Summer School on Optimal Transport

A variational regularity theory for optimal transportation

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

joint work with Michael Goldman (arXiv '17, Ann. ENS '20),
with MG & Martin Huesmann (to appear in CPAM),
with MH & Francesco Mattesini (arXiv)

version June 24st 2022

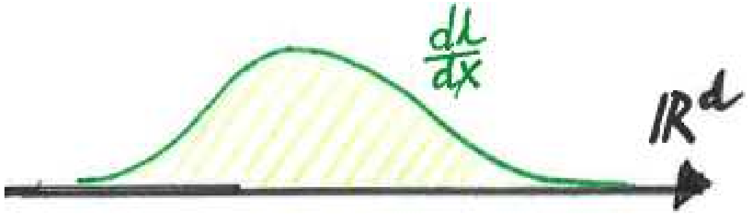
An application to the matching problem of our variational regularity theory for optimal transportation

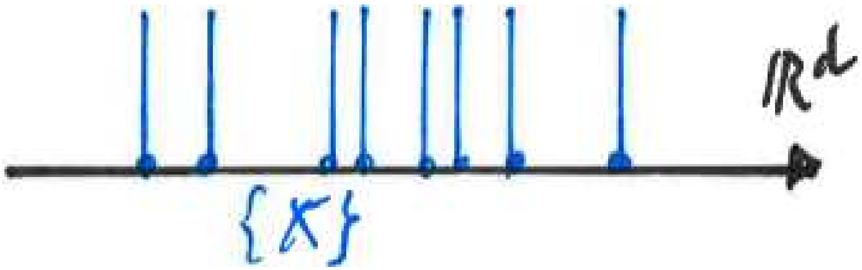
A natural application for optimal transportation ...

Given λ probability measure/law on $\mathbb{R}^d$,

draw N independent samples $X_1, \cdots, X_N$.

Consider empirical measure $\mu = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n}$.

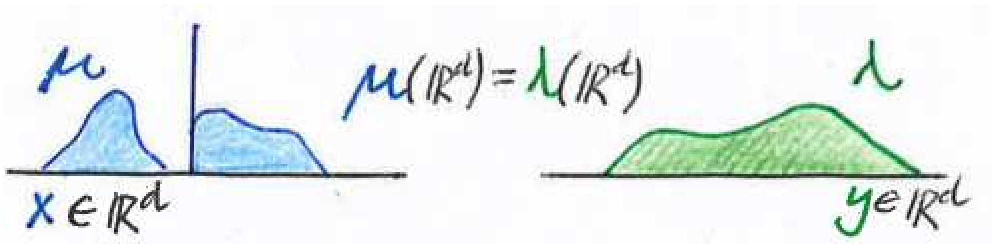


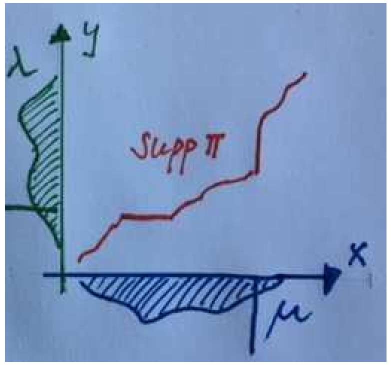


How close are λ and μ?

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 $\int_{\mathbb{R}^d\times\mathbb{R}^d}|y-x|^2\pi(dxdy)$;

$=$ coupling π that maximizes covariance.

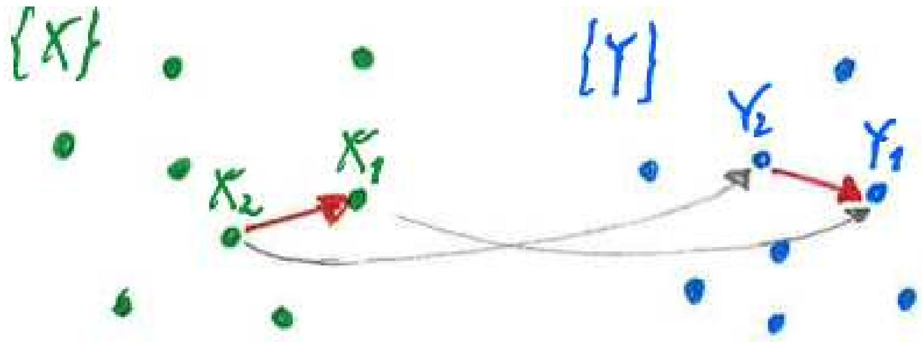
Minimum $=: W_2^2(\mu,\lambda)$ (squared) Wasserstein distance.

Optimal matching = Optimal transport

Two independent sets $\{X\}$, $\{Y\}$ of samples from λ.

Wasserstein distance between $\sum_{\{X\}} \delta_X$ and $\sum_{\{Y\}} \delta_Y$.

optimal transport
=optimal matching

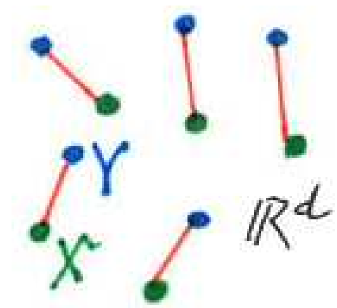


For the sake of discussion,
we also consider $|y-x|^p$ for $p \in (0,\infty)$ next to $|y-x|^2$.

Optimality for matching of infinite point clouds

Focus on mesoscopic behavior: $\lambda =$ Lebesgue $\Longrightarrow$ $\{X\}$, $\{Y\}$ indep. samples of Poisson point processes.

Matching of two infinite 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:

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

$$\sum_{n=1}^{N} |Y_n - X_n|^2 \leq \sum_{n=1}^{N} |Y_{n-1} - X_n|^2$$

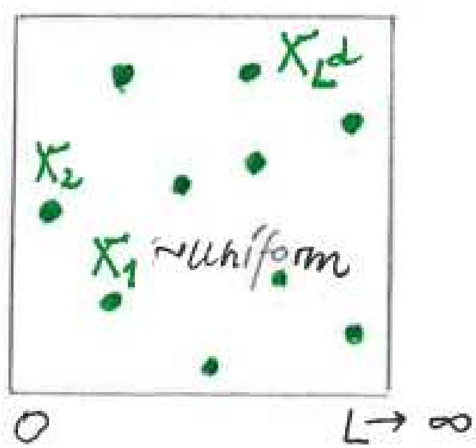
The Poisson point process

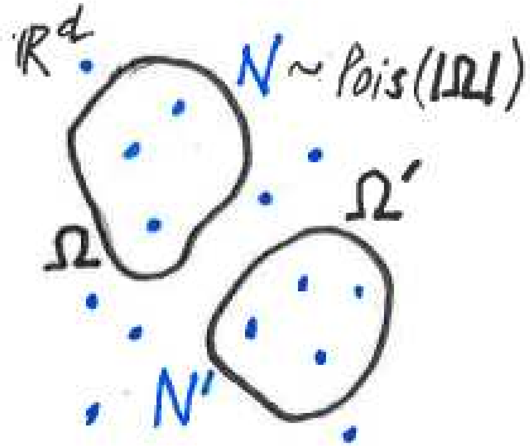
Locally finite point cloud via Poisson point process of unit intensity (means that distance between points $O(1)$)

canonical vs.

grand-canonical

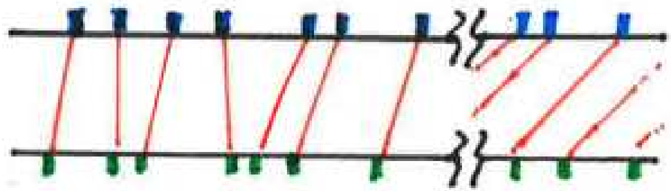
definition





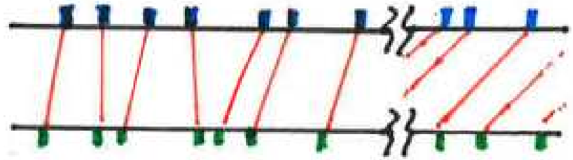
Seek optimal matching of two independent Poisson point processes – divergent behavior for $d = 1$ and $p > 1$

(implies monotonicity)

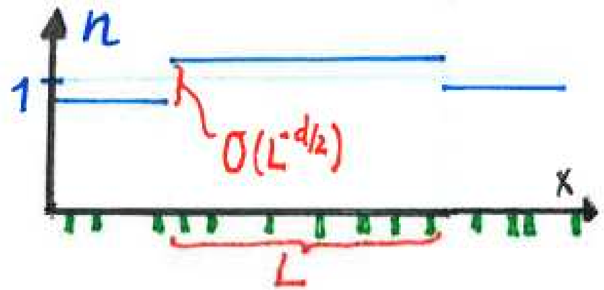


Matching depends on dimension d ...

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

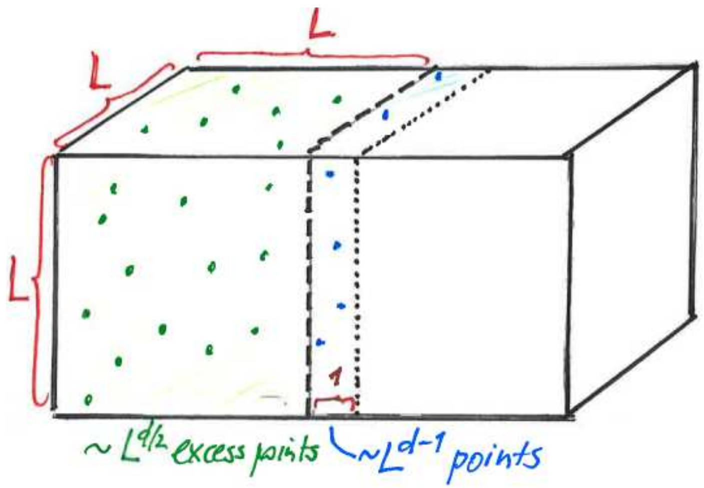


Fluctuations of number density n
$= O(L^{-\frac{d}{2}})$; lower for higher d.



Number of excess points $= O(L^{\frac{d}{2}})$,

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

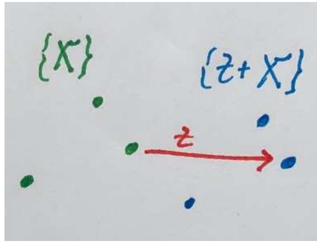


... critical dimension $d = 2$

Impose statistical translation invariance ("Stationarity") of matching

Optimal 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 optimal 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)

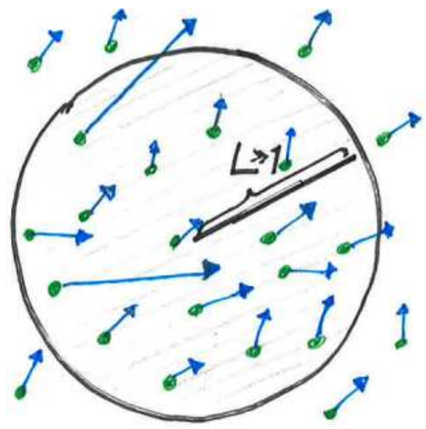
For $d \leq 2$ have non-existence.

Mesoscopic analysis of the matching problem via harmonic approximation of displacement, arising from our variational regularity theory

Indirect argument for non-existence in $d = 2$

Ergodicity: Most particles in B_L are moved at most $O(1)$.

Monotonicity: All particles in B_L are moved at most $o(L)$.



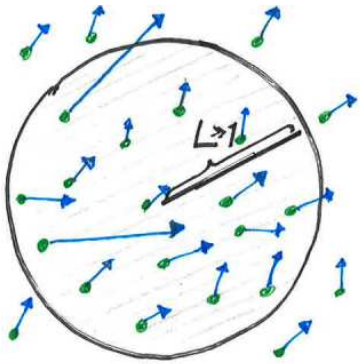
Local energy $E := L^{-d}\sum_{X\in B_L}|Y-X|^2 = o(L^2)$ is non-dimensionally small.

Local data $D := L^{-d}W_2^2(\sum_{X\in B_L}\delta_X, \text{uniform on } B_L) = O(\ln L)$ is non-dimensionally **much** smaller (same for $\sum_{Y\in B_L}\delta_Y$).

... **want to transmit smallness of D to E**

From D to E by harmonic approximation

Recall local energy and data size:



$$E := L^{-d}\sum_{X\in B_L}|Y\text{-}X|^2 = o(L^2),$$

$$D := L^{-d}W_2^2(\sum_{X\in B_L}\delta_X, \text{uniform on } B_L) = O(\ln L).$$

Harmonic approximation: $\exists\ \phi$ harmonic s. t.

$$L^{-d}\sum_{X\in B_L}|(Y\text{-}X)-\nabla\phi(X)|^2 \le o(E)+O(D).$$

By $\sup_{B_L}|\nabla\phi|^2 \lesssim L^{-d}\int_{B_{2L}}|\nabla\phi|^2$ and iteration $L\uparrow\infty$:

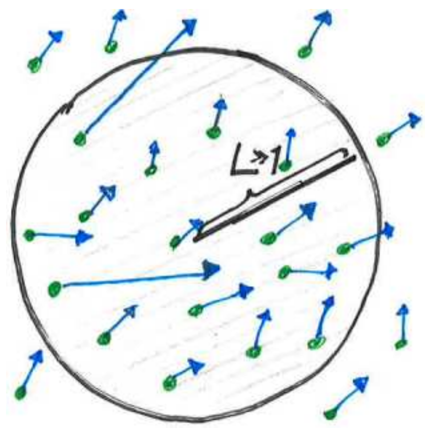
$$E = L^{-d}\sum_{X\in B_L}|Y\text{-}X|^2 \le O(\ln L).$$

... smallness of D does transmit to E

The final contradiction

Most particles in B_L
are moved at most $O(1)$,

Mean *square*
transportation distance in B_L
at most $O(\ln L)$.



Combines to:
Mean transportation distance in B_L at most $o(\ln^{\frac{1}{2}} L)$.

On the other hand:
Mean transportation distance in B_L at least $O(\ln^{\frac{1}{2}} L)$,
by $W_1(\sum_{X\in B_L}\delta_X, \text{uniform in } B_L) \gtrsim \ln^{\frac{1}{2}} L$.

Why is the result not trivial?

Huesmann-Sturm existence applies to general p;
have existence provided (morally speaking)

$$\lim_{R\uparrow\infty}\frac{1}{|B_R|}\mathbb{E}W_p^p(\mu\llcorner B_R,\kappa dx\llcorner B_R)<\infty,$$

i. e. the transportation cost per particle stays finite.

Our non-existence applies to $d=2$, $p=2$
($p\in(1,\infty)$ in preparation with L. Koch) and uses

$$\lim_{R\uparrow\infty}\frac{1}{|B_R|\ln^{\frac{1}{2}}R}\mathbb{E}W_1(\mu\llcorner B_R,\kappa dx\llcorner B_R)>0.$$

Holroyd-Pemantle-Peres-Schramm '09 for $d=1$, $p=\frac{1}{2}$
established existence of a stationary matching;
any stationary matching satisfies

$$\lim_{R\uparrow\infty}\frac{1}{|B_R|}\mathbb{E}W_{\frac{1}{2}}^{\frac{1}{2}}(\mu\llcorner B_R,\kappa dx\llcorner B_R)=\infty.$$

A variational approach
to the regularity theory
for optimal transportation;
At its core: harmonic approximation

From optimal transportation to Monge-Ampère

Minimize $\int_{\mathbb{R}^d \times \mathbb{R}^d} |y-x|^2 \pi(dxdy)$

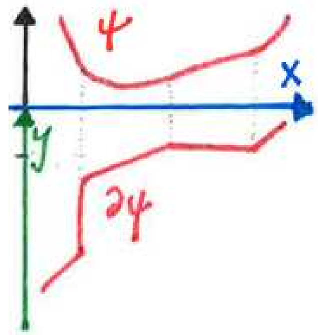
among all $\pi(dxdy)$ with marginals $\mu(dx)$ and dy.

Support of optimal transfer plan π

is cyclically monotone;

hence $\exists$ convex ψ

supp$\pi \subset \{\,(x,y)\,|\,y \in \text{sub-gradient } \partial\psi(x)\,\}$.



$\forall$ 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.

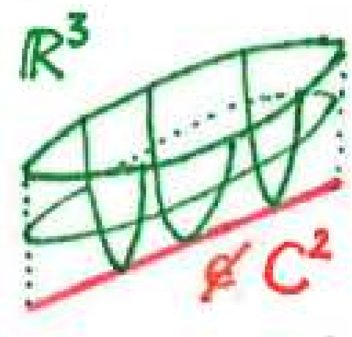
Nature of the Monge-Ampère equation

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.

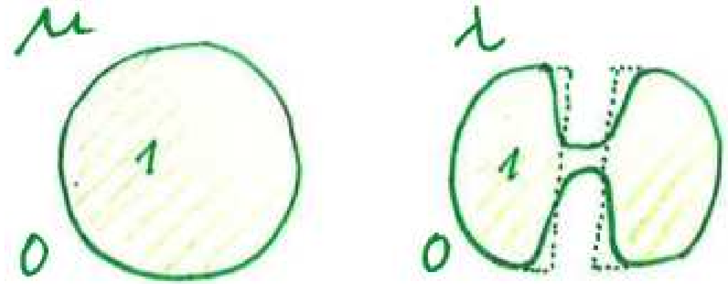


Pogorelov's example is worst case

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

Singularities in optimal transportation are generic

Caffarelli's example:



smooth data μ, λ
do not yield smooth $T = \nabla\psi$.

See also Loeper's example in Riemannian setting.

Thus ϵ-regularity is of interest (Figalli-Kim, DePhilippis-Figalli):

$\int |y\text{-}x|^2\pi(dxdy) \leq \epsilon$ locally, μ, λ smooth locally

$\Longrightarrow$ Kantorowicz potential ψ smooth locally.

Relies on harmonic approximation in our approach:

$\int |y\text{-}x|^2\pi(dxdy) \leq \epsilon$ locally, $\mu, \lambda \approx 1$ locally

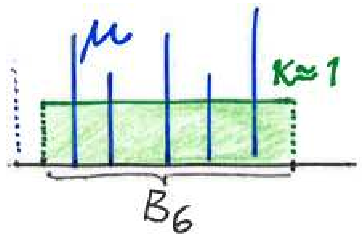
$\Longrightarrow$ displacement $(y\text{-}x)\pi(dxdy) \approx \nabla$harmonic locally.

Statement of our harmonic approximation result

Local energy $\quad E := \int_{(B_3\times\mathbb{R}^d)\cup(\mathbb{R}^d\times B_3)} |y-x|^2\pi(dxdy),$

Local data[2] $D := W_2^2(\mu\llcorner B_3, \kappa dx\llcorner B_3) + (\kappa\text{-}1)^2$

$+$ same for λ

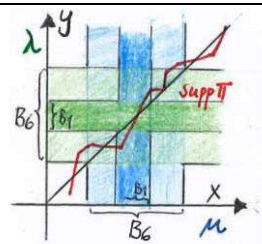


> **Proposition 1** (Goldman&Huesmann&O.)
>
> $\forall\ \theta>0 \quad \exists\ \epsilon(\theta,d)>0,\ C(\theta,d)<\infty \quad$ s. t. $\quad E+D\le\epsilon \Longrightarrow$
>
> $\exists\ \nabla\phi$ harmonic, $\quad \int_{B_1}|\nabla\phi|^2 \le C(E+D),$
>
> $$\int_{(B_1\times\mathbb{R}^d)\cup(\mathbb{R}^d\times B_1)} \big|y-x-\nabla\phi(x)\big|^2\pi(dxdy) \le \theta E + CD.$$

Amounts to:

Displacement $\quad y-x$

$\approx$ harmonic gradient $\quad \nabla\phi$



Plan for mini-course

Motivate connection between OT
and the Poisson equation

Elucidate the main ideas behind Proposition 1

Harmonic approximation: correct homogeneities ...

$E := \int_{(B_3\times\mathbb{R}^d)\cup(\mathbb{R}^d\times B_3)} |y-x|^2\pi(dxdy)$, quadratic in solution,

$D := W_2(\mu\llcorner B_3, \kappa_\mu dy\llcorner B_3) + (\kappa_\mu - 1)^2 +$ same for λ, quadratic in data.

Proposition 1

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

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

$$\int_{\{\exists t\in[0,1]\ X(t)\in\bar{B}_1\}} \int_\sigma^\tau \left|\dot{X}(t) - \nabla\phi(X(t))\right|^2 dt\,\pi(dxdy)$$

$$\le \theta E + C D.$$

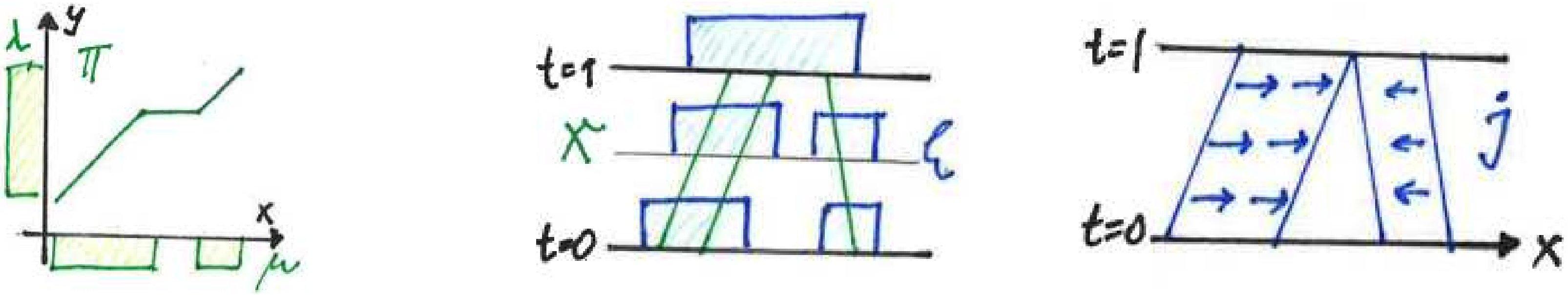
Compare to $\int_{B_1} L(\nabla u - \nabla\phi) \le \tau \int_{B_6} L(\nabla u) + C\int_{B_6} |f|^2$

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

... and correct metric

From Lagrangian to Eulerian (Benamou-Brenier)

Transport plan π, trajectories $X(t)$, density/flux (ρ, j)



Continuity eqn. $\partial_t \rho + \nabla \cdot j = 0$, kinetic energy $\int_{\mathbb{R}^d \times (0,1)} \frac{1}{\rho} |j|^2$.

$W^2(\mu, \lambda) = \inf\{ \int \frac{1}{\rho} |j|^2 \mid \partial_t \rho + \nabla \cdot j = 0,\ \rho(t{=}0){=}\lambda,\ \rho(t{=}1){=}\mu \}$

Kinetic energy density $(\rho, j) \mapsto \frac{1}{\rho} |j|^2$ is mostly strictly convex.

Linearization $\frac{1}{\rho} |j|^2 \rightsquigarrow |j|^2$ amounts to $W^2(\lambda, \mu) \rightsquigarrow \|\lambda - \mu\|^2_{\dot{H}^{-1}}$.

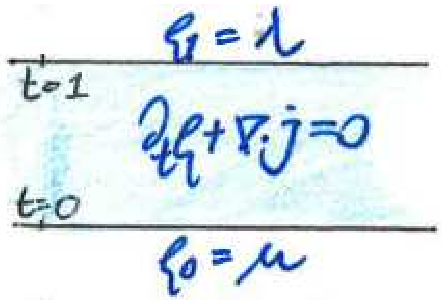
Analogy for minimal surfaces: varifolds vs. currents.

Eulerian version of harmonic approximation

Density/flux $(\rho, j) = (\rho_t dt, j_t dt)$ where

$\int \zeta d\rho_t = \int \zeta(ty+(1\text{-}t)x)\pi(dxdy), \ \int \xi \cdot dj_t = \int \xi(ty+(1\text{-}t)x)\cdot(y\text{-}x)\pi(dxdy)$

continuity equation



$$E \rightsquigarrow \int_{B_5\times(0,1)} \frac{1}{\rho}|j|^2$$

reveals strict convexity of variational problem

Proposition 1'

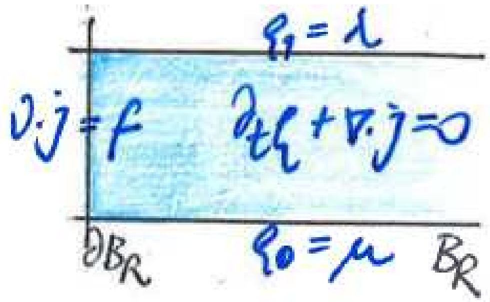
$\forall\ \theta > 0 \quad \exists\ \epsilon(\theta, d) > 0,\ C(\tau, d) < \infty$ s. t. $E + D \le \epsilon \Longrightarrow$

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

$$\int_{B_1\times(0,1)} \frac{1}{\rho}\left|j - \rho\nabla\phi\right|^2 \le \theta E + CD.$$

Amounts to: Eulerian velocity $\frac{j}{\rho} \approx \nabla\phi$ harmonic gradient

Construct $\nabla\phi$ via flux (Neumann) data



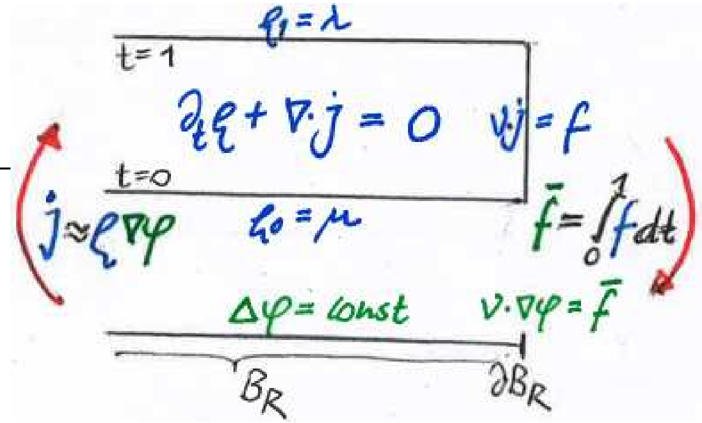
Normal flux $\nu \cdot j$ across bdry ∂B_R,

its time integral $\int_0^1 \nu \cdot j dt$.

Proposition 1" $\forall\, \theta > 0 \;\; \exists\, \epsilon > 0,\, C < \infty : \;\; E + D \le \epsilon \implies$ $\exists\, R \in [1,2]$ s. t. $\triangle\phi =$ const in B_R, $\nu \cdot \nabla\phi = \int_0^1 \nu \cdot j dt$ on ∂B_R satisfies

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

$$\int_{B_1 \times (0,1)} \frac{1}{\rho} \left| j - \rho\nabla\phi \right|^2 \le \theta E + CD.$$



cf. Dacorogna-Moser.

Choice of "good" radius R.

Analogies to De Giorgi's approach to regularity for 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 / time-averaged vs. time-resolved;
Lower-dimensional isoperimetric estimate:
error is of higher-order (choice of good radius).

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

ϵ-regularity as Schauder theory

Recall Hölder semi-norms $[u]_{\alpha,B} := \sup_{x\neq x'\in B}\frac{|u(x)-u(x')|}{|x-x'|^\alpha}$, $\alpha\in(0,1)$.

Suppose $\lambda = \lambda\, dx,\ \mu = \mu\, dy$

with Hölder continuous λ,μ and $\lambda(0)=\mu(0)=1$.

Monitor (the dimensionless) $E := \frac{1}{R^2|B_R|}\int_{\Omega\cap B_{2R}}|T-x|^2dx$

with $T=\nabla\psi$ Brenier map.

Monitor (the dimensionless) $D := R^{2\alpha}[\lambda]^2_{\alpha,B_{2R}} + R^{2\alpha}[\mu]^2_{\alpha,B_{2R}}$.

Theorem 1 (Goldman&O., à la DePhilippis&Figalli)

If $E + D \ll 1$ then $R^{2\alpha}[\nabla T]^2_{\alpha,B_R} \lesssim E + D$.

Amounts to $C^{2,\alpha}$-regularity for Monge-Ampère $\det D^2\psi = \lambda$ $(\mu\equiv 1)$.

Comparison: DePhilippis&Figalli vs. Goldman&O.

Theorem 1 (Goldman&O., à la DePhilippis&Figalli)
$E+[\lambda]^2_{\alpha,B_2}+[\mu]^2_{\alpha,B_2}\ll 1 \implies [D^2\psi-\mathrm{id}]^2_{\alpha,B_1}\lesssim E+[\lambda]^2_{\alpha,B_2}+[\mu]^2_{\alpha,B_2}$.

Perturbation around linear $\triangle\phi = const$,
as opposed to nonlinear $\det D^2\psi = 1$ (ϵ-regularity Figalli&Kim).

Get (immediately) linear homogeneities $[D^2\psi\text{-id}]_{\alpha,B_1}\lesssim[\lambda]_{\alpha,B_2}+[\mu]_{\alpha,B_2}$.

Get $\psi\in C^{2,\alpha}$ in one step,
as opposed to three-step bootstrap $C^{1,\alpha}$, $C^{1,1}$, $C^{2,\alpha}$.

Competitors in strictly convex variational problem,
instead of comparison principle.