum_{i=0,M-1} \frac{1}{dt} \|x_{i+1} - x_i\|^2 \right) dQ_{dt}(x_0, x_1, .., x_M)$$

$$\mathcal{E} = \{Q_{dt} \in \mathcal{P}(\otimes_{i=0,M} D_i),\ (e_{t_i})_\# Q_{dt} = \rho_i,\ i = 0, 1\}$$

Multi-Marginal OT

- General Form of MMOT :

$$\inf_{Q \in \mathcal{E}} \int_{\otimes_{i=0,M} D_i} c(x_0, x_1, .., x_M)\, dQ(x_0, x_1, .., x_M)$$

$$\mathcal{E} = \{Q \in \mathcal{P}(\otimes_{i=0,M} D_i),\ (e_{t_i})_\# Q = \rho_i,\ i = 0, 1, .., M\}$$

- Ex. : Density Functional Theory (Friesecke et al, Butazzo et al (...) , Pass, ...)

$$c \stackrel{\text{def.}}{=} \sum_{i<j} \frac{1}{||x_i - x_j||}$$

Margins : $(e_i)_\# Q = \bar{\rho},\ i = 0, .., M$

Existence of Maps open ...

- Generalized Euler Geodesics (Brenier) : last section.

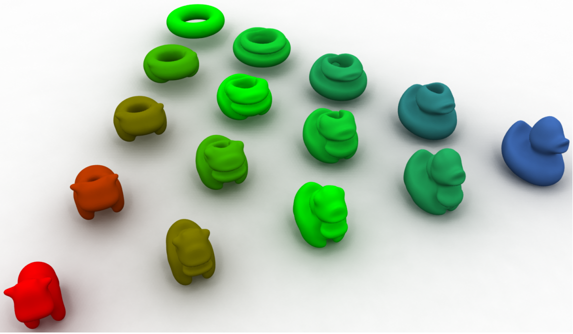
Multi-Marginal OT

- Ex. : Wassertein Barycenters (Agueh/Carlier (2011))

$$c \stackrel{\text{def.}}{=} \sum_i \lambda_i \|x_i - B(x_0, .., x_M)\|^2 \qquad B(x_0, .., x_M) \stackrel{\text{def.}}{=} \sum_i \lambda_i x_i$$

Margins : $(e_i)_\# Q = \rho_i,\ i = 0, .., M$ Barycenter : $B_\# Q$ ...

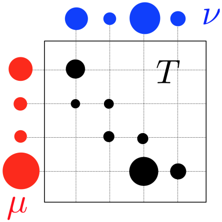
- Solomon et al (2015)



Summary

Entropic regularization of OT

See Christian Leonard surveys for the connection with the Schrödinger problem in the continuous setting.



Discretize in space

$D_0 : \{x_i\}$ and $D_1 : \{x_j\}$

$\rho_0 = \sum_i \mu_i \, \delta_{x_i}$ and $\rho_1 = \sum_j \nu_j \, \delta_{y_j}$

$c_{i\,j} = c(x_i, x_j)$

- **Entropic regularisation of MK :**

$$(MK_\varepsilon) \quad \min_{\gamma \in \mathcal{G}} \sum_{i\,j} \gamma^{\varepsilon}_{i\,j} \, c_{i\,j} + \varepsilon \, \gamma^{\varepsilon}_{i\,j} \, (\log \gamma^{\varepsilon}_{i\,j} - 1)$$

$$\mathcal{G} = \{\gamma \in \mathbb{R}^{N \times N}, \, \gamma^{\varepsilon}_{i\,j} \cancel{\geq 0}, \, \sum_j \gamma^{\varepsilon}_{i\,j} = \mu_i \;, \sum_i \gamma^{\varepsilon}_{i\,j} = \nu_j\}$$

- Set $\overline{\gamma}^{\varepsilon}_{i\,j} = e^{-\frac{c_{i\,j}}{\varepsilon}}$

$$(MK_\varepsilon) \qquad \min_{\gamma \in \mathcal{G}} \sum_{i\,j} KL(\gamma^{\varepsilon}_{i\,j} | \overline{\gamma}^{\varepsilon}_{i\,j})$$

$$KL(f|g) = f \, (\log(\frac{f}{g}) - 1)$$

Iterative Proportional Fitting Procedure

Sinkhorn (67) Ruschendorf (95) Galichon (09) Cuturi (13) ...

$$\min_{\gamma^\varepsilon_{i\,j}} \max_{\{\varphi^\varepsilon_i,\psi^\varepsilon_j\}} \sum_{ij} \psi^\varepsilon_j\,\nu_j + \varphi^\varepsilon_i\,\mu_i + \gamma^\varepsilon_{i\,j}\left(c_{i\,j} - \psi^\varepsilon_j - \varphi^\varepsilon_i + \varepsilon\left(\log\gamma^\varepsilon_{i\,j} - 1\right)\right)$$

- Optimal plan is a scaling : $\boxed{\gamma^{\star,\varepsilon}_{i\,j} = a^\varepsilon_i\; b^\varepsilon_j\; \overline{\gamma}^\varepsilon_{i\,j}}$

  where $a^\varepsilon_i = e^{\frac{\varphi^\varepsilon_i}{\varepsilon}}$ and $b^\varepsilon_j = e^{\frac{\psi^\varepsilon_j}{\varepsilon}}$.

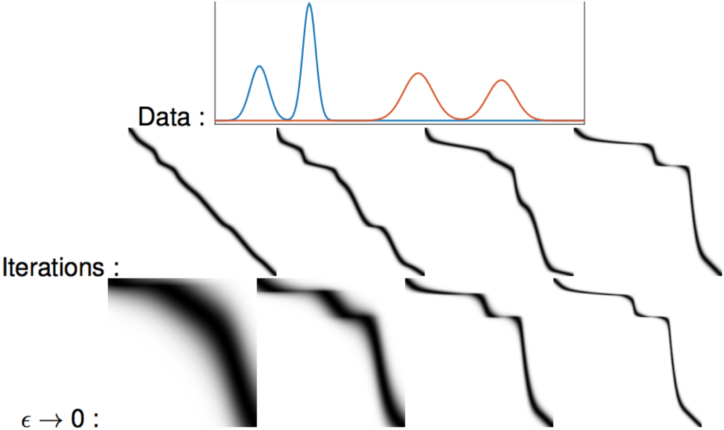
- Margin constraints give :

$$\boxed{a^\varepsilon_i = \frac{\mu_i}{\sum_j \overline{\gamma}^\varepsilon_{i\,j}\; b^\varepsilon_j} \text{ and } b^\varepsilon_j = \frac{\nu_j}{\sum_i \overline{\gamma}^\varepsilon_{i\,j}\; a^\varepsilon_i}.}$$

- IPFP is the relaxation :

$$\boxed{a_i^{\varepsilon,k+\frac{1}{2}} = \frac{\mu_i}{\sum_j \overline{\gamma}^\varepsilon_{i\,j}\; b_j^{\varepsilon,k}} \qquad b_j^{\varepsilon,k+1} = \frac{\nu_j}{\sum_i \overline{\gamma}^\varepsilon_{i\,j}\; a_i^{\varepsilon,k+\frac{1}{2}}}.}$$

1-D IPFP/Sinkhorn



Remarks on IPFP

- It. are contractions in the Hilbert metric $d_H(p,q) = \log(\frac{max_i(\frac{p_i}{q_i})}{min_i(\frac{p_i}{q_i})})$.
- Convergence with ε (Cominetti San Martin (94) , Carlier et al (15)) .
- On a cartesian grid and $d \geq 2$ ($x_i = \{x^1_{i_1}, x^2_{i_2}\}$)

$$\boxed{\overline{\gamma}^{\varepsilon}_{i\,j} = e^{-\frac{||x_i - x_j||^2}{\varepsilon}} = e^{-\frac{||x^1_{i_1} - x^1_{j_1}||^2}{\varepsilon}}\ e^{-\frac{||x^2_{i_2} - x^2_{j_2}||^2}{\varepsilon}}}$$ is separable

  Store $(\sqrt{N} \times \sqrt{N})$ matrices. One Iteration costs $O(N^{1.5})$.
- # iterations increase with $\frac{1}{\varepsilon}$. Stability problems can be fixed.
- Many Generalizations including MMOT check B. et al (2015) Chizat et al (2017) .

IPFP for the MMOT

- M scalings : $\boxed{Q^{\star,\varepsilon}_{i_1,i_1,..,i_M} = u^1_{i_1}\, u^2_{i_2} \dots u^M_{i_M}\, e^{-\frac{c(x_1,..,x_M)}{\varepsilon}}}$

$$c(x_1,..,x_M) = \frac{\|x_{i_2} - x_{i_1}\|^2 + \|x_{i_3} - x_{i_2}\|^2 + .. + \|x_{i_M} - x_{i_{M-1}}\|^2}{dt}$$

- IPFP algebra amounts to

$$\boxed{u^{m,(k)}_{i_m} = \frac{\mu_{i_m}}{\sum_{i_1,..,i_{m-1},\cancel{i_m},i_{m+1},..,i_M} \{.\}}}$$

$$\{.\} \stackrel{\text{def.}}{=} u^{1,(k)}_{i_1} \,..\, u^{m-1,(k)}_{i_{m-1}} \cancel{u^m_{i_m}}\, u^{m+1,(k-1)}_{i_{m+1}} \,..\, u^{M,(k-1)}_{i_M}\, \overline{Q}_{i_1,..,i_M}$$

- Cost again separable (also along dimensions)

$$\boxed{\overline{Q}_{i_1,...,i_M} = \prod_{m=1}^{M-1} \xi_{i_m\, i_{m+1}}} \qquad \xi_{i\,j} = e^{-\frac{\|x_i - x_j\|^2}{dt\,\varepsilon}}$$

- Store (M $\sqrt{N} \times \sqrt{N}$) matrices - one iteration costs $O(M\,N^{1.5})$

Summary

Euler Equations (1755)

$$(E)\begin{cases}\partial_t v + (v\cdot\nabla)v = -\nabla p\\ div(v) = 0\\ \partial_\nu v = 0 \qquad on \quad \partial D\\ v(0,x) = v_0(x)\end{cases}$$

Global in time weak solution open for $d = 3$

(see survey `http://cvgmt.sns.it/paper/1714/` Danieri-Figalli)

Lagrangian formulation

$$\begin{cases}\dot{G}(t,x_0) = v(t, G(t,x_0))\\ \\ G(0,x_0) = x_0 \qquad x_0 \in D\end{cases}$$

- $x_0 \mapsto G(t,x_0) \in \mathbb{S}\mathit{diff}(D)$.

$$\boxed{\mathbb{S}\mathit{diff}(D) \stackrel{\text{def.}}{=} \{S \in L^2(D;D);\ \mathit{diff}.\ s.t.\ det(\nabla S) = 1\}}$$

- $\ddot{G}(t,x_0) = -\nabla p(t, G(t,x_0)) \perp T_{G(t,x_0)}\mathbb{S}\mathit{diff}(D)$

Arnold's geodesics formulation (1966)

$$G = Arg \inf_{S \in \mathcal{H}diff} \frac{1}{2} \int_{[0,1]\times D} \|\dot{S}(t, x_0)\|^2 \, dx_0 \, dt$$

$$\mathcal{H}diff \overset{\text{def.}}{=} \{S \in H^1([0,1], \mathbb{S}diff(D)),\ S(0,.) = Id,\ S(1,.) = S^*\}$$

- Pbm : Lack of completeness of $\mathbb{S}diff(D) \subset L^2(D)$ (Shnirelman 85) : pathological examples with no minimizers in $\mathbb{S}diff$.
- Set of measure preserving mapping

  $$\mathbb{S}(D) := \{S \in L^2(D; D);\ S_\# \mathcal{L}_D = \mathcal{L}_D\}$$

  $(= \overline{\mathbb{S}diff(D)})$ for $d \geq 3$)
- Idea (Brenier (89)) : Replace $\mathcal{H}diff$ with $\mathcal{H} = \{S \in H^1([0,1], \mathbb{S}(D)),\ S(0,.) = Id,\ S(1,.) = S^*\}$ and approximate $\mathbb{S}(D)$ with permutations

1-D Computations with permutations - Brenier (89)

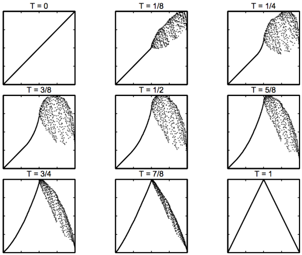


FIG. 1: APPROXIMATE GEODESIC FOR MAP 1

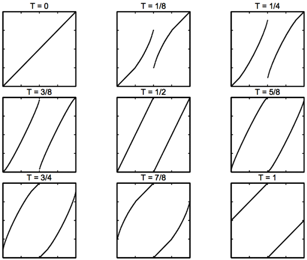


FIG. 2: APPROXIMATE GEODESIC FOR MAP 2

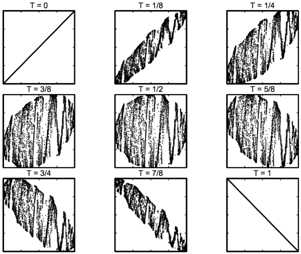


FIG. 3: APPROXIMATE GEODESIC FOR MAP 3

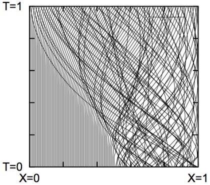
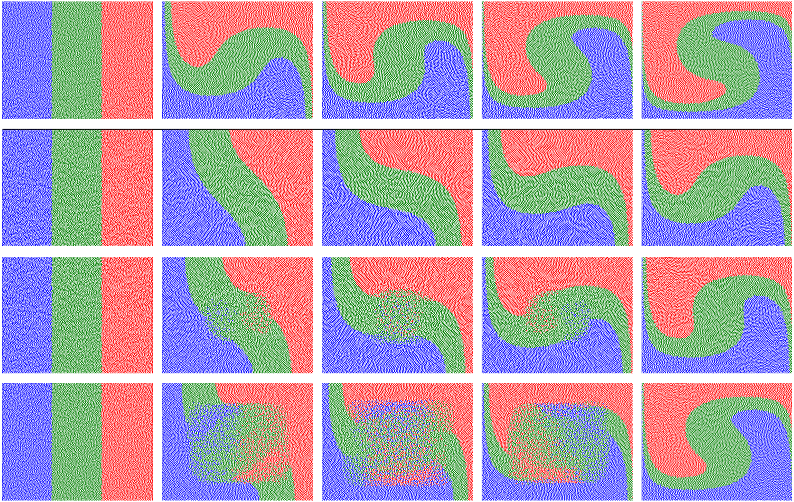


FIG. 4: TRAJECTORIES FOR MAP 1

Beltrami flow on $D = [0,1]^2$

$V(x,y) = \{-\cos(\pi x)\sin(\pi y), \sin(\pi x)\cos(\pi y)\}$

$p(x,y) = \frac{1}{4}(\sin^2(\pi x) + \sin^2(\pi y))$



Generalized incompressible flows Brenier (89)

- Back to $\Omega(D) = C([0,1]; D)$ and $Q \in \mathcal{P}(\Omega(D))$.
- Incompressibility: $\boxed{(e_t)_\# Q = \mathcal{L}_D \text{ for all } t}$

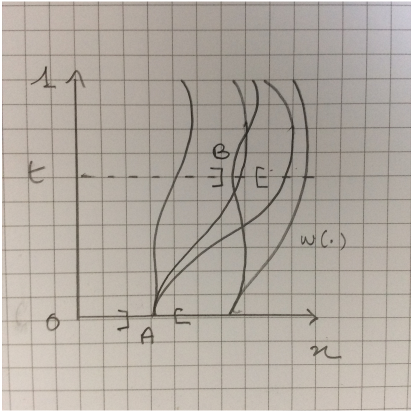
  $(e_t)_\# Q(A) = Q(\{\omega \in \mathcal{P}(\Omega(D)), \omega(t) \in A)\}$ forall $A \subset D$.

Probability of transition $0 \to t$:

$(e_0,\ e_t)_\# Q(A, B) =$

$Q(\{\omega,\ \omega(0) \in A \text{ and } \omega(t) \in B\})$

forall $(A, B) \subset D_0 \times D_t$.



Generalized incompressible flows Brenier (89)

- Encoding time boundary conditions :
  with a deterministic transport plan

$$(Id,\ S^*)_\# \mathcal{L}_D(B_0, B_1) = \mathcal{L}_D(\{x_0 \in B_0,\ \text{s.t. } S^*(x_0) \in B_1\})$$

- Finally :

$$\min \begin{cases} Q \in \mathcal{P}(\Omega(D)) \text{ s. t.} \\ (e_t)_\# Q = \mathcal{L}_D \text{ for all } t \\ (e_0,\ e_1)_\# Q = (Id,\ S^*)_\# \mathcal{L}_D \end{cases} \int_{\Omega(D)} \int_{[0,1]} \|\frac{1}{2}\dot{\omega}(t)\|^2 \, dt \, dQ(\omega)$$

- Q is "deterministic" if there is a a map $t \mapsto G_t \in \mathbb{S}$ s.t.

$$(e_0, e_t)_\# Q = (Id,\ G_t)_\# \mathcal{L}_D$$

Generalized incompressible flows : properties

- $\mathbb{S}\mathit{diff}$ dense in $\mathbb{S}$ for $d \geq 3$ but not for $d = 2$. Pathological examples with no minimizers in $\mathbb{S}\mathit{diff}$ (and $\mathbb{S}$) (Shnirelman 85).
- Q-Minimizers exists and there is a unique pressure.
- Consistency (Brenier 89) : given an admissible $Q \in \mathcal{P}(\Omega(D))$ and pressure p s. t.

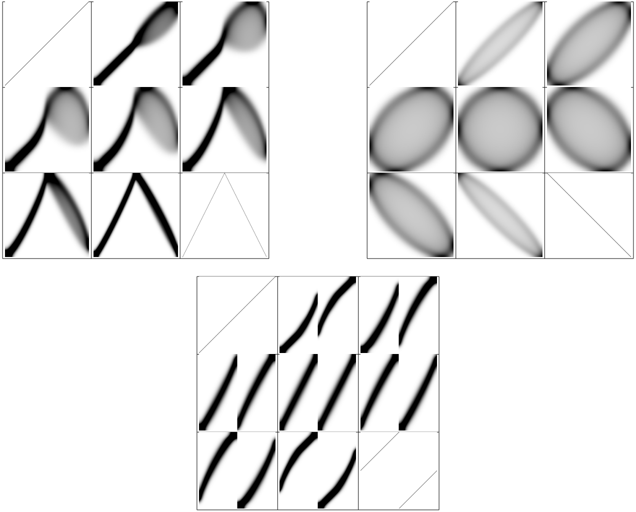
$$\ddot{\omega}(t) = -\nabla p(t, \omega(t)), \quad Q - a.e.$$

$$\sup_{(t,x)\in[0,T]\times D} \nabla_x^2 p(t,x) \leqslant \frac{\pi^2}{T^2} Id$$

  then Q solves (P_γ).
- If $< \leftarrow \leqslant$ above, Q is unique and deterministic.

Tests 1D (B. et al 2015)

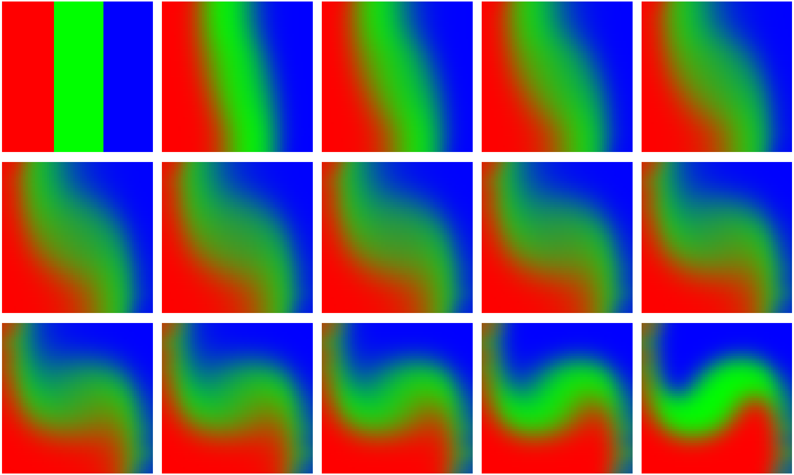
$$\boxed{(e_0, e_t)_\# Q^{\varepsilon,*} \in \mathcal{P}(D_0 \times D_t)}$$



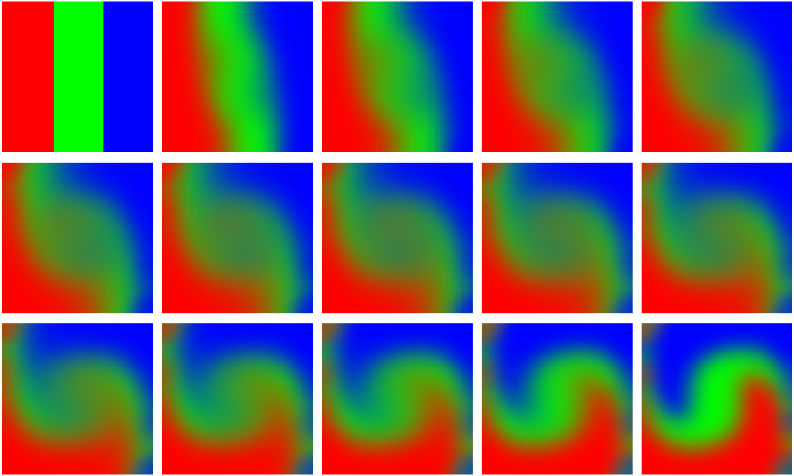
Test Beltrami $T = 0.9$ - Nenna (2016)

$R\,(G, B) \stackrel{\text{def.}}{=}$ red (green, blue) subset of D_0

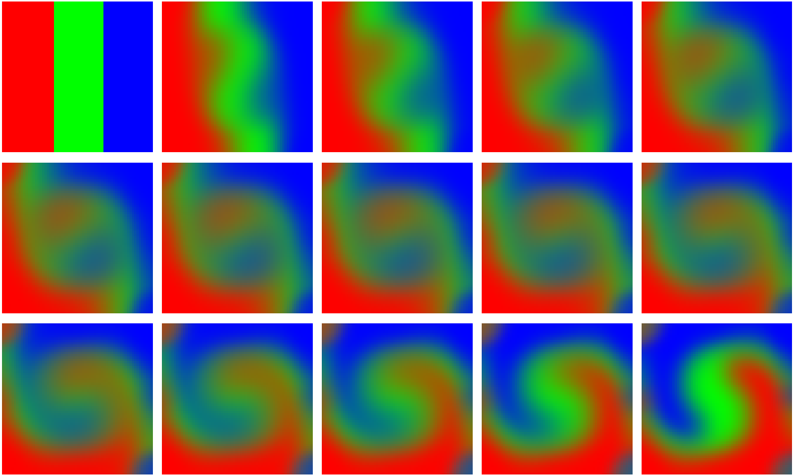
$$(e_0, e_t)_\# Q^{\varepsilon,*}(R/G/B, .) \in \mathcal{P}(D_t)$$



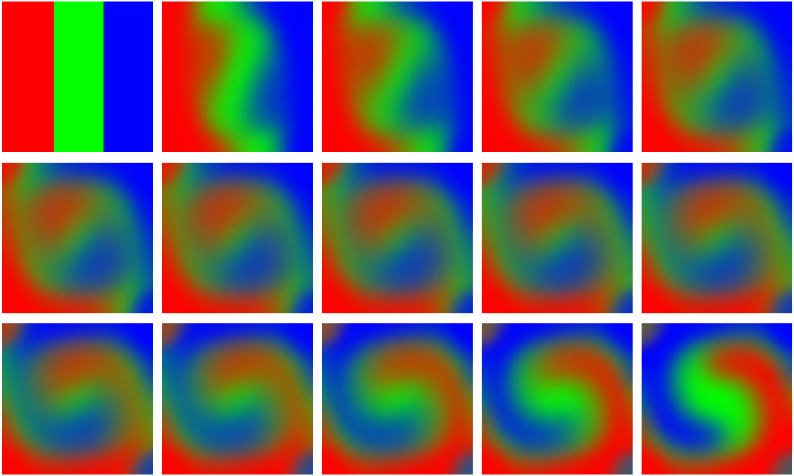
Test Beltrami $T = 1.1$



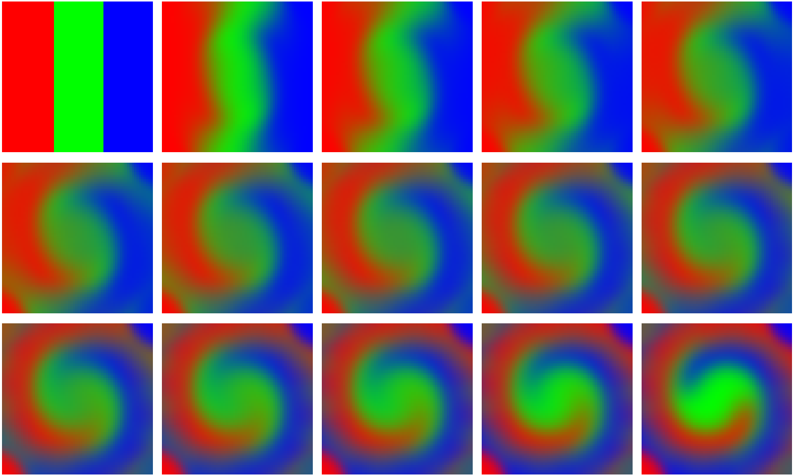
Test Beltrami $T = 1.3$



Test Beltrami $T = 1.5$



Test Beltrami $T = \pi$



GRACIAS por su atención !