The Abel Symposium

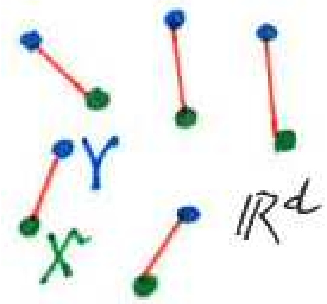
A variational regularity theory for optimal transportation and its application to matching

Felix Otto,
Max Planck Institute for Mathematics in the Sciences,
Leipzig, Germany

joint work with Michael Goldman (Ann. ENS '20),
with MG & Martin Huesmann (CPAM '23),
with MH & Francesco Mattesini (submitted PTRF).

version June 16th 2023

Optimal matching of random point clouds



Matching of two locally finite point clouds $\{X\}$ and $\{Y\} \subset \mathbb{R}^d$, amounts to a pairing $\{(X, Y)\} \subset \mathbb{R}^d \times \mathbb{R}^d$.

Optimality means cyclical monotonicity

$\forall$ matched finite subset
$\{X_1, \cdots, X_N\}$
$\{Y_1, \cdots, Y_N = Y_0\}$.

$$\Sigma_{n=1}^N |Y_n - X_n|^2$$
$$\leq \Sigma_{n=1}^N |Y_{n-1} - X_n|^2$$
$$\iff \Sigma_{n=1}^N (Y_n - Y_{n-1}) \cdot X_n \geq 0.$$

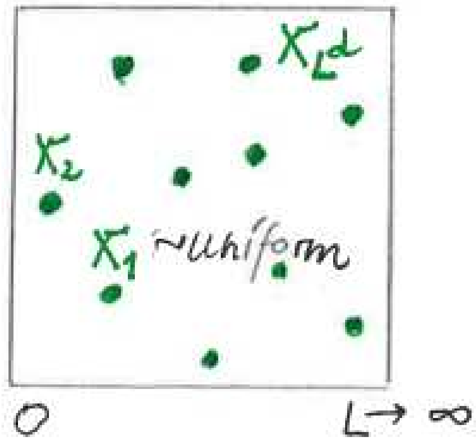
Simplest *random* setting:

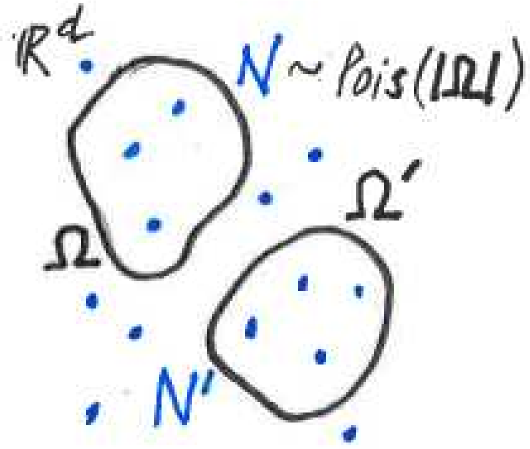
$\{X\}$, $\{Y\}$ indep. samples of **Poisson point processes**.

The Poisson point process

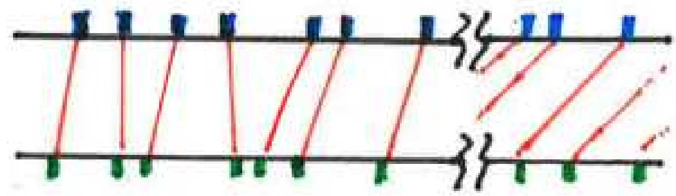
Locally finite point cloud via Poisson point process of unit intensity (means that distance between points $O(1)$)

canonical vs.
grand-canonical
definition



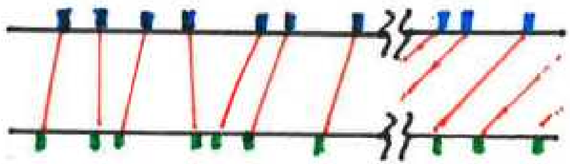


Seek cyclically monotone matching of two independent Poisson point processes
– mean distance diverges in $d = 1$.

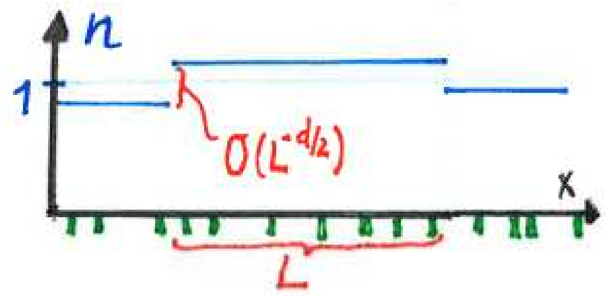


Matching depends on dimension d ...

Cyclically monotone matching of two independent Poisson point processes – distances diverge like square-root for $d = 1$.

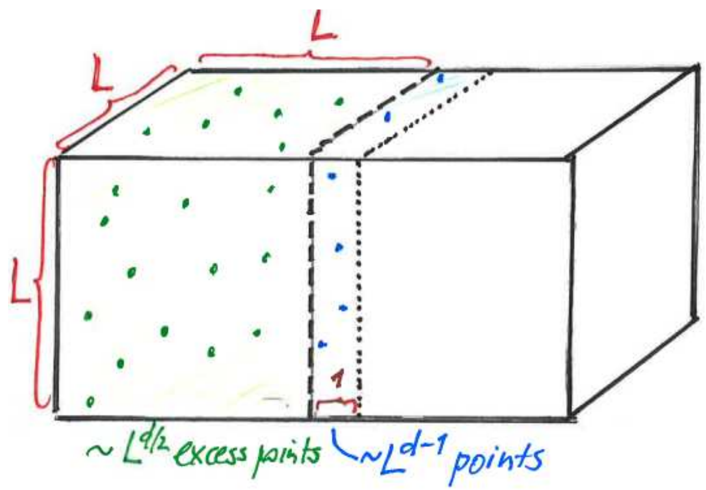


Cause of divergence:
mesoscopic fluctuations
of number density n
$= O(\frac{1}{\sqrt{L^d}})$; lower for higher d.



Number of excess points $= O(\sqrt{L^d})$,

number of points in
(width one) boundary layer $= O(L^{d-1})$.

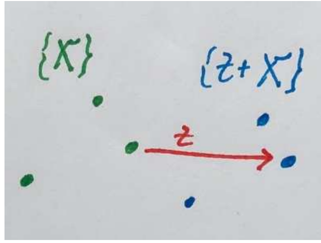


... critical dimension $d = 2$

Impose statistical translation invariance ("Stationarity") of matching

Cyclically monotone matching of
two independent Poisson point processes $\{X\}$, $\{Y\}$.

Poisson point process is *stationary*:
$\forall$ shift vectors $z \in \mathbb{R}^d$ $\quad \{z + X\} =_{\text{law}} \{X\}$.



Seek random point cloud $\{(X, Y)\}$ in $\mathbb{R}^d \times \mathbb{R}^d$ s. t.
marginals are independent Poisson point processes,
coupling is cyclically monotone almost surely,
and $\forall\ z \in \mathbb{R}^d \quad \{(z + X, z + Y)\} =_{\text{law}} \{(X, Y)\}$.

Critical dimension $d = 2$ rigorously captured

Interest in Combinatorics (eg. Ajtai et al. '84),
Probability Theory (Talagrand '92+, Holroyd-Peres '11+),
Physics (eg. Parisi et al. '14),
Analysis (eg. Ambrosio et al. '16+).

Seek random point cloud $\{(X, Y)\}$ in $\mathbb{R}^d \times \mathbb{R}^d$ s. t.
marginals are independent Poisson point processes,
coupling is cyclically monotone almost surely,
and $\forall\, z \in \mathbb{R}^d \quad \{(z + X, z + Y)\} =_{\text{law}} \{(X, Y)\}$.

Theorem (Huesmann&Sturm '13)

For $d > 2$ have existence.

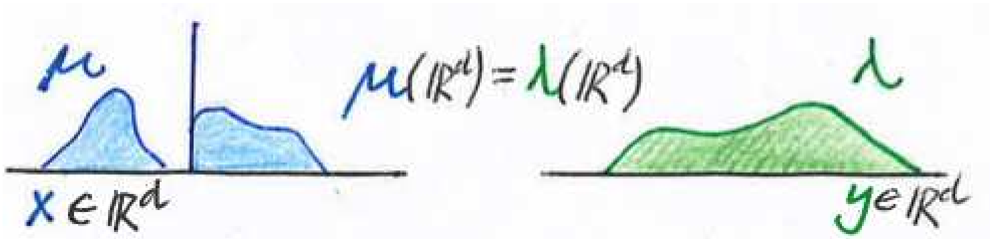
Theorem (H.&Mattesini&O. '21)

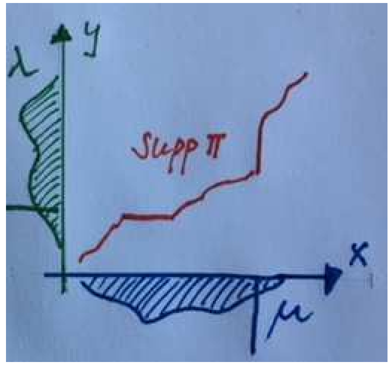
For $d \leq 2$ have non-existence.

Proof via continuum/PDE perspective

Optimal Transportation in Kantorowicz' formulation

Given two measures





seek transfer plan π, i. e. $\pi(U \times \mathbb{R}^d) = \mu(U)$, $\pi(\mathbb{R}^d \times V) = \lambda(V)$

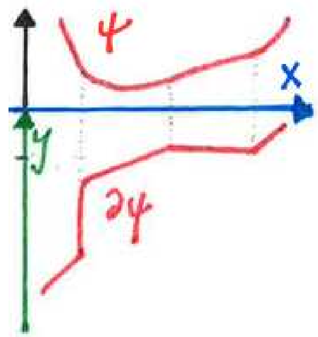
that minimizes Euclidean transport cost $\displaystyle\int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^2 \pi(dxdy)$.

Minimum $=: W_2^2(\mu, \lambda)$ (squared) Wasserstein distance.

From optimal transportation to Monge-Ampère

Minimize $\int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^2 \pi(dxdy)$
among all $\pi(dxdy)$ with marginals $\mu(dx)$ and dy.

Support of optimal transfer plan π
is cyclically monotone;
hence $\exists$ convex ψ
supp$\pi \subset \{ (x, y) \mid y \in$ sub-gradient $\partial\psi(x) \}$.



$\forall$ test functions ζ $\quad \int \zeta(\nabla\psi(x))\mu(dx) = \int \zeta(y)dy$.

In smooth case, this amounts to $\det D^2\psi = \mu$,
an instance of the Monge-Ampère equation.

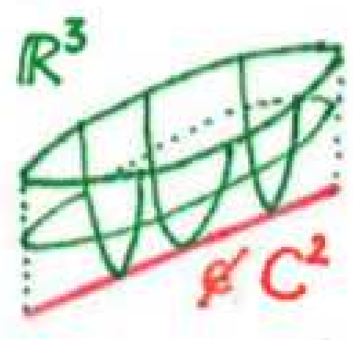
Nature of the Monge-Ampère equation

Recall Monge-Ampère: $\det D^2\psi = 1$.
Fully non-linear with $F(A) := \det A - 1$.

However elliptic: $F(A) > F(A')$ for $A > A' \geq 0$;
satisfies comparison principle.

However degenerate: $\leftrightarrow$ affine invariant (non-compact SL(d)).
Cf. Laplacian $F(A) = \mathrm{tr} A$: rotation invariant (compact SO(d)).

Caffarelli's '90 breakthrough:
comparison principle,
affine invariance,
compactness.



Pogorelov's example is worst case

Monge-Ampère equation at crossroads
of fully nonlinear and variational.

Parisi's heuristics for semi-discrete matching

$\mu = \sum_{x\in X}\delta_x$ Poisson, $\lambda = dy$ Lebesgue

$\det D^2\psi\text{-}1 = \mu\text{-}1 \approx_{\text{law}} \xi$ on scales $R \gg 1$, $\xi =$ white noise,
meaning $\int \eta_R(\mu\text{-}1) \approx_{\text{law}} \int \eta_R \xi$, with $\eta_R(x) = \frac{1}{R^d}\eta(\frac{x}{R})$, $\eta \in C_0^\infty$ fixed.

Consider $\triangle\phi = \xi$ (so that $\nabla\phi$ behaves as Gaussian free field)
$\xi \sim R^{-\frac{d}{2}} \ll 1$ on scales $R \gg 1 \Longrightarrow D^2\phi \ll 1$ on scales $\gg 1 \Longrightarrow$
$\det(D^2(\phi+\frac{1}{2}|x|^2))\text{-}1 = \det(D^2\phi+\text{id})\text{-}1 \approx \text{tr}D^2\phi = \xi$
on scales $\gg 1$.

Hence expect $\nabla\psi - x \approx_{\text{law}} \nabla\phi$ on scales $\gg 1$;
in terms of "displacement" $(y\text{-}x)\pi(dxdy)$:

$$\frac{\int \eta_R(x)(y\text{-}x)\pi(dxdy)}{\int \eta_R(x)\pi(dxdy)} \approx_{\text{law}} \int \eta_R \nabla\phi.$$

Quantitative large-scale linearization

Compare: $\pi(dxdy)$ optimal for $\mu(dx)$ and dy
with $\quad -\triangle\phi = \mu - 1$.

Rate function $D(R)\uparrow$, $\frac{D(R)}{R}\downarrow$
in strengthened sense of Dini $\sum_{k=0}^{\infty}\frac{D(2^kR)}{2^kR} \le 4\frac{D(R)}{R}$.

Theorem 1 (Goldman&Huesmann&O.) Provided
$\frac{1}{|B_R|}W_{B_R}^2(\mu,\kappa_R)+R^2(\kappa_R\text{-}1)^2 \le D(R)$ for all $R \ge 1$,
then $\left|\frac{\int\eta_R(x)(y-x)\pi(dxdy)}{\int\eta_R(x)\pi(dxdy)} - \int\eta_R\nabla\phi\right| \lesssim \frac{D(R)}{R}$ for all $R \ge 1$.

Confirms (deterministic part of) Parisi's heuristics.

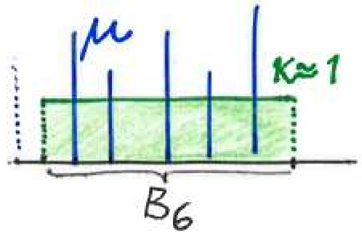
Relies on large-scale regularity theory.

Large-scale regularity via harmonic approximation

Consider $\pi(dxdy)$ optimal for $\mu(dx)$ and $\lambda(dy)$.

Local energy $\quad E := \int_{(B_6\times\mathbb{R}^d)\cup(\mathbb{R}^d\times B_6)} |y-x|^2\pi(dxdy)$,

Local data size^2 $\quad D := W^2_{B_6}(\mu, \kappa{=}\mathrm{const}) + (\kappa\text{-}1)^2$
$\quad + \text{same for } \lambda$



Theorem 2 (Goldman&Huesmann&O.)

$\forall\ \tau > 0 \quad \exists\ \epsilon(\tau,d) > 0,\ C(\tau,d) < \infty \quad$ s. t. $\quad E+D \le \epsilon \implies$

$\exists\ \nabla\phi$ harmonic, $\quad \int_{B_1} |\nabla\phi|^2 \le C(E+D)$,

$$\int_{(B_1\times\mathbb{R}^d)\cup(\mathbb{R}^d\times B_1)} \left|y-x-\nabla\phi(y)\right|^2\pi(dxdy) \le \tau E + CD.$$

Harmonic approximation: correct homogeneities ...

$E := \int_{(B_6\times\mathbb{R}^d)\cup(\mathbb{R}^d\times B_6)} |y-x|^2\pi(dxdy)$, quadratic in solution,

$D := W^2_{B_6}(\mu, \kappa{=}\text{const}) + (\kappa-1)^2 + \text{same for } \lambda$, quadratic in data.

Theorem 2

$\forall\ \tau > 0 \quad \exists\ \epsilon(\tau,d) > 0,\ C(\tau,d) < \infty \quad$ s. t. $\quad E + D \le \epsilon \implies$

$\exists\ \nabla\phi$ harmonic, $\quad \int_{B_1} |\nabla\phi|^2 \le C(E+D)$,

$$\int_{(B_1\times\mathbb{R}^d)\cup(\mathbb{R}^d\times B_1)} \Big|y - x - \nabla\phi(y)\Big|^2\pi(dxdy) \le \tau E + C D.$$

Compare to $\int_{B_1} L(\nabla u - \nabla\phi) \le \tau \int_{B_6} L(\nabla u) + C\int_{B_6} |f|^2$

for $\ -\nabla\cdot DL(\nabla u) = \nabla\cdot f\ $ with uniformly convex L .

... and correct metric

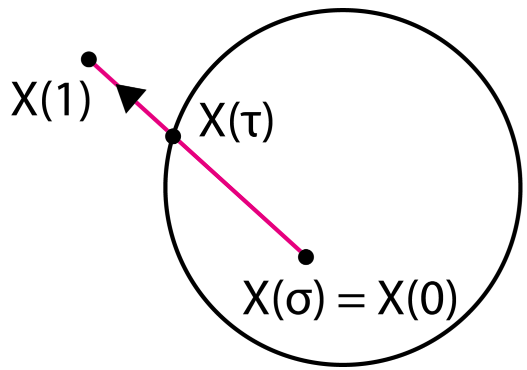
Definition of $\nabla\phi$ via Neumann-Poisson problem

Consider trajectories $[0,1] \ni t \mapsto X(t) := ty + (1-t)x$.

Consider when & where trajectories $\binom{\text{enter}}{\text{exit}}$ B_R:

$\binom{\sigma}{\tau} := \binom{\min}{\max}\{\, t \in [0,1] \,|\, X(t) \in \bar{B}_R \,\}$,

$$\int \zeta dg = \int \zeta(X(\sigma))\pi(dxdy),$$
$$\int \zeta df = \int \zeta(X(\tau))\pi(dxdy).$$



Define $\nabla\phi$ by the Neumann-Poisson problem:

$-\triangle\phi = \frac{(\lambda-\mu)(B_R)}{|B_R|}$ in B_R and $\nu \cdot \nabla\phi = f - g$ on ∂B_R.

Optimize in $R \in [2,3]$.

A numerical illustration in case of matching

Matching of $\{X\}$ and $\{Y\}$; consider a square Q

Define $\nabla\phi$ by the Neumann-Poisson problem:

$$-\Delta\phi = \Sigma_{Y\in Q}\,\delta_Y - \Sigma_{X\in Q}\,\delta_X \quad \text{in } Q,$$

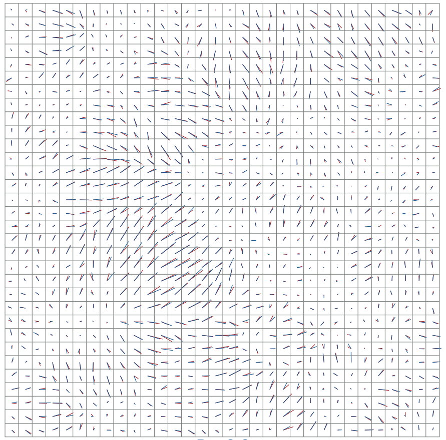
$$\nu\cdot\nabla\phi = \Sigma_{Y\in Q, X\notin Q}\,\delta_Z - \Sigma_{Y\notin Q, X\in Q}\,\delta_Z \quad \text{on } \partial Q.$$

A numerical illustration in case of matching

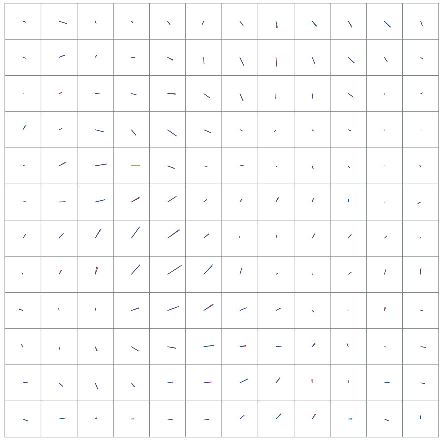
Recall $-\triangle\phi = \sum_{Y\in Q}\delta_Y - \sum_{X\in Q}\delta_X$ in Q,

$\nu\cdot\nabla\phi = \sum_{Y\in Q, X\notin Q}\delta_Z - \sum_{Y\notin Q, X\in Q}\delta_Z$ on ∂Q.

Check $\nabla\phi \approx Y - X$ when averaged on scale R;



R = 0.8



R = 2.2

R larger – better agreement; side-length of $Q = 75$.

Finite Element discretization, simulations by R. Kriemann

Analogies to minimal surfaces (Schoen&Simon '82)

Approximate minimal surface by harmonic graph /
approximate displacement by harmonic gradient.

Use: Object is minimizing under compact perturbations.
Don't use: Euler-Lagrange equation (= first variation).

Mismatch of type of boundary condition
for construction of harmonic competitor:
graph vs. non-graph / flux vs. displacement;
choice of good radius.

Use of strict convexity to convert energy gap
into distance ("approximate orthogonality");
need to smooth out boundary data.

Summary and outlook

Variational regularity theory for optimal transportation,
mimics ϵ-regularity theory for minimal surfaces,
is more robust than maximum principle-based approach,
provides large-scale regularity theory for matching.

Extend non-existence result and regularity theory
from $p=2$ to $p\in(1,\infty)$ (& L. Koch),
and to entropic regularization (& R. Gvalani).

Seek quantitative coupling of shot noise $\mu-1$ with
white noise ξ in $d>2$ (Komlós&Major&Tusnády '75).