On recent progress in the optimal matching problem

Joint works with L. Ambrosio, D. Trevisan, M. Huesmann, F. Otto

Michael Goldman

CNRS, CMAP, École Polytechnique

The Gaussian isoperimetric problem

For $m \in (0,1)$, let $\mathcal{U}(m)$ be the Gaussian perimeter of a half space of Gaussian measure m. In 97 Bobkov proved that

$$\mathcal{U}\left(\int f d\gamma\right) \leq \int \sqrt{\mathcal{U}^2(f) + |\nabla f|^2} d\gamma.$$

As a consequence he derived the Gaussian isoperimetric inequality

$$\mathcal{U}(m) \leq P_\gamma(E) \qquad \forall E \text{ with } \gamma(E) = m.$$

In 2012 we proved with M. Novaga that in the Wiener space (here bar denotes lsc enveloppe)

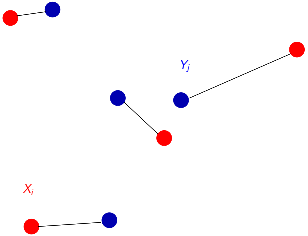
$$\bar{P}_\gamma(f) = \int \sqrt{\mathcal{U}^2(f) + |\nabla f|^2} d\gamma$$

The bi-partite optimal matching problem

For X_i and Y_j iid uniformly distributed in $[0,1]^d$ and $p \geq 1$, the optimal matching problem is

$$\min_{\sigma} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p$$

where the min is among all permutations σ.



Vastly studied combinatorial problem, related to TSP, MST...
Typical questions: **expectation of the energy**, description of the optimal σ. For the second aspect, see Ambrosio-Glaudo-Trevisan, Clozeau-Mattesini or G-Huesmann-Otto.

Reformulation as an optimal transport problem

If $\mu(\mathbb{R}^d) = \lambda(\mathbb{R}^d)$, the $p-$Wasserstein distance is defined as

$$W_p^p(\mu, \lambda) = \inf_{\pi \in \Pi(\mu,\lambda)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\pi,$$

where $\pi \in \Pi(\mu, \lambda)$ if $\pi_1 = \mu$ and $\pi_2 = \lambda$.
Rk: this is a **linear programming** problem.

Let $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $\lambda = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$. By Birkhoff,

$$\frac{1}{n} \min_{\sigma} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^p = W_p^p(\mu, \lambda).$$

Matching to the reference measure

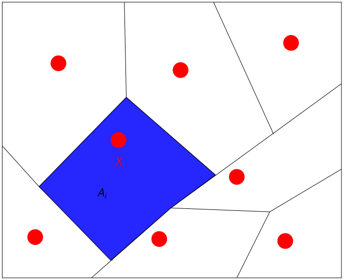
A natural variant is the matching between $\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ and the reference measure i.e.

$$W^p_{[0,1]^d}(\mu, 1),$$

where more generally for $\Omega \subset \mathbb{R}^d$,

$$W^p_{\Omega}(\mu,\kappa) = W^p_p(\mu \llcorner \Omega, \kappa\chi_\Omega)$$

with $\kappa = \frac{\mu(\Omega)}{|\Omega|}$.



Notice that this problem may be seen as an optimal tessellation (Laguerre cells) problem since

$$W^p_{[0,1]^d}(\mu, 1) = \min_{A_i \cap A_j = \emptyset, |A_i| = \frac{1}{n}} \sum_{i=1}^{n} \int_{A_i} |x - X_i|^p dx.$$

Because of their connections to random graph theory, probability (empirical measures), theoretical physics etc.., both problems as well as many variants have received a lot of attention (Yukich, Talagrand, Steel, Ajtai-Komlós-Tusnády, Caracciolo and al., Barthe-Bordenave, Ledoux, Bobkov...).

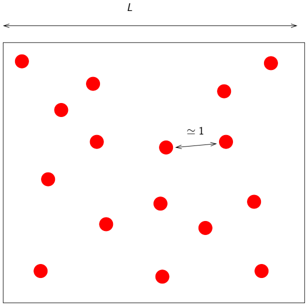
In most of the talk we focus on the **matching to the reference measure**. Also $d = 1$ is very special (see Bobkov-Ledoux) so we only focus on $d \geq 2$.

Rescaling

Letting $L = n^{\frac{1}{d}}$ and making the change of variables $y = Lx$, the original matching problem is equivalent to

$$\frac{1}{L^d} W^p_{[0,L]^d}(\mu, 1)$$

where $\mu = \sum_{i=1}^{L^d} \delta_{X_i}$ with X_i iid uniformly distributed in $[0, L]^d$.



In these variables the typical distance between points is of order 1 (the microscale) and the size of the system is $L \gg 1$ (the macroscale).

Scaling laws

Since the distance between points is order 1, one can expect $\mathbb{E}[\frac{1}{L^d} W^p_{[0,L]^d}(\mu, 1)] \sim 1$ for $L \gg 1$. However,

Theorem (Ajtai-Komlós-Tusnády, Talagrand, Bobkov-Ledoux...)

For $p \geq 1$ and $L \gg 1$,

$$\mathbb{E}\left[\frac{1}{L^d} W^p_{[0,L]^d}(\mu, 1)\right] \sim \begin{cases} \log^{\frac{p}{2}} L & \textit{if } d = 2 \\ 1 & \textit{if } d \geq 3 \end{cases}$$

Question: Are these bounds asymptotically sharp?

The PDE ansatz of Caracciolo and al.

At scales $\gg 1$, $\mu = \sum_{i=1}^{L^d} \delta_{X_i} \sim 1$ and thus all quantities which depend on mesoscopic or macroscopic scales are well described by the linearized problem.
Recall that for $p = 2$, by Brenier Theorem, optimal coupling given by $\pi = (T \times \mathrm{Id})\#1$, where $T = \nabla\psi$ is the gradient of a convex function which solves (formally) the Monge-Ampère equation

$$\det D^2\psi = \frac{1}{\mu}.$$

For large scales $\mu \simeq 1$ and writing that $\psi = \frac{1}{2}|x|^2 - \phi$ with ϕ small,

$$\det D^2\psi = \det(\mathrm{Id} - D^2\phi) \simeq 1 - \Delta\phi, \qquad \frac{1}{\mu} = \frac{1}{1 + (\mu - 1)} \simeq 2 - \mu$$

$$\Longrightarrow \boxed{\Delta\phi \simeq \mu - 1 \qquad \text{and} \qquad \int_{[0,L]^d} |T - x|^2 \simeq \int_{[0,L]^d} |\nabla\phi|^2.}$$

i.e. linearization of W_2 around 1 is H^{-1}.

Using this ansatz, Caracciolo and al. predict in particular:

$$\mathbb{E}\left[\frac{1}{L^d} W^2_{[0,L]^d}(\mu, 1)\right] = \begin{cases} \frac{1}{2\pi}\log L + f_{2,2} + o(1) & \text{if } d = 2 \\ f_{2,d} + \frac{\zeta_d(1)}{4\pi^2}\frac{1}{L^{d-2}} + o\left(\frac{1}{L^{d-2}}\right) & \text{if } d \geq 3, \end{cases}$$

where $f_{2,d}$ are numerically estimated constants and

$$\zeta_d(s) = \sum_{n \in \mathbb{Z}^d \setminus \{0\}} \frac{1}{|n|^{2s}}.$$

Also some related predictions for $p \neq 2$.

Rigorous results on the thermo. limit

Theorem (Ambrosio and al., Barthe-Bordenave, Dereich and al., G-T, G-H-O)

For $p \geq 1$ and $L \gg 1$,

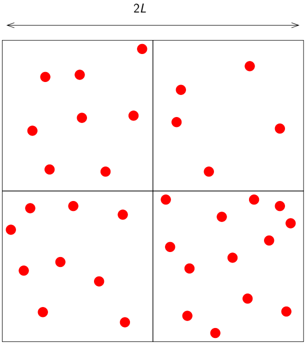
$$\mathbb{E}\left[\frac{1}{L^d} W^p_{[0,L]^d}(\mu, 1)\right] = \begin{cases} \frac{1}{2\pi}\log L + O(\log\log L) & \textit{if } d = p = 2 \\ f_{p,d} + o(1) & \textit{if } d \geq 3, p \geq 1. \end{cases}$$

Rk: the error bound $O(\log\log L)$ improves on Ambrosio-Glaudo by a factor $\log^{1/2} L$. Consequence of quantitative linearization down to microscale from GHO.

The case $d \geq 3$

We expect that to leading order cost is given by small scale behavior $\Longrightarrow$ ansatz by Caracciolo and al. cannot be directly used.
Idea: use subbaditivity. Indeed,
$W^p(\mu + \mu', \lambda + \lambda') \leq W^p(\mu, \lambda) + W^p(\mu', \lambda')$.

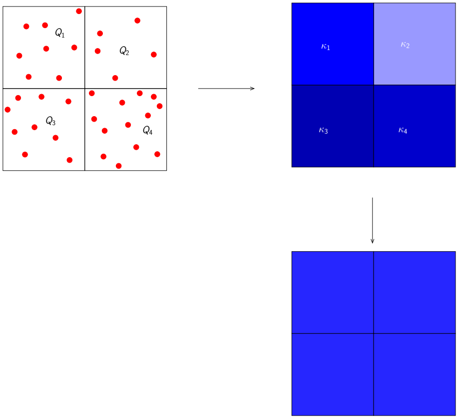
Restriction of μ from $[0, 2L]^d$ to $[0, L]^d$ very similar. Best seen by replacing deterministic number of points L^d by Poisson random variable N (grand canonical vs canonical)
Problem: In general not the same number of points in each subcube.
Solution: Relax the problem.



Idea: treat the defect in number of points as a local mass defect. Use first transport to send points to a locally constant distribution and then use for instance PDE to adjust the mass. Let $\kappa = \frac{\mu([0,L]^d)}{L^d}$ and

$$f_{p,d}(L) = \mathbb{E}\left[\frac{1}{L^d} W^p_{[0,L]^d}(\mu, \kappa)\right].$$

Divide $[0,2L]^d$ in 2^d cubes Q_i. Set $\kappa_i = \frac{\mu(Q_i)}{L^d}$.



By Δ inequality for W and subbaditivity of W^p, for every $\varepsilon \ll 1$

$$f_{p,d}(2L) \leq (1+\varepsilon) f_{p,d}(L) + \frac{C}{\varepsilon^{p-1}} \mathbb{E}\left[\frac{1}{L^d} W^p_{[0,2L]^d}\left(\sum_i \kappa_i \chi_{Q_i}, \kappa\right)\right]$$

The Global term: the Caracciolo and al. ansatz.

Recall

$$\|f\|^p_{W^{-1,p}} = \min_{\nabla \cdot j = f} \int |j|^p$$

and the Benamou-Brenier formula

$$W^p(\mu,\lambda) = \inf_{(\rho,j)} \left\{ \int \int_0^1 \frac{1}{\rho^{p-1}} |j|^p \ : \ \partial_t \rho + \nabla \cdot j = 0, \rho_0 = \mu, \rho_1 = \lambda \right\}.$$

If $\nabla \cdot j = \mu - 1$, set $\rho = (1-t)\mu + t \implies$

$$W^p(\mu,1) \lesssim \|\mu - 1\|^p_{W^{-1,p}} \left(\lesssim \mathrm{diam}(\Omega)^p \int |\mu - 1|^p \right).$$

PDE ansatz: take $j = \nabla \phi$ with $\Delta \phi = \mu - 1$.

The Global term

By previous slide:

$$W^p_{[0,2L]^d}\left(\sum_i \kappa_i \chi_{Q_i}, \kappa\right) \lesssim \|\sum_i \kappa_i \chi_{Q_i} - \kappa\|^p_{W^{-1,p}}$$
$$\lesssim L^p \sum_i \int_{Q_i} |\kappa_i - \kappa|^p \simeq L^{p+d} \sum_i |\kappa_i - \kappa|^p$$

Now $|\kappa_i - 1| \sim L^{-d/2}$ and thus

$$\mathbb{E}\left[\frac{1}{L^d} W^p_{[0,2L]^d}\left(\sum_i \kappa_i \chi_{Q_i}, \kappa\right)\right] \lesssim L^p L^{-\frac{pd}{2}} = L^{-\frac{p}{2}(d-2)}$$

Conclusion

All in all,

$$f_{p,d}(2L) \leq (1+\varepsilon) f_{p,d}(L) + \frac{C}{L^{\frac{p}{2}(d-2)} \varepsilon^{p-1}}.$$

Optimizing in ε yield

$$f_{p,d}(2L) \leq f_{p,d}(L) + \frac{C}{L^{\frac{d-2}{2}}}$$

from which $\lim_{L\to\infty} f_{p,d}(L)$ exists.

On the case of non-uniform densities

Ω connected bounded Lipschitz domain.

Theorem (Ambrosio-G-Trevisan)

For every Hölder continuous probability density ρ bounded above and below on Ω, if X_i are iid distributed according to ρ,

$$\lim_{n\to\infty} \frac{n}{\log n} \mathbb{E}\left[W_2^2 \left(\frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}, \rho \right) \right] = \frac{|\Omega|}{4\pi}$$

This answers a conjecture of Benedetto and Cagliotti (which proved the upper bound when $\Omega = (0,1)^2$.

Strategy of proof: Upper bound

Use subbaditivity and the result of Ambrosio-Stra-Trevisan for $(0,1)^2$.
Difficulty: in general cannot divide Ω in a finite number of cubes.
Solution: use a Whitney partition which refines close to the boundary. This forces a much more careful treatment of the 'global term' ($\|f\|_{W^{-1,p}} \lesssim \|f\|_{L^p}$ is not enough).

Lower bound

Consider a superaditive quantity Wb (see Figalli-Gigli) and show that for $\mu = \sum_{i=1}^{L^d} \delta_{X_i}$ with X_i iid uniformly distributed in $(0, L)^d$

$$\lim_{L\to\infty} \frac{1}{\log L} \mathbb{E}\left[\frac{1}{L^2} W^2_{[0,L]^2}(\mu, 1)\right]$$
$$= \lim_{L\to\infty} \frac{1}{\log L} \mathbb{E}\left[\frac{1}{L^2} Wb^2_{[0,L]^2}(\mu, 1)\right] = \frac{1}{2\pi}.$$

Rk: other case when Dirichlet cost=Neumann cost: $d = 1$, $p \in (0, 1/2)$, see G-Trevisan. General case $p \geq 1$, $d \geq 3$: open since Barthe-Bordenave.

Extension to more general bipartite combinatorial problems

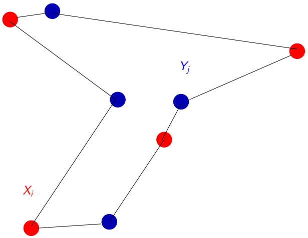
The previous analysis strongly relied on the connection between matching and optimal transport. What about the case of more general combinatorial problems (TSP, MST...)?

- The case $p < d/2$ has been treated by Barthe-Bordenave using 'soft' subbaditivity arguments.
- For non-bipartite problems the result is only known for $p < d$.

An example, the bipartite TSP

Set $\mathbf{x} = (X_i)_{i=1}^n$ and $\mathbf{y} = (Y_i)_{i=1}^n$.
For $p \geq 1$, the TSP is

$$\mathrm{TSP}^p(\mathbf{x}, \mathbf{y}) = \min_{\sigma,\tau} \sum_{i=1}^{n} |X_{\sigma(i)} - Y_{\tau(i)}|^p + |X_{\sigma(i)} - Y_{\tau(i+1)}|^p.$$



Can extend it for $|\mathbf{x}| \neq |\mathbf{y}|$ by requesting that the tour contains as many points as possible.

Let Ω be a bounded connected domain either smooth or convex and ρ Hölder continuous probability density ρ bounded above and below on Ω.

Theorem (G-Trevisan)

For every $p \in [1, d)$ there exists $\beta_p \in (0, \infty)$ such that if X_i, Y_j are iid distributed according to ρ,

$$\lim_{n\to\infty} n^{\frac{p}{d}-1}\mathbb{E}\left[\mathsf{TSP}^p(\mathbf{x}, \mathbf{y})\right] \leq \beta_p \int_\Omega \rho^{1-\frac{p}{d}}.$$

Moreover if ρ is constant then equality holds.

As in B-B, AGT, etc... first proved for $\Omega = (0, 1)^d$ and $\rho = 1$, then use it as a building block for the general case. As in AGT, we need a Whitney partition in this second step.

Sketch of proof

Both steps rely on the combination of

- Subbaditivity: If $\Omega = \cup_k \Omega_k$ and for each k, $\mathbf{x}_k, \mathbf{y}_k \subset \Omega_k$ with $|\mathbf{x}_k| = |\mathbf{y}_k|$ and $\mathbf{x}_0, \mathbf{y}_0 \in \Omega$ then with $\mathbf{x} = \mathbf{x}_0 \cup \bigcup_k \mathbf{x}_k$ and similarly for $\mathbf{y}$,

$$\mathsf{TSP}^p(\mathbf{x},\mathbf{y}) \leq \sum_k \mathsf{TSP}^p(\mathbf{x}_k,\mathbf{y}_k) + C\mathsf{TSP}^p(\mathbf{x}_0,\mathbf{y}_0) + lot.$$

- Growth: For every $\mathbf{x}, \mathbf{y} \in \Omega$,

$$\mathsf{TSP}^p(\mathbf{x},\mathbf{y}) \lesssim \operatorname{diam}(\Omega)^p \min\left\{|\mathbf{x}|^{1-\frac{p}{d}}, |\mathbf{y}|^{1-\frac{p}{d}}\right\} + \mathsf{M}^p(\mathbf{x},\mathbf{y}),$$

where M^p is the matching cost.

Growth:

Follows from (assume $|\mathbf{x}| \leq |\mathbf{y}|$)

$$\mathsf{TSP}^p(\mathbf{x}, \mathbf{y}) \lesssim \mathsf{TSP}^p(\mathbf{x}) + \mathsf{M}^p(\mathbf{x}, \mathbf{y}),$$

which is proven in Capelli and al. and

$$\mathsf{TSP}^p(\mathbf{x}) \lesssim \mathrm{diam}(\Omega)^p |\mathbf{x}|^{1-\frac{p}{d}}.$$

Given Subbaditivity +Growth we write $\mathbf{x} = \mathbf{x}^{1-\eta} \cup \mathbf{x}^{\eta}$ where $|\mathbf{x}^{1-\eta}| = (1-\eta)n$ (same for $\mathbf{y}$).

- Let $(\mathbf{x}_k, \mathbf{y}_k)$ be minimizers of $\mathrm{TSP}^p(\mathbf{x}^{1-\eta} \cap \Omega_k, \mathbf{y}^{1-\eta} \cap \Omega_k)$
- Set $\mathcal{U} = \mathbf{x}^{1-\eta} \setminus (\cup_k \mathbf{x}_k)$, $\mathcal{V} = \mathbf{y}^{1-\eta} \setminus (\cup_k \mathbf{y}_k)$
- and $\mathbf{x}_0 = \mathbf{x}^{\eta} \cup \mathcal{U}$, $\mathbf{y}_0 = \mathbf{y}^{\eta} \cup \mathcal{V}$.

We get

$$\begin{aligned}\mathsf{TSP}^p(\mathbf{x},\mathbf{y}) \le \sum_k \mathsf{TSP}^p(\mathbf{x}_k,\mathbf{y}_k) \\ + C\left(|\mathbf{x}^\eta|^{1-\frac{p}{d}} + |\mathcal{U}|^{1-\frac{p}{d}} + \mathsf{M}^p(\mathbf{x}^\eta \cup \mathcal{U}, \mathbf{y}^\eta \cup \mathcal{V})\right).\end{aligned}$$

Notice that $|\mathbf{x}^\eta|^{1-\frac{p}{d}} = (\eta n)^{1-\frac{p}{d}}$ and $|\mathcal{U}|^{1-\frac{p}{d}} \lesssim n^{\frac{1}{2}(1-\frac{p}{d})}$. Thus the term in parenthesis is small provided

$$\mathbb{E}[\mathsf{M}^p(\mathbf{x}^\eta \cup \mathcal{U}, \mathbf{y}^\eta \cup \mathcal{V})] \lesssim (\eta n)^{1-\frac{p}{d}} + C_\eta n^{\frac{1}{2}(1-\frac{p}{d})}.$$

Since $\mathbf{x}_\eta$ are iid points but $\mathcal{U}$ are not, this requires extending bounds for the matching to the case where most but not all points are iid.

Notice that the proof extends to all combinatorial problems satisfying Subbaditivity $+$Growth. This covers essentially all the examples from B-B.

Thank you for your attention and happy birthday Sergey.