On multi-marginal optimal transport

Guillaume Carlier [a] **.**

Brenier's fest, Calculus of Variations and Optimal Transport, Institut Henri Poincaré, January 2017.

[a] CEREMADE, Université Paris Dauphine and MOKAPLAN, Inria-Paris

Yann Brenier and (applied) transport

Transports BRENIER

Aim of this talk is to review a few results on some multi-marginal problems i.e. problems of the form:

$$\inf_{\gamma\in\Pi(\mu_1,\cdots,\mu_N)} \int_{X_1\times\cdots\times X_N} c(x_1,\cdots,x_N)\mathrm{d}\gamma(x_1,\cdots,x_N), \tag{1}$$

where $\Pi(\mu_1,\cdots,\mu_N)$ is the set of probability measures on $X_1\times\cdots\times X_N$ having $\mu_1,\cdots,\mu_N$ as marginals. Much less is known than in the two-marginals case (e.g. Monge solution for twisted costs, Brenier, McCann, Gangbo theory, regularity...).

Important motivation: Brenier's (CPAM, 1999) relaxation of Arnold's interpretation of incompressible Euler as a geodesic problem on the group of measure preserving diffeomorphisms.

$$\partial_t u + \nabla \cdot (u \otimes u) + \nabla p = 0,\ \nabla \cdot u = 0,\ t \in (0,T),\ x \in \mathbf{T^d} \quad (2)$$

The two-endpoints problem asks that the flow X of u at the final time T is a prescribed measure-preserving map h. At least formally (2) is the Euler-Lagrange equation for the minimization of

$$\int_0^T \int_{\mathbf{T^d}} |\dot{X}(t,x)|^2 \mathrm{d}x\mathrm{d}t$$

subject to $X_0 = \mathrm{id}$, $X_T = h$ and of course that X_t is measure-preserving for every t.

No minimizer in general (Shnirelman). Yann's relaxation (in a nutshell):

$$\inf_{Q\in\Gamma(h)} I(Q) := \int_\Omega E(\omega)\ \mathrm{d}Q(\omega) \tag{3}$$

where $E(\omega) = \int_0^T |\dot\omega(t)|^2 \mathrm{d}t$ and $\Gamma(h)$ is the set of probability measures on $\Omega = C([0,T], \mathbf{T^d})$ such that

$$e_{t\#}Q = \mathcal{L},\ t \in [0,T],\ (e_0, e_T)_\# Q = (\mathrm{id}, h)_\# \mathcal{L},$$

where e_t is the evaluation map a time t and $\mathcal{L}$ is Lebesgue's measure on $\mathbf{T^d}$.

It is an infinitely many-marginals limit case $N \to \infty$ of (1),

$$\inf_{\gamma \in \Pi(\mathcal{L}, \cdots, \mathcal{L})} \int_{(\mathbf{T^d})^\mathbf{N}} C_N(x_1, \cdots, x_N) \mathrm{d}\gamma(x_1, \cdots, x_n)$$

with

$$C_N(x_1, \cdots, x_N) = N \sum_{k=1}^{N-1} |x_{k+1} - x_k|^2 + N|x_N - h(x_1)|^2.$$

Without the last term: quadratic multi-marginal OT by Gangbo-Święch. (1996) (I'll come back to this in relation with Wasserstein barycenters).

Outline

① Matching for teams

② Wasserstein barycenters

③ Limit behavior

④ Numerics

Matching for teams

Matching for teams Joint with I. Ekeland. Market for houses, quality $z \in Z$. For one house z to be available, need for one buyer and a team of producers (mason, plumber, electrician). We shall denote by the index $i \in \{1, \cdots, I\}$ the different populations (buyers, plumbers, electricians, masons...), the agents in each population are hetererogeneous, they are characterized by a certain type which affects their (quality dependent) cost function.

More precisely for each i, we are given a compact metric space of types X_i and a cost function $c_i \in C(X_i \times Z, \mathbf{R})$ with the interpretation that $c_i(x_i, z)$ is the cost for an agent of population i with type x_i to work in a team that produces good z. The distribution of type x_i in population i is known and given by some Borel probability measure $\mu_i \in \mathcal{P}(X_i)$. We look for an equilibrium i.e. a system of transfers (paid by the buyer to the producers) which clears the markets: everybody is in a team and the supply and demands for houses coincides.

A system of transfers is a collection of continuous functions $\varphi_1, \ldots \varphi_I \colon Z \to \mathbf{R}$ where $\varphi_i(z)$ is the amount paid to i by the other members of the team for producing z. An obvious equilibrium requirement is that teams are self-financed i.e.

$$\sum_{i=1}^{I} \varphi_i(z) = 0, \ \forall z \in Z. \tag{4}$$

Given transfers $\varphi_1, \ldots \varphi_I$, an agent from population i with type $x_i \in X_i$, gets a net minimal cost given by c_i-transform of φ_i:

$$\varphi_i^{c_i}(x_i) := \min_{z \in Z} \{c_i(x_i, z) - \varphi_i(z)\} \tag{5}$$

Agents are rational: they choose cost miniminizing qualities i.e. a $z \in Z$ such that

$$\varphi_i^{c_i}(x_i) + \varphi_i(z) = c_i(x_i, z). \tag{6}$$

The last unknown is a collection of plans $\gamma_i \in \mathcal{P}(X_i \times Z)$ such that $\gamma_i(A_i \times A)$ represents the probability that an agent in population i has a type in A_i and belongs to a team that produces a quality in A. At equilibrium the first marginal of γ_i should be μ_i (this is equilibrium on the i-th labor market) and the second marginal of γ_i should not depend on i (this is equilibrium on the quality good market), this common marginal represents the equilibrium quality line.

Equilibrium: transfer system $(\varphi_1, \dots \varphi_I) \in C(Z, \mathbf{R})^I$, probability measures $\gamma_i \in \mathcal{P}(X_i \times Z)$ and a probability measure $\nu \in \mathcal{P}(Z)$ such that

- teams are self-financed i.e. (4) holds,
- $\gamma_i \in \Pi(\mu_i, \nu)$ for $i = 1, \dots, I$ (equilibrium on the labor markets and on the good market),
- (6) holds on the support of γ_i for $i = 1, \dots, I$ (i.e. agents choose cost minimizing qualities).

Variational characterization of equilibria

$$\inf_{\nu \in \mathcal{P}(Z)} J(\nu) := \sum_{i=1}^{I} W_{c_i}(\mu_i, \nu) \tag{7}$$

and its dual (concave maximization) formulation

$$\sup \left\{ \sum_{i=1}^{I} \int_{X_i} \varphi_i^{c_i}(x_i)\mu_i(dx_i) \, : \, \sum_{i=1}^{I} \varphi_i = 0 \right\}. \tag{8}$$

Theorem 1 *$(\varphi_i, \gamma_i, \nu)$ is an equilibrium if and only if:*

- *ν solves (7),*
- *the transfers $(\varphi_1, \dots \varphi_I)$ solve (8),*
- *for $i = 1, \dots, I$, γ_i solves the Monge-Kantorovich problem $W_{c_i}(\mu_i, \nu)$.*

Wasserstein barycenters

Given $(\mu_1, \ldots, \mu_I) \in \mathcal{P}_2(\mathbf{R}^d)$ and weights $\lambda_i > 0$ summing to 1, consider:

$$\inf_{\nu \in \mathcal{P}_2(\mathbf{R}^d)} J(\nu) := \sum_{i=1}^{I} \lambda_i W_2^2(\mu_i, \nu) \tag{9}$$

Existence is obvious. Special case of Fréchet mean. Special case of the matching for teams problem (quadratic costs).
Extensions to Riemannian manifolds Kim, Pass.

Uniqueness holds as soon as one of the μ_i's does not give mass to small sets (in this case $\nu \mapsto W_2^2(\mu_i, \nu)$ is strictly convex).

Proposition 1 *Assume that there is an index $i \in \{1, \ldots I\}$ such that μ_i vanishes on small sets. Then (10) admits a unique solution $\overline{\nu}$.*

As soon as one of the μ_i's vanishes on small sets , this therefore enables one to define unambiguously the barycenter $(\text{bar}(\mu_i, \lambda_i)_{i=1,\ldots,I})$ of the μ_i's with weights λ_i. It is known that Fréchet means are unique on nonpositively curved metric spaces (Sturm) BUT the Wasserstein space is not nonpositively curved! Here, it is standard convexity which matters.

Characterization by duality. Dual of (9):

$$\sup\left\{F(f_1,...,f_I)=\sum_{i=1}^{I}\int_{\mathbf{R}^d} S_{\lambda_i}f_i d\mu_i \ : \sum_{i=1}^{I} f_i=0, \right\} \qquad (10)$$

where

$$S_\lambda f(x) := \inf_{y\in\mathbf{R}^d}\left\{\frac{\lambda}{2}|x-y|^2 - f(y)\right\}, \ \ \forall x\in\mathbf{R}^d,\ \lambda>0.$$

Both the infimum in (9) and the supremum in (10) are attained and values coincide.

Let $(f_1, ..., f_I)$ be a solution of (10) and define the convex potentials:

$$\lambda_i \phi_i(x) := \frac{\lambda_i}{2}|x|^2 - S_{\lambda_i} f_i(x), \tag{11}$$

by duality, if $\overline{\nu}$ solves (9) and γ_i is an optimal transport plan between μ_i and $\overline{\nu}$ then:

- the support of γ_i is included in $\partial\phi_i$,
- $\sum_i \phi_i^*(y) \leq \frac{1}{2}|y|^2$ for all $y \in \mathbf{R}^d$ with an equality on the support of $\overline{\nu}$.

These duality conditions lead to the following characterization

Proposition 2 *Assume that* μ_i *vanishes on small sets for every* $i = 1, .., p$*, then the following conditions are equivalent:*

1. $\overline{\nu} = (\text{bar}(\mu_i, \lambda_i)_{i=1,\dots,I})$,
2. *There exist convex potentials* ϕ_i *such that* $\nabla\phi_i$ *is Brenier's map transporting* μ_i *to* $\overline{\nu}$*, and a constant* C *such that*

$$\sum_{i=1}^{p} \lambda_i \phi_i^*(y) \leq C + \frac{|y|^2}{2},\ \forall y \in \mathbf{R}^d, \text{ with equality } \overline{\nu}\text{-a.e.} \quad (12)$$

The optimality conditions for the barycenter $\overline{\nu}$, at least formally take the form of an obstacle-like system of Monge-Ampère equations. Set $\psi_i = \phi_i^*$, so $\mu_i = \nabla\psi_{i\#}\overline{\nu}$, therefore

$$\overline{\nu} = \mu_i(\nabla\psi_i)\det(D^2\psi_i),\ i = 1,\dots,I \tag{13}$$

which is supplemented with

$$\sum_{i=1}^{I} \lambda_i \psi_i(y) \leq \frac{|y|^2}{2} \tag{14}$$

with equality on the support of $\overline{\nu}$, yielding in particular (formally)

$$\sum_{i=1}^{I} \lambda_i \nabla\psi_i = \mathrm{id}\ \ \overline{\nu}\text{-a.e.}.$$

There is another characterization of the barycenter in terms of a multi-marginal quadratic optimal transport problem which was solved by Gangbo-Święch. Adopting this viewpoint one finds the following structure for the barycenter:

$$\overline{\nu} = \left(\lambda_1 \mathrm{id} + \sum_{i=2}^{I} \lambda_i \nabla u_i^* \circ \nabla u_1 \right)_{\#} \mu_1$$

where each u_i is strongly convex ($D^2 u_i \geq \lambda_i \mathrm{id}$) and such that $\nabla u_i^* \circ \nabla u_1$ transports μ_1 to μ_i.

With M. Agueh, we observed that this gives integral estimates for barycenters. Indeed, take $F : [0, \infty) \to \mathbf{R}$, $F(0) = 0$ continuous, satisfying McCann's condition
$(0, \infty) \ni t \mapsto t^d F(t^{-d})$ is convex and nonincreasing, then

Proposition 3 *The internal energy* $E(\rho) = \int_{\mathbf{R}^d} F(\rho(x))\, dx$ *is convex along barycenters i.e.*

$$E(\mathrm{bar}(\mu_i, \lambda_i)_{i=1,\dots,I}) \le \sum_{i=1} \lambda_i E(\mu_i).$$

Proof. Write $\overline{\nu} := \mathrm{bar}(\mu_i, \lambda_i)_{i=1,\dots,I}) = (\sum_{i=1}^{I} \nabla u_i^*)_{\#}\nu_1$ with $\nu_1 := \nabla u_{1\#}\mu_1$, so $T_i := \nabla u_i^*$ transports ν_1 to μ_i and $T := \sum_{i=1}^{I} \lambda_i T_i$ transports ν_1 to $\overline{\nu}$, so

$$E(\nu) = \int_{\mathbf{R}^d} F\Big(\frac{\nu_1}{\det DT}\Big) \det DT$$

and $\det DT^{1/d} \geq \sum_{i=1}^{N} \lambda_i \det(DT_i)^{1/d}$; one gets the result by McCann's condition. □

In particular for $F(t) = t^p$, $p \in (1, +\infty)$ this gives L^p bounds:

$$\int_{\mathbf{R}^d} \overline{\nu}(x)^p \mathrm{d}x \leq \sum_{i=1}^{N} \lambda_i \|\mu_i\|_{L^p}^p.$$

as well as the limiting cases: $\mu_i \in L^1 \Rightarrow \overline{\nu} \in L^1$, and as for the L^∞ case:

Theorem 2 *Let* $(\mu_1, ..., \mu_I)$ *be probability measures with finite second moments and let* $(\lambda_1, ..., \lambda_I)$ *be positive reals that sum to* 1*. If* $\mu_1 \in L^\infty$*, then* $\overline{\nu} := \mathrm{bar}((\mu_i, \lambda_i)) \in L^\infty$ *and more precisely:*

$$\|\overline{\nu}\|_{L^\infty} \leq \frac{1}{\lambda_1^d} \|\mu_1\|_{L^\infty}. \tag{15}$$

It is not kown if there is more regularity? Like $\mu_i \in C^{k,\alpha} \Rightarrow \overline{\nu} \in C^{k,\alpha}$ (in the periodic case say)? Also is $\overline{\nu}$ Lipschitz in the weights?

Examples

- $I = 2$ the barycenter of $(\mu_0, (1-t))$ and (μ_1, t) is McCann's (or Wasserstein geodesic) interpolant:

$$\nu_t := ((1-t)\mathrm{id} + t\nabla\phi)_{\#}\mu_0 = (t\mathrm{id} + (1-t)\nabla\phi^*)_{\#}\mu_1$$

  where $\nabla\phi$ is the Brenier's map between μ_0 and μ_1.

- $d = 1$, in dimension one $\mathrm{bar}(\mu_i, \lambda_i)_i$ is simply obtained as the image of μ_1 by the linearly interpolated transport map $\sum_i \lambda_i T_i^1$ where T_i^1 is the monotone transport from μ_1 to μ_i.

- Gaussian case μ_i is a gaussian measure with mean 0 and covariance matrix S_i, $\overline{\nu} = \mathcal{N}(0, \overline{S})$ where $\overline{S}$ is the unique positive definite root of the matrix equation

$$\sum_{i=1}^{p} \lambda_i \left(S^{1/2} S_i S^{1/2}\right)^{1/2} = S. \tag{16}$$

Limit behavior

Given a well-behaved $m \in \mathcal{P}(\mathcal{P}_2(\mathbf{R}^d))$ (for instance supported on regular measures supported on some ball), Bigot and Klein have considered the barycenter of m, $\overline{\nu} = \text{bar}(m)$ as the minimizer of

$$\nu \mapsto \int_{\mathcal{P}_2(\mathbf{R}^d)} W_2^2(\mu, \nu) \mathrm{d}m(\mu)$$

Assume now that for some (large N) we are given an i.i.d. sample $\hat{\mu}_1, \cdots, \hat{\mu}_N$ drawn according to m and define (the random measure)

$$\hat{\nu}_N := \text{bar}\Big((\hat{\mu}_1, \cdots, \hat{\mu}_N, \frac{1}{N}) \Big).$$

From what we saw before, mild conditions on m ensure that $\overline{\nu}$ and $\hat{\nu}_N$ have densities.

Bigot and Klein derived from the usual law of large numbers a LLN for Wasserstein barycenters, namely that

$$W_2^2(\hat{\nu}_N, \overline{\nu}) \to 0 \text{ a.s.}$$

in other words the $L^2(\overline{\nu})$ random map $\hat{T}_N$ that transports optimally the true barycenter $\overline{\nu}$ to its empirical counterpart $\hat{\nu}_N$ converges to the identity map a.s.. It is tempting to go one step further through some sort of CLT, which is the validity of

$$\hat{h}_N := \sqrt{N}(\hat{T}_N - \mathrm{id})$$

converging in law to some gaussian on $\mathcal{N}(0, \Sigma)$ for a certain trace class, positive self adjoint Σ on $L^2(\overline{\nu}, \mathbf{R}^d)$.

This seems difficult to establish in general but with M. Agueh, we could prove validity in some cases

Proposition 4 *The Wasserstein CLT holds in the following cases:*

- *m is a Benoulli between two $\mathcal{P}_2(\mathbf{R}^d)$ measures, one of which vanishes on small sets,*
- *$d = 1$ and m is concentrated on a bounded (for W_2) set of nonatomic measures,*
- *$m = \sum_{i=1}^{I} \lambda_i \delta_{\mu_i}$ for nondegenerate gaussian μ_i's.*

Numerics

By now, there are fast solvers for entropic-regularized approximation of multi-marginals OT following the seminal work of Cuturi, Benamou et al., Nenna. Take the matching for teams example:

$$\inf_{\gamma:=(\gamma_1,\cdots,\gamma_I)\in C_1\cap C_2} \left\{ \sum_{i=1}^{I} F(\gamma) := \int_{X_i\times Z} c_i(x,z)\mathrm{d}\gamma_i(x_i,z) \right\}$$

where

$$C_1 := \{(\gamma_1,\ \cdots,\gamma_I)\ :\ \pi_{X_i\#}\gamma_i = \mu_i,\ \forall i\}$$

and

$$C_2 := \{(\gamma_1,\ \cdots,\gamma_I)\ : \exists \nu \in \mathcal{P}(Z)\ \pi_{Z\#}\gamma_i = \nu,\ \forall i\}.$$

Entropic regularization, $\varepsilon > 0$

$$\inf_{\gamma \in C_1 \cap C_2} F(\gamma) + \varepsilon \sum_{i=1}^{I} \int_{X_i \times Z} \gamma_i(x_i, z) \ln(\gamma_i(x_i, z)) \, \mathrm{d}x_i \mathrm{d}z$$

which is the same as

$$\inf_{\gamma \in C_1 \times C_2} \sum_{i=1}^{N} \mathrm{KL}(\gamma_i | \theta_i)$$

where $\theta_i := e^{-c_i/\varepsilon}$ and

$$\mathrm{KL}(\gamma_i | \theta_i) := \int_{X_\times Z} \gamma_i(x_i, z) \ln\left(\frac{\gamma_i(x_i, z)}{\theta_i(x_i, z)}\right) \, \mathrm{d}x_i \mathrm{d}z$$

Bregman iterative projections (aka IPFP, Sinkhorn...) start from $\gamma_0 = \theta$ and KL alternately project onto C_1 and C_2. The projection onto C_1 is close form, the projection onto C_2 a well, $\gamma := \mathrm{proj}^{C_2}_{KL}(\overline{\gamma})$ is indeed given by

$$\gamma_i(x_i, z) = \overline{\gamma}_i(x_i, z) a_i(z)$$

where

$$a_i(z) := \frac{(\overline{\nu}_1(z) \times \cdots \times \overline{\nu}_I(z))^{\frac{1}{I}}}{\overline{\nu}_i(z)}$$

and

$$\overline{\nu}_i(z) := \int_{X_i} \overline{\gamma}_i(x_i, z) \mathrm{d}x_i.$$

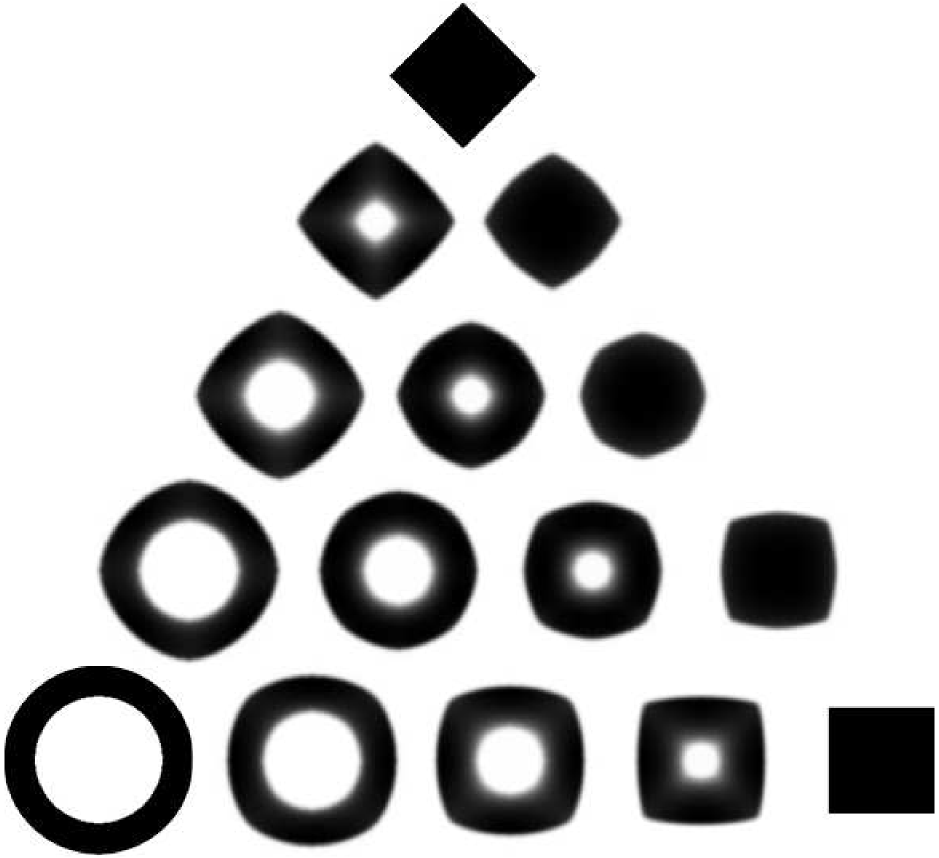
Convergence is well undertsood (Contraction for the Hilbert projective metric etc...).

Old idea (goes back too Schrödinger), related to several streams of research:

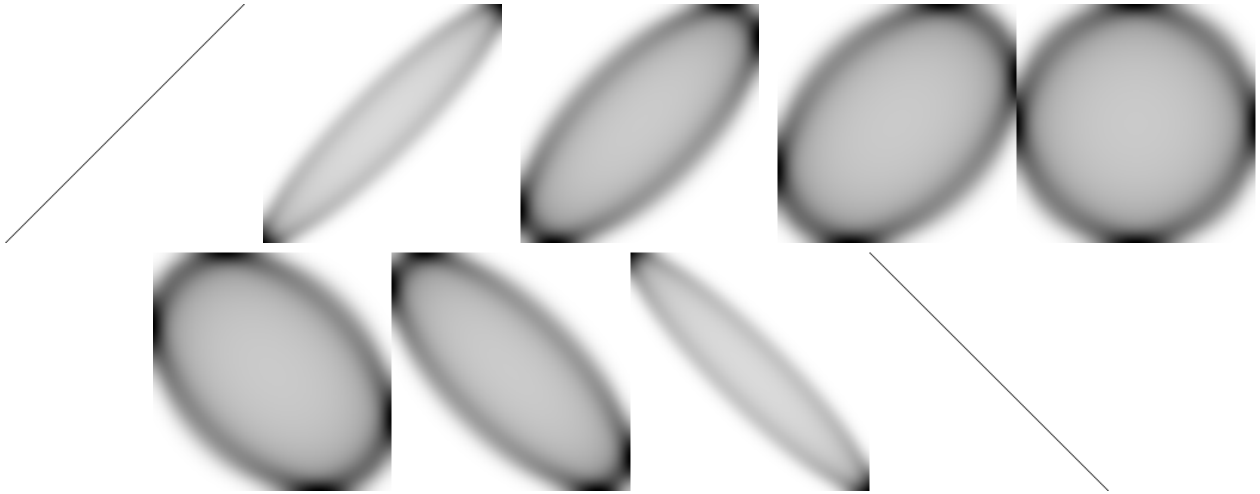
- probability in statistics (Csisczar, Dykstra, Léonard, Föllmer, Rüschendorf...),
- optimization (Bregman, Bauschke, Lewis, Cominetti, San-Martin...).

OT framework: Galichon, Salanié (economics) and Cuturi (machine learning, link with Sinkhorn). Easily adapts to multi-marginals, partial transport, barycenters.... cf. Benamou, C., Cuturi, Nenna and Peyré.

An application of entropic regularization to W_2 barycenters:



An application to Euler in 1d (see the beautiful simulations of L. Nenna in higher dimensions):



On repulsive costs

So far, I have only addressed easy cases where the structure of the cost is mainly attractive, there are many other relevant cases which are far from being fully understood and where the cost is rather repulsive. The most challenging example is the Coulomb cost which arises in density functional theory:

$$\inf_{\gamma\in\Pi(\rho,\cdots,\rho)} \int_{(\mathbf{R}^d)^N} \sum_{i<j} \frac{1}{|x_i - x_j|} \mathrm{d}\gamma(x_1,\cdots,x_N), \qquad (17)$$

N electrons, ρ: single-electron density. Received a lot of attention in recent years (Friesecke, Cotar, Pass, Klüppelberg, Buttazzo, De Pascale, Gori-Giorgi, Colombo, Di Marino, Nenna...). Symmetric problem (Ghoussoub-Moameni, Ghoussoub-Maurey). Determinant cost (C.-Nazaret).

A striking example due to S. Di Marino, L. Nenna and A. Gerolin (2015), take $N = 3$, $X_1 = X_2 = X_3 = [0, 1]$, $\mu_1 = \mu_2 = \mu_3 = \mu$ is the Lebesgue measure on $[0, 1]$ and $c(x_1, x_2, x_3) = (x_1 + x_2 + x_3)^2$ (repulsive harmonic case) then there exists an optimal plan of the form $(\mathrm{id}, T, T \circ T)_{\#}\mu$ with $T \circ T \circ T = \mathrm{id}$ and T is fractal (nowhere smooth). This is also a challenge for numerics.

Happy birthday Yann!! C'est gééénial (the other Benamou-Brenier formula).