Numerical Optimal Transportation

Jean-David Benamou (INRIA)
https://team.inria.fr/mokaplan/

Contents

Brief Recap of theory

- `http://math.berkeley.edu/~evans/Monge-Kantorovich.survey.pdf`
- `http://www.stochastik.uni-freiburg.de/~rueschendorf/papers/MKTransProbOptCouplings.pdf`
- `https://www.ceremade.dauphine.fr/~carlier/IMA-transport-Lecture-Notes.pdf`
- `http://cvgmt.sns.it/paper/832/` (F. Santambrogio)
- `http://www.math.toronto.edu/mccann/papers/FiveLectures.pdf`
- ....

Data (C2C) (D2D) (D2C)

We are given $\mu \in P(X)$ and $\nu \in P(Y)$ (probability measures on X and Y).
We will discuss the numerical resolution of Optimal Transportation problems in the following three cases :
C2C : They may be absolutely continuous w.r.t. the Lebesgue Measure : $\mu = f(x)dx$ where $f \in \mathcal{L}^1(X)$ and $\nu = g(y)dy$ where $g \in \mathcal{L}^1(Y)$. We will use a classic FD or FE discretization of f and g.
D2D : They may be both in (discrete) atomic form $\mu = \sum_{i=1}^{M} \mu_i \delta_{x_i}$ where $\mu_i > 0$ and $\delta_{x_i}(A) = 1$ if $x_i \in A$ and 0 else (for all A measurable subset of X) and likewise $\nu = \sum_{j=1}^{N} \nu_j \delta_{y_j}$.
C2D (or $D2C$) : Finally, $\mu = f(x)dx$ and $\nu = \sum_{j=1}^{N} \nu_j \delta_{y_j}$.

Tansport Maps and PushForwards

A measurable map $T : X \mapsto Y$ **Pushes Forward** μ to ν , Noted
$T\#\mu = \nu$ if
$\nu(B) = \mu(T^{-1}(B)$ for all B measurable subset of Y,
or
$\int_Y \phi \, d\nu = \int_X \phi \circ T \, d\mu$ for all $\phi \in \mathcal{C}(Y)$,
or
$(g \circ T) \, det(DT) = f$ if T has smoothness.

Nota : $T(X) \subset Y$ and $T(X) = Y$ if $\nu > 0$
this is a **State Constraint** on T.

Monge Problem

$$(MP) \inf_{T:X\mapsto Y,\, T\#\mu=\nu} \int_X c(x, T(x))\, d\mu(x)$$

- Classical Monge "Ground cost" : $c(x, y) = \|x - y\|$.
  http://images.math.cnrs.fr/Gaspard-Monge,1094.html.
- Quadratic MK cost : $c(x, y) = \frac{\|x-y\|^2}{2}$.
  http://images.math.cnrs.fr/_Brenier-Yann_.html
  http://images.math.cnrs.fr/pdf2004/Villani.pdf

- Coulomb cost (DFT) : $c(x, y) = \frac{1}{\|x-y\|}$.
  C.Cotar, G.Friesecke, C.Klüppelberg, Density functional theory and optimal transportation with Coulomb cost, in Comm. Pure Appl. Math., 2012
- Far field Reflector Problem : $X, Y \in \mathcal{S}^2$ and
  $c(x, y) = -log(\|x - y\|)$.
  L. Caffarelli, S. Kochengin, and VI Oliker, On the numerical solution of the problem of reflector design with given far-field scattering data, Contemporary Mathematics 226 (1999), 1332.

Brenier Theorem (91)

- There is a unique Transport Map $\overline{T} = \nabla u$, u convex minimising Monge Problem for the quadratic cost.
- The result was extended to $c(\|x - y\|)$, c strictly convex by Gangbo and McCann ...
- Corollary : u is a weak "Brenier" solution of the Monge-Ampère equation
  $(g \circ \nabla u)\, det(D^2 u) = f$ with "BC" $\nabla u(X) \subset Y$
  (2nd BVP for the (MA) equation).

1D is (almost) explicit

- $g(u'(x))u''(x) = f(x)$
- $G(u'(x)) = F(x)$
- $u'(x) = G^{-1}(F(x))$

Transport Plans

- $\Pi(\mu,\nu)$: **Transport Plans**, the set of $\gamma \in \mathcal{P}(X \times Y)$ s.t.

  $$\nu(B) = \gamma(X \times B) \text{ and } \mu(A) = \gamma(A \times Y)$$

  for all (A, B) measurable subsets of $X \times Y$.

- or :
  $\int_{X\times Y}(\phi(x) + \psi(y))\, d\gamma(x, y) = \int_X \phi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y)$
  for all $(\phi, \psi) \in \mathcal{C}(X) \times \mathcal{C}(Y)$.

- $\mu \otimes \nu$ is a transport plan .

- $\gamma_T = (Id \times T)\#\mu$ is a transport plan if T is a transport map.

- ...

The Monge-Kantorovich problem and its dual

- Primal :
  $\inf_{\{\gamma \in \Pi(\mu,\nu)\}} \int_{X\times Y} c(x,y)\, d\gamma(x,y)$
- Saddle Point Formulation
  $\inf_{\{\gamma \geq 0} \sup_{(\phi,\psi)\in \mathcal{C}(X)\times\mathcal{C}(Y)\}} \{\int_X \phi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y)$
  $+ \int_{X\times Y} (c(x,y) - \phi(x) - \psi(y))\, d\gamma(x,y)\}.$
- Dual :
  $\sup_{\{(\phi,\psi)\in \mathcal{C}(X)\times\mathcal{C}(Y), s.t. c(x,y)\geq \phi(x)+\psi(y), \forall (x,y)\}}$
  $\left\{\int_X \phi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y)\right\}$
- Dual again :
  $\sup_{\{\phi\in \mathcal{C}(X)\}} \left\{\int_X \phi(x)\, d\mu(x) + \int_Y \phi^c(y)\, d\nu(y)\right\}$

Monge Kantorovich

- Continuity of c is sufficient for the existence of a solution and Primal $=$ Dual.
- If $c(x, y) = h(\|x - y\|)$, h, strictly convex then the solution is given as a Transport Map : $\overline{\gamma} = (Id \times \overline{T})\#\mu$ and the relative transport is given by $\overline{T} - x = -(\nabla h)^{-1}(\nabla\phi(x))$ ($\nabla h = Id$ for the quadratic cost).
- Lots of theory there (twist condition) Ma-Wang Trudinger ....
- stability results ...

Monge Kantorovich for **D2D** data : Linear Programming

- $\mu = \sum_{i=1}^{M} \mu_i \delta_{x_i}$ where $\mu_i > 0$ and $\nu = \sum_{j=1}^{N} \nu_j \delta_{y_j}$ where $\nu_j > 0$. Also $\sum_{i=1}^{M} \mu_i = \sum_{j=1}^{N} \nu_j = 1$.
- Transport Plans are restricted to $\gamma = \sum_{i=1}^{M} \sum_{j=1}^{N} \gamma_{ij} \delta_{x_i} \delta_{y_j}$ and the problem becomes a linear problem
  $\inf_{\{\gamma_{ij} \geq 0, \sum_j \gamma_{ij} = \mu_i \;\; , \sum_i \gamma_{ij} = \nu_j\}} \sum_{i,j} \gamma_{ij}\, c_{ij}$
  where $\mu_i = \mu(x_i)$, $\nu_j = \nu(y_j)$ and $c_{ij} = c(x_i, y_j)$.
- the Dual is of the form
  $\sup_{\{(\phi_i, \psi_j),\, s.t.\, c_{ij} \geq \phi_i + \psi_j\}} \sum_{i=1}^{N} \phi_i\, \mu_i + \sum_{j=1}^{M} \psi_j\, \nu_j$
- Standard Linear Problems. N^2 unknown - $2 * N$ constraints

The Assignment problem

Assume from now on $M = N$.

- When $\mu_i = \nu_j = \frac{1}{N}$ we get (normalize)
  $\inf_{\{\gamma_{ij} \geq 0, \sum_j \gamma_{ij}=1 \ , \sum_i \gamma_{ij}=1\}} \sum_{i,j} \gamma_{ij} \, c_{ij}$

- (Birkhoff) : Extreme points of bistochastic matrices are permutations matrices.
  $\inf_{\sigma \in \mathcal{S}^{\mathcal{N}}} \sum_i c_{i\sigma(i)}$

- Large litterature, Hungarian method, Bertsekas Auction but $O(N^3)$ (see Q. Merigot slides). Close to the $C2D$ numerical methods discussed in session #2.

- Back to MK D2D problem.

A Matlab implementation of the LP approach

- Doc Linprog .
- $\inf_{x,\, A\,x=b}\ f^t\, x$.
- Attention : Equality constraint matrix has rank $2*N-1 \rightarrow$ remove one line ($\sum_i \mu_i = \sum_j \nu_j$) .

- E. Schrodinger. Uber die umkehrung der naturgesetze. Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math., 144:144153, 1931
- R. Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. Amer. Math. Monthly, 74:402405, 1967
- L. Ruschendorf. Convergence of the iterative proportional fitting procedure. The Annals of Statistics, 23(4):11601174, 1995.
- D. Bosc PhD
`https://pastel.archives-ouvertes.fr/pastel-00721674`
- A. Galichon and B. Salanié. Matching with trade-offs: Revealed preferences over competing characteristics. Preprint ssrn-1487307, Preprint SSRN-1487307, 2009.
- M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal trans- port. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, Proc. NIPS, pages 22922300, 2013.
- M. Cuturi and A. Doucet. Fast computation of wasserstein barycenters. In Proc. ICML, 2014.

Entropy regularization and IPFP for the impatient

ϵ is a temperature

- $\min_{\{\gamma^{\epsilon}_{ij}\geq 0, \sum_j \gamma^{\epsilon}_{ij}=\mu_i \ , \sum_i \gamma^{\epsilon}_{ij}=\nu_j\}} \sum_{i,j} \gamma^{\epsilon}_{ij}\, c_{ij} + \epsilon\, \gamma^{\epsilon}_{ij}\, (log \gamma^{\epsilon}_{ij} - 1)$
- $\min_{\{\gamma^{\epsilon}_{ij}\geq 0\}} \max_{\{\phi^{\epsilon}_i, \psi^{\epsilon}_j\}} \sum_{ij} \psi^{\epsilon}_j\, \nu_j + \phi^{\epsilon}_i\, \mu_i$
  $+\gamma^{\epsilon}_{ij}\, (c_{ij} - \psi^{\epsilon}_j - \phi^{\epsilon}_i + \epsilon\, (log \gamma^{\epsilon}_{ij} - 1))$
- $\gamma^{\epsilon}_{ij} = e^{\frac{\phi^{\epsilon}_i}{\epsilon}}\, e^{-\frac{c_{ij}}{\epsilon}}\, e^{\frac{\psi^{\epsilon}_j}{\epsilon}}$

Entropy regularization and IPFP for the impatient

ϵ is a temperature

- set $\overline{\gamma}^{\epsilon}_{ij} = e^{-\frac{c_{ij}}{\epsilon}}$ $a^{\epsilon}_i = e^{\frac{\phi^{\epsilon}_i}{\epsilon}}$ and $b^{\epsilon}_j = e^{\frac{\psi^{\epsilon}_j}{\epsilon}}$.
- Use margin constrains to get $a^{\epsilon}_i = \frac{\mu_i}{\sum_j \overline{\gamma}^{\epsilon}_{ij} b^{\epsilon}_j}$ and $b^{\epsilon}_j = \frac{\nu_j}{\sum_i \overline{\gamma}^{\epsilon}_{ij} a^{\epsilon}_i}$.
- IPFP $=$ relaxation
  $b^{\epsilon,0} = 1$
  $$a_i^{\epsilon,k+\frac{1}{2}} = \frac{\mu_i}{\sum_j \overline{\gamma}^{\epsilon}_{ij} b_j^{\epsilon,k}} \qquad b_j^{\epsilon,k+1} = \frac{\nu_j}{\sum_i \overline{\gamma}^{\epsilon}_{ij} a_i^{\epsilon,k+\frac{1}{2}}}.$$

Bregman iterative projection for KL divergence

L.M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR computational mathematics and mathematical physics, 7(3):200217, 1967.

- $KL(\gamma|\overline{\gamma}) = \sum_{i \in I} \gamma_i \, (log(\frac{\gamma_i}{\overline{\gamma}_i}) - 1) = \epsilon$ times what we had above
- We look for $P^{KL}_{\gamma \in C}(\overline{\gamma}) = Argmin_{\gamma \in C} KL(\gamma|\overline{\gamma})$ where $C = C_\mu \cap C_\nu = \{\gamma, \sum_j \gamma^\epsilon_{ij} = \mu_i\} \cap C_\mu = \{\gamma, \sum_i \gamma^\epsilon_{ij} = \nu_j\}$ is the intersection of linear (Convex) subspaces ..
- Bregman iterative alternate projection converge (Bauske Lewis) $\gamma^0 = \overline{\gamma}$ then $\forall k \; \gamma^{k+1} = P^{KL}_{\gamma \in C_\mu}(\gamma^k)$ and $\gamma^{k+2} = P^{KL}_{\gamma \in C_\nu}(\gamma^{k+1})$.

Bregman iterative projection for KL divergence

- Projections are explicit
  $\gamma_{ij}^{k+1} = \mu_i \frac{\gamma_{ij}^k}{\sum_j \gamma_{ij}^k}$ and $\gamma_{ij}^{k+2} = \nu_i \frac{\gamma_{ij}^{k+1}}{\sum_i \gamma_{ij}^{k+1}}$

- One can recover the IPFP method by setting $\gamma_{ij}^k = a_i^{\epsilon,k-\frac{1}{2}} \overline{\gamma}_{ij}^{\epsilon} b_j^{\epsilon,k}$
  $\gamma_{ij}^{k+1} = a_i^{\epsilon,k+\frac{1}{2}} \overline{\gamma}_{ij}^{\epsilon} b_j^{\epsilon,k}$

Convergence Theory ..

- Convergence with ϵ (interior point method convergence)
  Roberto Cominetti, Jaime San Martin: Asymptotic analysis of the exponential penalty trajectory in linear programming. Math. Program. 67: 169-187 (1994)
- Convergence of IPFP iterates (Ruschendorf '95).
- Toy implementation in Matlab.

– features of IPFP / KL Bregman projections

- Costs $N^2 \times \#\textit{iterations}$, store N^2 matrices
- $\#$ iterations increase with ϵ.
- Not always easy to recover the Transport Map.

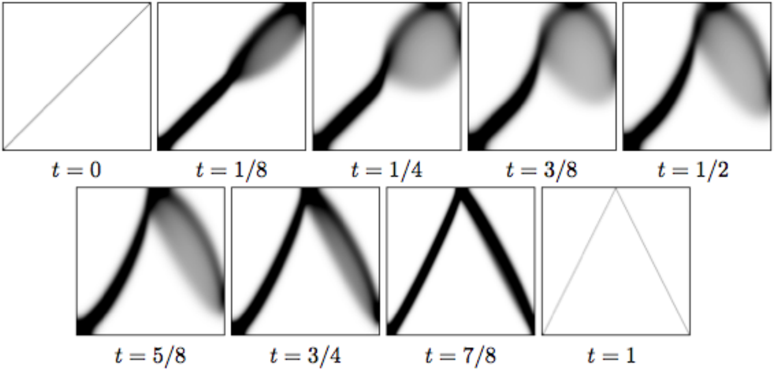
+ features of IPFP / KL Bregman projections

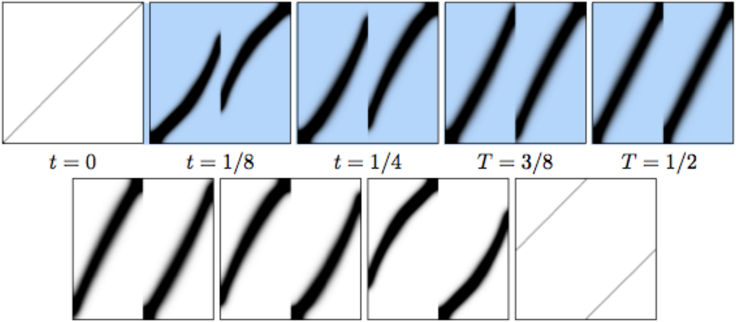
- Easy implementation (does not depend on discretisation or dimension of the problem).
- Parallelization ...
- Applies to general cost matrix c_{ij}.
- Generalizes to many variants of OT (cf G. Peyré / L. Nenna talk).

Generalized Euler Flows -Multimarginal OT

- Y. Brenier. Generalized solutions and hydrostatic approximation of the Euler equations. Phys. D, 237(14-17) :1982-1988, 2008.
- Cost function $C_{j_1,j_2,\ldots,j_K} = \sum_{i=1,..K-1} \|x_{j_{i+1}} - x_{j_i}\|^2 + \|x_{\sigma(j_1)} - x_{j_k}\|^2$
- Tansport plan $\gamma_{j_1\ldots j_K}$ has Lebesgue Marginals and 2-Couplings $(T_{1,k})_{s,w} = \sum_{j_i \neq j_1, j_k} \gamma_{s,j_2,\ldots,j_{k-1},w,j_{k+1},\ldots,j_K}, \quad s, w = 1, ..., N$. give fluids particles ”trajectories”.
- (from B. Carlier Cuturi Peyré Nenna ...))

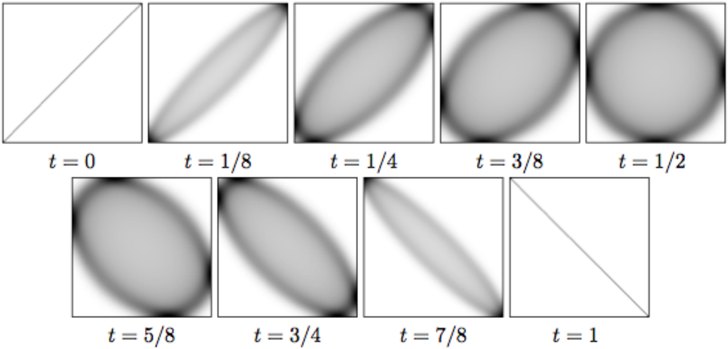
Numerical computations of Transport Plans - *D2D* data





$t = 0$ $t = 1/8$ $t = 1/4$ $T = 3/8$ $T = 1/2$

$t = 5/8$ $t = 3/4$ $t = 7/8$ $t = 1$


t = 0
t = 1/8
t = 1/4
t = 3/8
t = 1/2
t = 5/8
t = 3/4
t = 7/8
t = 1

Contents

(C2D) data Pogorelov Solution Biblio

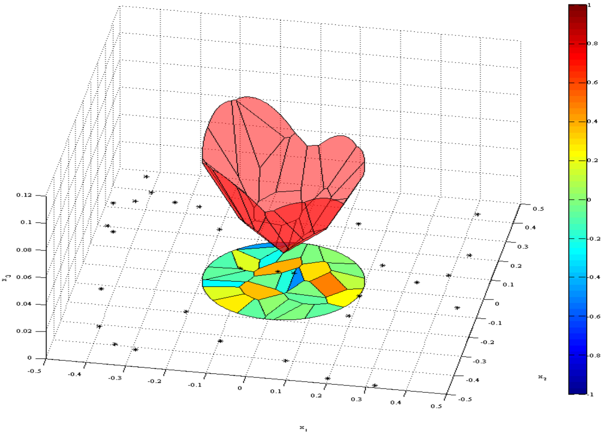
- Pogorelov, ...
- VI Oliker and LD Prussner, On the numerical solution of the equation $\frac{\partial^2 z}{\partial^2 x}\frac{\partial^2 z}{\partial^2 y}?(\frac{\partial^2 z}{\partial x \partial y)^2} = f$, Numerische Mathematik 54 (1988), no. 3, 271–293.
- MJP Cullen and RJ Purser, An extended lagrangian theory of semi-geostrophic frontogenesis., Journal of Atmospheric Sciences 41 (1984), 1477–1497.
- Mérigot, Quentin, A comparison of two dual methods for discrete optimal transport, Geometric science of information, P 389-396, 3642400191 , 2013 , Springer Berlin Heidelberg.
- Damien Bosc PhD `https://pastel.archives-ouvertes.fr/pastel-00721674.`
- Bruno Levy `http://arxiv.org/abs/1409.1279`

- $X = \{\|x\| < 1\}$, $\mu = 1dx$, $\nu = \sum_{k=1}^{K} \alpha_k \delta_{d_k}$, $\sum_k \alpha_k = 2\pi$
- The (convex) Brenier Solutions takes an explicit form Pogorelov solutions (1964).

$$\phi_V(x) = \sup_k \{x \cdot d_k - v_k\}$$

- $C_k = \{x \in X, \nabla \phi_V(x) = d_k\}$ is the support of the cell being mapped to the Dirac at d_k.
- $V = \{v_k\}$ need to be adjusted such that $\mu(C_k) = |C_k| = \alpha_k, \forall k$ (Rearrangement mapping).

Not converged example



- ”Classical” num. methods rely on Laguerre (Power) Diagrams (computation in $O(K\ logK)$ for $d = 2$) .

$$
\begin{aligned}
C_k &= \{x \in X, x \cdot d_k - v_k \geq x \cdot d_j - v_j,\ \forall j \neq k\} \\
&= \{x \in X, \frac{\|x-d_k\|^2}{2} + w_k \leq \frac{\|x-d_j\|^2}{2} + w_j,\ \forall j \neq k\}
\end{aligned}
$$

with $w_i = \frac{\|d_i\|^2}{2} + v_i$.
($w_i = cst$ corresponds to Voronoi diagrams) .

SD OT is a finite dimensional optimization problem

- $\phi(x) = \sup_k \{x \cdot d_k - v_k\}$ for $x \in B(0,1)$ (and extended with $+\infty$ outside).
  $\phi^*(y) = \sup_{x \in B(0,1)} \{y \cdot x - \phi(x)\}$.
- ϕ convex and lsc $\Leftrightarrow \quad \phi = \phi^{**}$.
- Subgradients : $\partial\phi(x) = \{y \in \mathbb{R}^2 \mid \phi(z) \geq u(x) + y \cdot (z - x) \forall z\}$
- $y \in \partial\phi(x) \Leftrightarrow x \in \partial\phi^*(y)$
- We have an explicit form for $\phi^* = (\min_k \{v_k + \|y - d_k\|\})^{**}$ and in particular $\phi^*(d_k) = v_k$.

SD OT is a finite dimensional optimization problem

- Recall Dual : $\sup_{\phi \in \mathcal{C}(X)} \left\{ \int_X \phi(x)\, d\mu + \int_Y \phi^c(y)\, d\nu(y) \right\}$ We are going to construct a similar problem ...
- We have $\phi(x) = x \cdot d_k - v_k$ for $x \in C_k$ and $\phi^*(d_k) = v_k$
- and $\mu = 1dx$, $\nu = \sum_{k=1}^{K} \alpha_k \delta_{d_k}$.
- Replacing, we get : $\sup_{V=\{v_k\}} \Phi(V)$
- $\Phi(V) = \sum_k \left\{ \int_{C_k} (x \cdot d_k - v_k)\, d\mu + \alpha_k\, v_k \right\}$

Properties of Φ

Assuming all C_ks have mass.

- Gradient : $\frac{\partial \Phi(V)}{\partial v_k} = -C_k(V) + \alpha_k$.
- Hessian : $\frac{\partial^2 \Phi(V)}{\partial v_k \, \partial v_{k'}} = -\frac{|\Sigma_{k\,k'}|}{\|d_k - d'_k\|}$, $\frac{\partial^2 \Phi(V)}{\partial v_k^2} = \sum_{k' \neq k} \frac{|\Sigma_{k\,k'}|}{\|d_k - d'_k\|}$
- $\Sigma_{k\,k'} = C_k \cap C'_k$

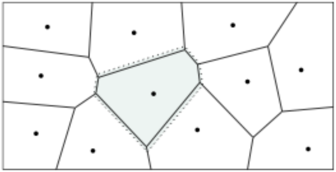


Fig. 1. A Laguerre (or Power) diagram of a planar point set. In light gray, the evolution of a Laguerre cell when the weight of the corresponding point is decreased by ε.

(from Mérigot)

Properties of Φ Continued

- Gradient $\frac{\partial \Phi(W)}{\partial w_k}$ is K Lipschitz $K \leq \frac{C}{\min_{k' \neq k} \|d_k - d'_k\|}$.
- Hessian is singular (sum of lines $= 0$) but setting $v_0 = 0$ (fix the constant) we get a (strictly) diagonally dominant matrice $\rightarrow$ Hessian is positive definite.
- Φ is concave.
- Open for optimization.

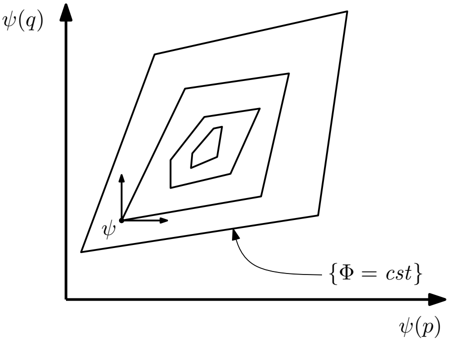
Classical ”Pogorelov” technique (Oliker-Prussner)

- Coordinate-wise descent.
- Need Laguerre Cells area computation
  (Toy implementation with MPT
  `http://people.ee.ethz.ch/~mpt/3/`
  (Better use CGAL for performance `https://www.cgal.org/`
- Choose V^0 such that $|C_k| \leq \alpha_k + \delta$ for $k \neq 0$.
- Repeat : while $\exists k \neq 0$ s.t $|C_k| \leq \alpha_k - \delta$ decrese v_k s.t.
  $|C_k| \in [\alpha_k, \alpha_k + \delta]$

δ enforces convergence in K^2 operations (see Mérigot).

Coordinate-wise Maximization of Ψ

Is it a good idea ?



Bertsekas' solution: impose a minimum bid ε

Toy Implementation

Aleksandrov solutions (change of notations ...)

Definition

- Subgradients :
  $\partial u(x) = \{p \in \mathbb{R}^2 \mid u(z) \geq u(x) + p \cdot (z - x)\ \forall z \in \mathbb{R}^2\}$
- Subgradient of a set E : $\partial u(E) = \bigcup_{y \in E} \partial u(y)$.
- MA equation in the Aleksandrov sense

$$\rho_Y(\partial u(E)) = \rho_X(E), \ \ \textit{for all}\ E \subset X$$

- For smooth functions and densities, make the change of variables $y = \nabla u(x)$:

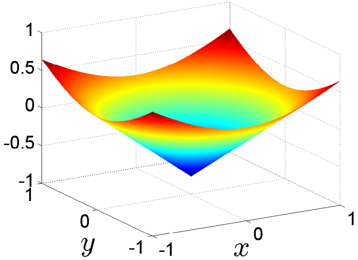
$$\int_E \rho_X(x)\,dx = \int_{\partial u(E)} \rho_Y(y)\,dy = \int_E \rho_Y(\nabla u(x))\,det(D^2 u(x))dx$$

to recover the MA equation.

Example

$u(x) = \|x - x_0\|$ is the Aleksandrov solution of

$$det(D^2 u) = 2\,\pi\delta_{x_0}, \qquad \nabla u(X) \subset Y = B(x_0, 1)$$



$$\partial u(x) = \begin{cases} Y \text{ if } x = x_0, \\ \frac{x - x_0}{|\|\|x - x_0\|} \text{ else.} \end{cases}$$

, u NOT a viscosity solution.

D2C Aleksandrov are dual to *C2D* Pogorelov solutions

- Recall the LFT def. : $u^*(y) = \sup_{y \in B(0,1)} \{y \cdot x - u(x)\}$.
- We have an explicit formula for u_V^*

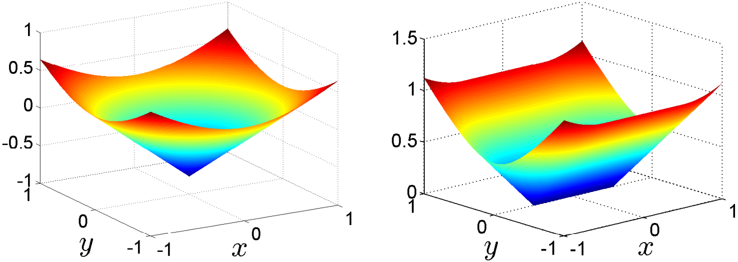
$$u_V(x) = \sup_k \{x \cdot d_k - v_k\}$$

$(u_V^* = w^{**},$ with $w(y) = \min_k \{v_k + \|y - d_k\|\}.))$

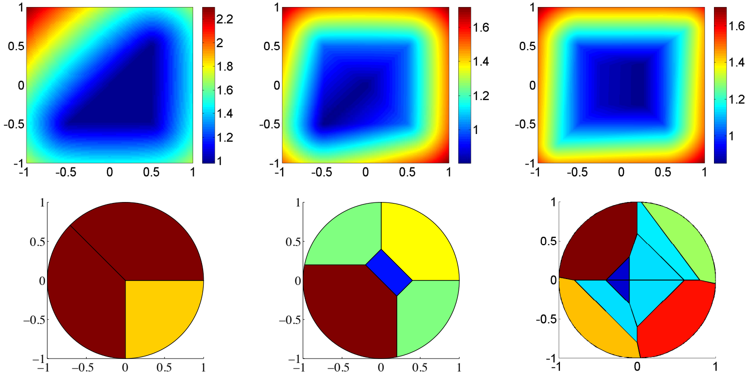
- u_V^* is the Aleksandrov solution of

$$det(D^2\psi) = \sum_{k=1}^{K} \alpha_k \delta_{d_k}, \qquad \nabla\psi(Y) \subset X = B(x_0, 1)$$

(check this)


1
0.5
0
-0.5
-1
1
0
y
-1
-1
x
0
1
1.5
1
0.5
0
1
0
y
-1
-1
x
0
1

3,5 and 10 diracs, u_V^* and the C_ks



- (Brenier, Knott-Smith, McCann, Gangbo, ...) There is a unique Lipschitz convex potential u such that $\overline{M} = \nabla u$
- Remember the Jacobian equation : $det(DM(x))\rho_Y(M(x)) = \rho_X(x)$.
- $\Rightarrow$ u is a weak ("Brenier") solution of the Elliptic Monge-Ampère equation

$$(MA) \qquad det(D^2u(x)) = \frac{\rho_X(x)}{\rho_Y(\nabla u(x))}. \; x \in X$$

- Boundary conditions are replaced by state constraints :

$$(BV2) \qquad \nabla u(\overline{X}) \subset \overline{Y}.$$

This is called the "Second Boundary Value problem" for the Monge-Ampère equation.

Illustration 2D

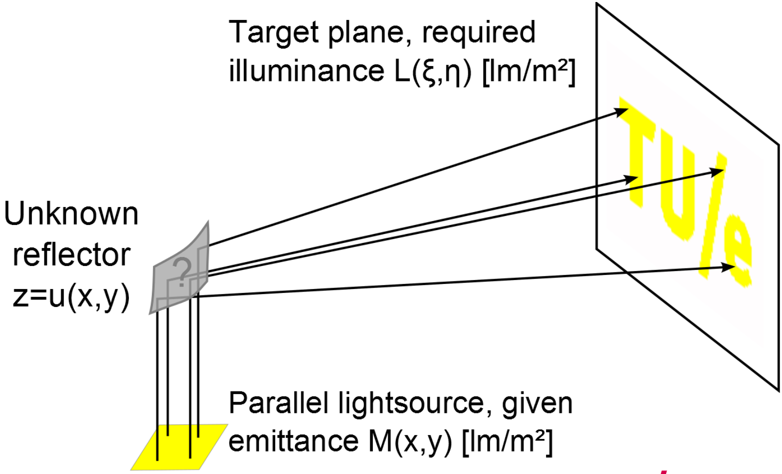
$$(t \mapsto \rho(t,.) = \{(1-\frac{t}{T})x + \frac{t}{T}\overline{M}(x)\}\#\rho(0,.), \;\; t \in]0,T[)$$

The regularity theory for the MA/OT problem is well developed

- Classical solution studied in Delanoë, Urbas, Caffarelli.
- Caffarelli : if X, Y are convex and $\lambda < \rho_{X,Y} < \frac{1}{\lambda}$ $\lambda > 0$, then Brenier Solutions have some regularity : $\psi \in C^{1,\alpha}_{loc}$.
- Caffarelli gave a counter-example where Brenier solution are not C^1.
- Recent results for non convex domains by Figalli (2010).

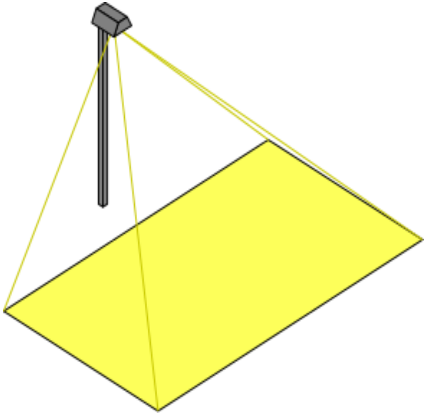
Why do we care about weaker solutions ?
Application to reflectors (C. Prins PhD Dissertation)
S-G application (Cullen ...)
Displacement interpolation

The reflector problem

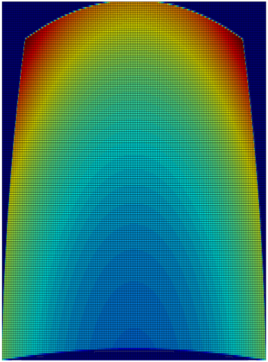


Example: lamppost

Lamppost



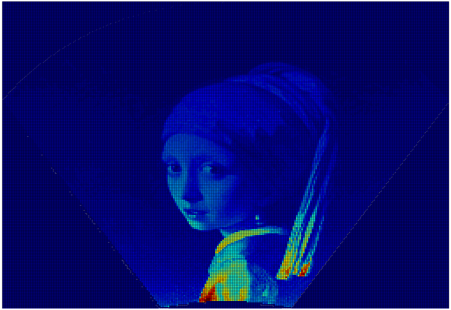
Target intensity in (ux,uy)-space



Girl with a pearl earring

Original

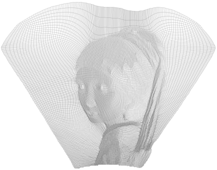
Target intensity in (ux,uy)-space



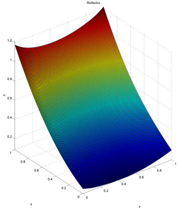
Technische Universiteit **Eindhoven** University of Technology

Reflector calculation

Mapping

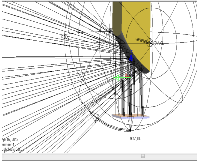


Reflector

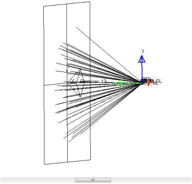


Verification with forward raytracing

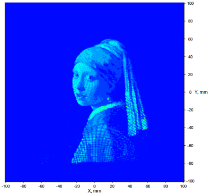
Reflector



Receiver



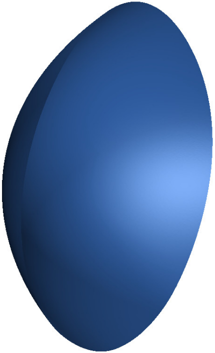
Illuminance on receiver



Numerical results (2)

$\nu = \sum_{i=1}^{N} \nu_i \delta_{x_i}$ obtained by discretizing a picture of G. Monge.

$\mu =$ uniform measure on half-sphere $\mathcal{S}^2_+$ $\qquad N = 15000$



solution to the far-field reflector problem: $R(\kappa_{\mathrm{sol}})$

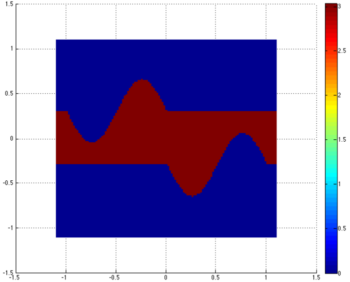
Numerical results (2)

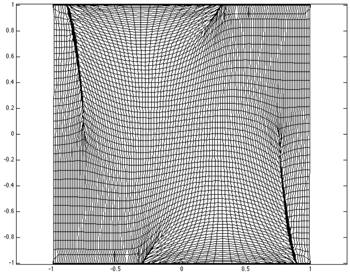
$\nu = \sum_{i=1}^{N} \nu_i \delta_{x_i}$ obtained by discretizing a picture of G. Monge.

$\mu =$ uniform measure on half-sphere $\mathcal{S}^2_+$ $\qquad N = 15000$

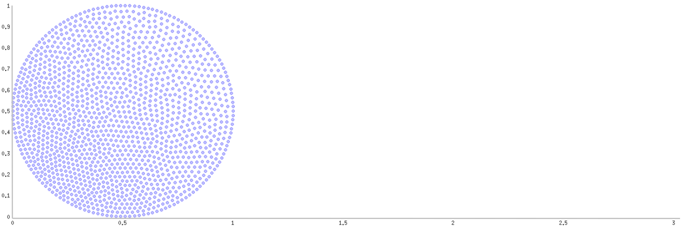
rendering of the image reflected at infinity (using LuxRender)

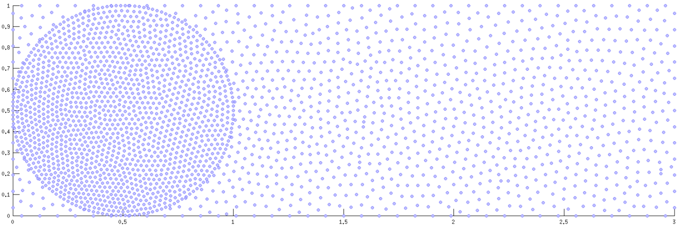
Toy Semi-Geostrophic case



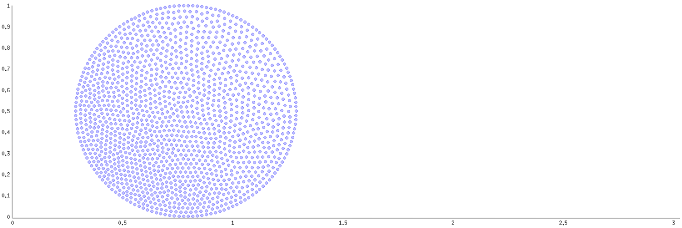


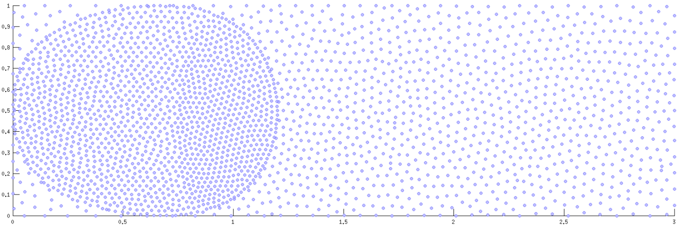
Effect of Background Non-zero Density



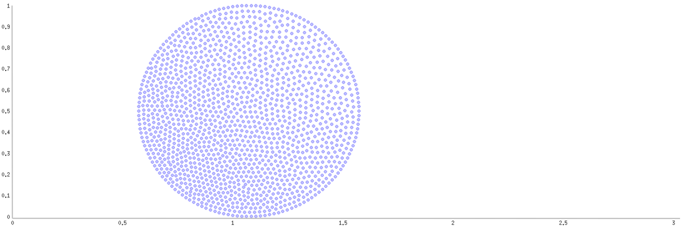


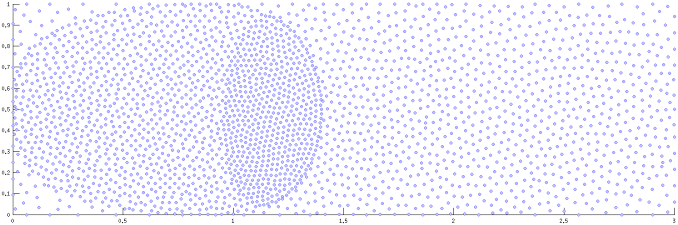
Effect of Background Non-zero Density



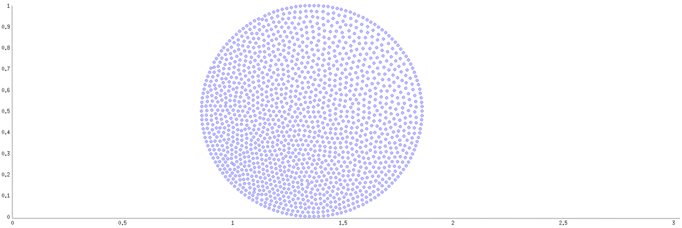


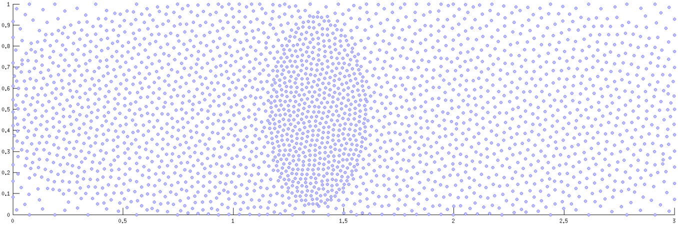
Effect of Background Non-zero Density



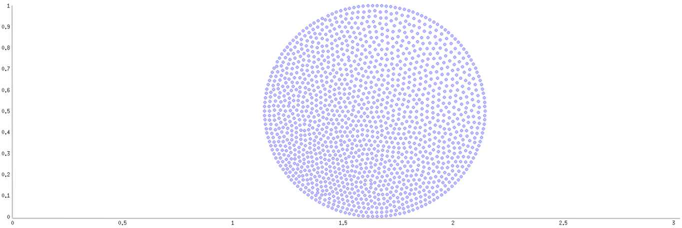


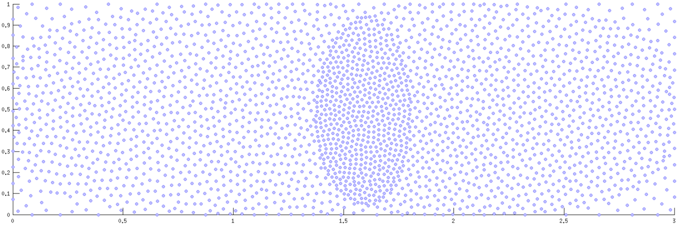
Effect of Background Non-zero Density



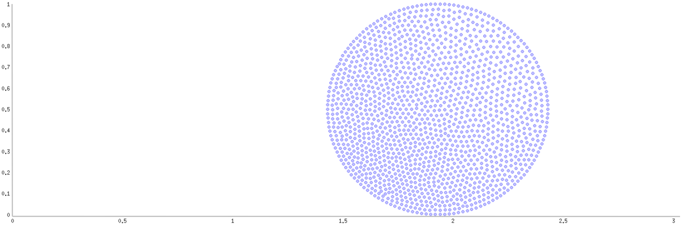


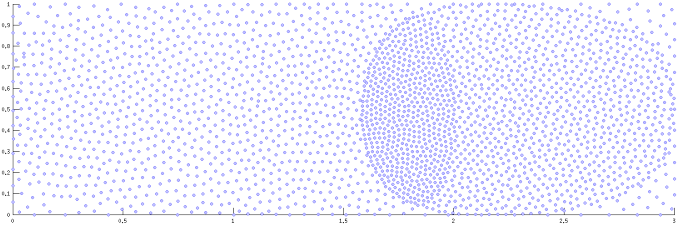
Effect of Background Non-zero Density



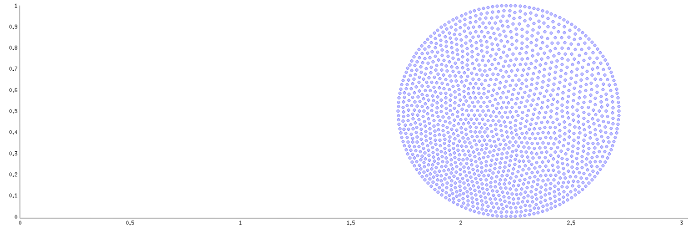


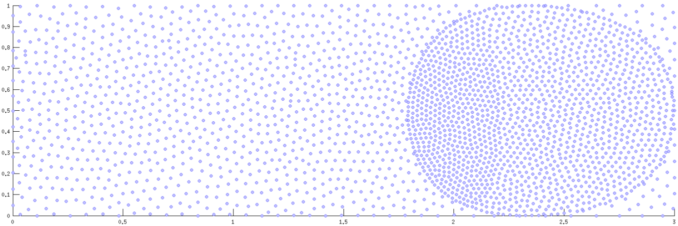
Effect of Background Non-zero Density



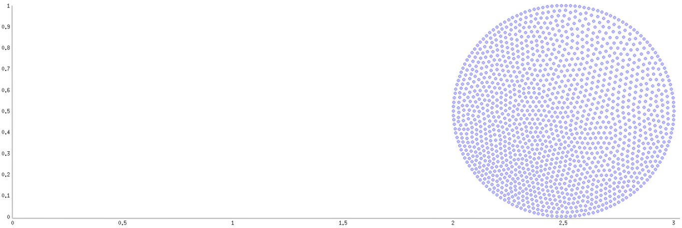


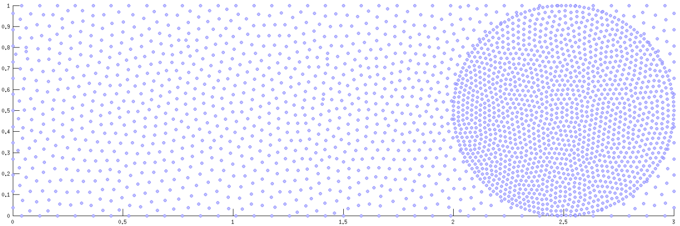
Effect of Background Non-zero Density





Effect of Background Non-zero Density

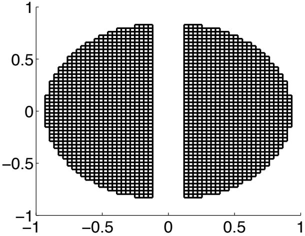


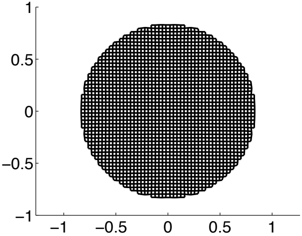


Discussion on Caffarelli counter example :

Constant density - one non convex domain

$\rho_X = \rho_Y = 1$

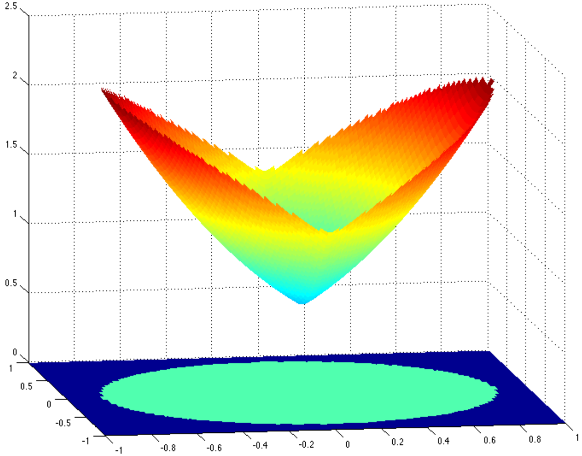




$\leftarrow$: Brenier (not Aleksandrov) Solution.

$\rightarrow$: Aleksandrov + Viscosity solution (non strictly convex).

Caffarelli counter example : a Brenier (not Aleksandrov) sol.



Aleksandrov Solutions

- Subgradients :
  $\partial u(x) = \{p \in \mathbb{R}^2 \mid u(z) \geq u(x) + p \cdot (z - x) \ \forall z \in \mathbb{R}^2\}$
- Subgradient of a set E : $\partial u(E) = \bigcup_{y \in E} \partial u(y)$.
- MA equation in the Aleksandrov sense

$$\rho_Y(\partial u(E)) = \rho_X(E), \ \textit{for all } E \subset X$$

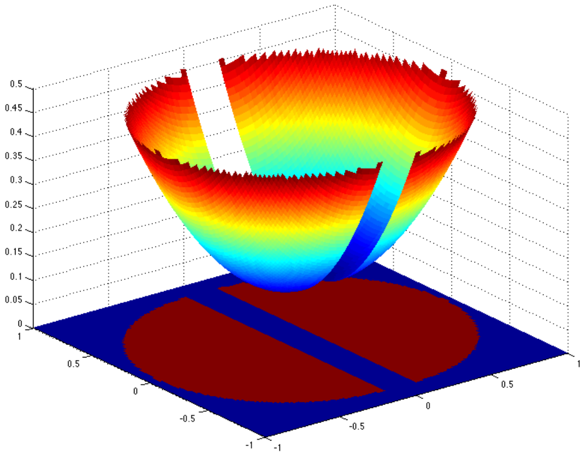
- For smooth functions and densities, make the change of variables $y = \nabla u(x)$:

$$\int_E \rho_X(x)\, dx = \int_{\partial u(E)} \rho_Y(y)\, dy = \int_E \rho_Y(\nabla u(x))\, det(D^2 u(x)) dx$$

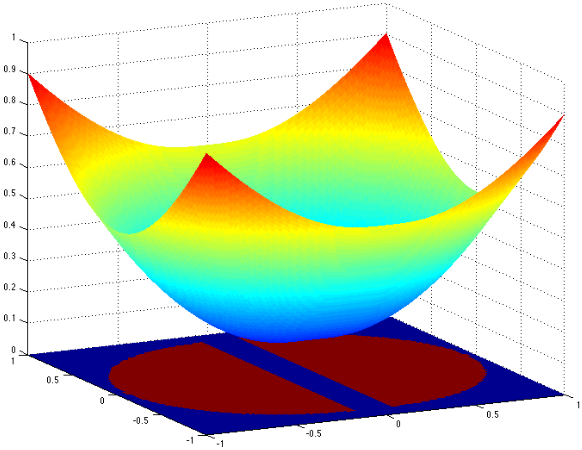
  to recover the MA equation.

Note that the previous solution is not an Aleksandrov solution.

Inverse Caffarelli counter example



The convex envelope (padd with 0s : Aleksandrov + Viscosity solution)



Def of Viscosity solution for MA

Convex Viscosity solutions are defined by testing locally against Convex Quadratic test functions $\phi(x) =< A\,x, x > + < p, x > +C$ ($A \in S_2^+$) (See Gutterez book)

- u is a viscosity sub(super) -solution at x_0 if

$$(u - \phi)(x) \geq (\leq)(u - \phi)(x_0)\ \forall x \in \mathcal{N}(x_0) \Rightarrow$$

$$det(A) \geq (\leq)\rho_X(x_0)$$

- Note that Viscosity solutions of MA allows for $\rho_X = 0$ ... only flat test functions are admissible. Convex envelopes with flat faces/zones are viscosity solutions.

Aleksandrov solutions may not be viscosity solutions (cone example)

Viscosity Solutions See User's guide by Crandall, Ishii, Lions (1992)

$F(x, \nabla u(x), D^2 u(x)) = 0$ in X , $H(x, \nabla u(x)) = 0,$ on ∂X .

$(F(x,p,A) = -det(A) + \dfrac{\rho_X(x)}{\rho_Y(p)})$ and

$(H(x,p) = sup_{\{\|n\|=1,\, n \cdot n_x \geq 0\}}\{p \cdot n - H^*(n)\})$ (BFO reformulation of BV2)

Main hypothesis for existence and uniqueness :

- (Degenerate Ellipticity) $F(x,p,A) \leq F(x,p,B)$ if $A \geq B$. True if u convex ($det(A) = \lambda_1 \lambda_2$).
- $D_p H(x,p) \cdot n_x \geq 0$ for n_x the normal at $x \in \partial x$. (Obliqueness).
- Restrict to $\rho_y > \alpha > 0$ and Lipschitz.
- When ρ_X has compact/non connected support in the computational domain. The Viscosity solution is the convex envelope of Brenier solution.

Convergent Monotone Schemes (Barles Souganidis 1991)

$u_i = u(x_i)$ for $(x_i)_{i \in I}$ a grid.

The Non-Linear Scheme : $S(x_i, u_i, (u_j)_{j \neq i}) = 0, \; \forall i \in I$.

Main hypothesis for solvability and convergence

- (Stability) + (Consistency)
- (Monotonicity) : $S(x_i, \uparrow, (\downarrow)_{j \neq i})$ discrete DE.

Jacobian Positive Definite - Damped Newton Alg. works to solve $S = 0$.
$u^{k+1} = u^k - \tau \, (\nabla S[u^k])^{-1} \, S[u^k]$, $(\nabla S[u^k]) > 0)$.

Stencils and FD of directional derivatives

- Grid discretization $X := \Omega \cap h\mathbb{Z}^2$.
- If $x \pm e \in X$ then $\Delta_e u(x) := u(x+e) - 2u(x) + u(x-e)$ (else use BCs).
- Degenerate Elliptic Schemes (Oberman '06) are $S(\Delta_{e_1} u(x_i), \Delta_{e_2} u(x_i), ...) = 0$ s.t. $S(\downarrow, \downarrow, ...)$ are monotone.
- Stencil $V(x)$=collection of direction or points e in X.
- Classical 2-D FD of $-det(D^2 u)$: $-\Delta_{(1,0)}\Delta_{(0,1)} + (\Delta_{(1,1)} - \Delta_{(1,-1)})^2/16.$ is not monotone.

WS idea (Froese-Oberman '11)

See Also Bonnans-Zidani '03 in a different context.

- Hadamard's theorem : for all $M \in S_2^+$, and any pair $(f, g) \in (\mathbb{R}^2)^2$ of non-zero orthogonal vectors, one has $< f, Mf >< g, Mg > \geq \|f\|^2 \|g\|^2 \det(M)$, with equality iff f and g are eigenvectors of M.
- $S_{FO} = \min_{\{(f,g) \in V^2, s.t. <f,g>=0\}} \frac{\Delta_f^+ u(x)}{\|e\|^2} \frac{\Delta_g^+ u(x)}{\|f\|^2}$ (DE if u remains "convex").
- Adapted to treat the 2nd Transport BC, B.-Froese-Oberman (2014).
- Only approximate consistency in practice. Consider for exemple $u_M x = \frac{1}{2} < x, M\, x >$ with eigenvectors not in V.

Monge-Ampere with Lattice Basis Reduction

Lattice Basis Reduction is the study of preferred coordinate systems for lattices (discrete subgroups of $\mathbb{R}^d$).

Definition (Superbase of $\mathbb{Z}^2$)

A superbase is a triplet $(e, f, g) \in (\mathbb{Z}^2)^3$ such that $e + f + g = 0$ and $|\det(f, g)| = 1$. It is said M-obtuse, where $M \in S_2^+$, iff $\langle e, Mf\rangle \leq 0$, $\langle f, Mg\rangle \leq 0$, $\langle g, Me\rangle \leq 0$.

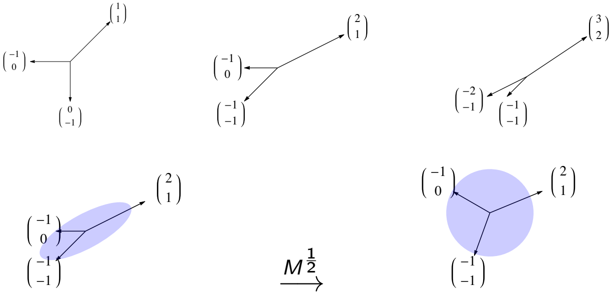


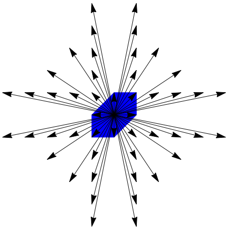
Figure : Left: An M-obtuse superbase, and the unit ball $\{\langle e, Me\rangle \leq 1\}$. Right: Likewise under change of coordinates $M^{\frac{1}{2}}$.

Definition (MA-LBR scheme with finite stencil $V \subset \mathbb{Z}^2$)

$$\mathcal{D}_V u(x) := \min_{\substack{\{e,f,g\}\subset V \\ superbase}} h(\Delta_e^+ u(x), \Delta_f^+ u(x), \Delta_g^+ u(x)).$$

where $h : \mathbb{R}_+^3 \to \mathbb{R}_+$ is

$$h(a,b,c) = \begin{cases} bc \text{ if } a \geq b + c \text{ (and likewise permuting a,b,c)}, \\ \frac{1}{4}(2ab + 2bc + 2ca - a^2 - b^2 - c^2) \text{ otherwise.} \end{cases}$$



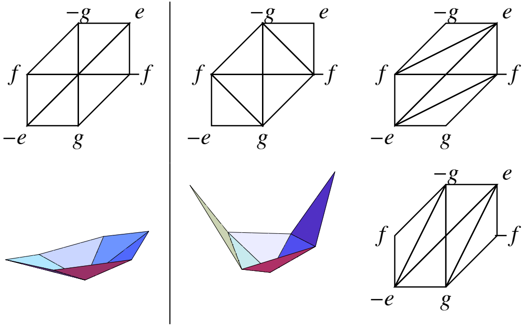


Figure : $h(a, b, c)$ can be interpreted as a subgradient measure.

Proposition (Consistency)

For any $M \in S_2^+$, $x \in X$, $\mathcal{D}_V u_M(x) \geq \det(M)$, *with equality iff* V *contains an* M *obtuse superbase.*

Relative Consistency error

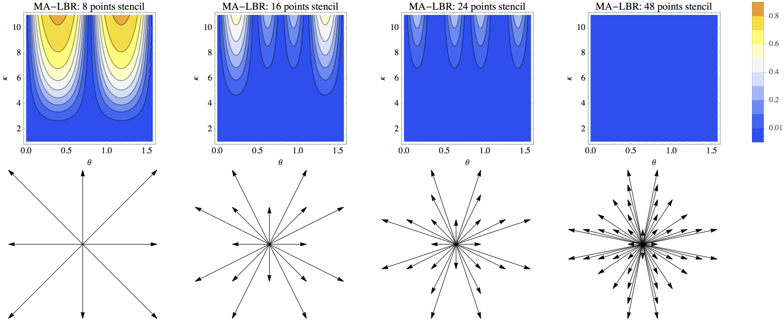


Figure : Relative consistency error $(\mathcal{D}_V(u_M) - \det(M))/\mathcal{D}_V(u_M)$, with several stencils V. Matrix $M \in S_2^+$ has condition number $\kappa^2 := \|M\|\|M^{-1}\|$ and eigenvector $(\cos\theta, \sin\theta)$.

Relative Consistency error

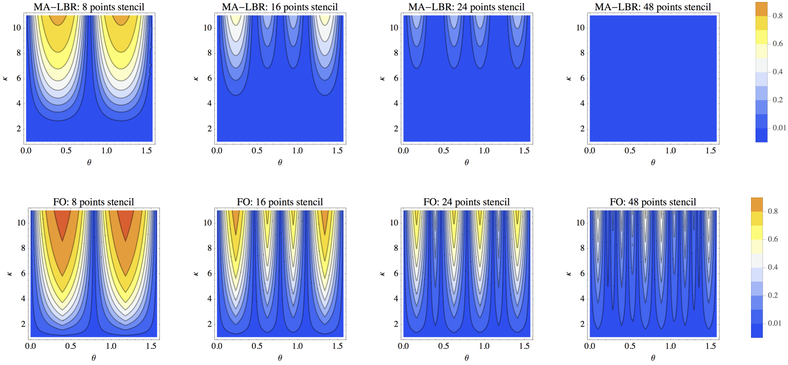


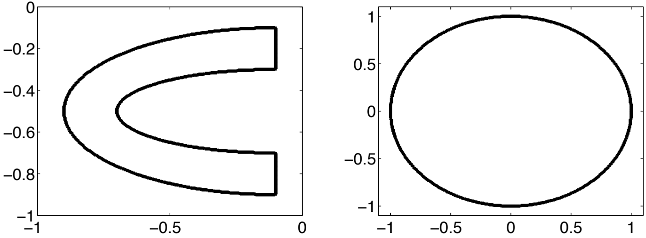
Figure : Relative consistency error $(\mathcal{D}_V(u_M) - \det(M))/\mathcal{D}_V(u_M)$, with several stencils V. Matrix $M \in S_2^+$ has condition number $\kappa^2 := \|M\|\|M^{-1}\|$ and eigenvector $(\cos\theta, \sin\theta)$.

Inverse Caffarelli example

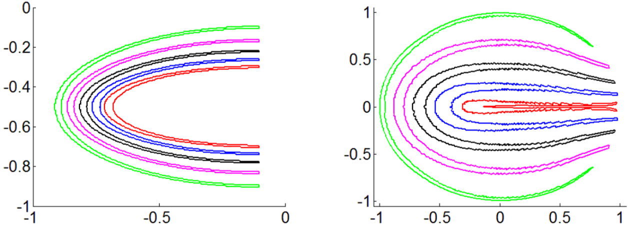
N_X	Max. Error N_Y 32	64	128	256	512	Iterations	Time (s)
32	0.0280	0.0284	0.0286	0.0286	0.0286	3	0.2
64	0.0158	0.0164	0.0165	0.0165	0.0165	3	0.4
128	0.0092	0.0093	0.0092	0.0092	0.0092	3	1.3
256	0.0047	0.0036	0.0036	0.0036	0.0036	4	8.3
512	0.0049	0.0040	0.0034	0.0033	0.0033	5	51.7

Table : Exact gradient error $\#$ Newton itération de Newton, timing for $N_y = 512$. Wide-Stencil : 9pts

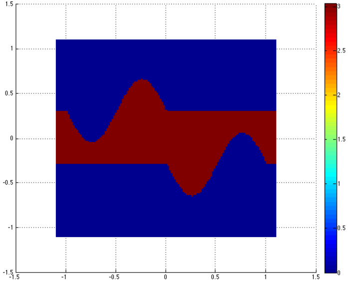
Non convex $\rho_X > 0$ support

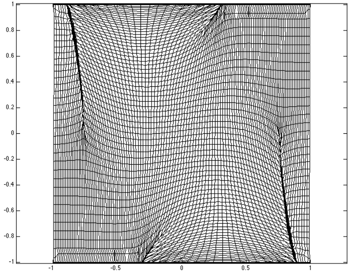


Markers displacements

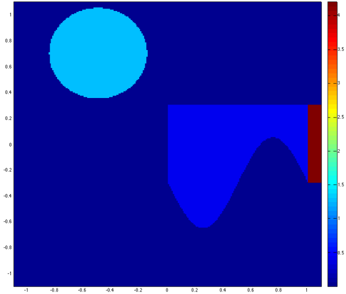


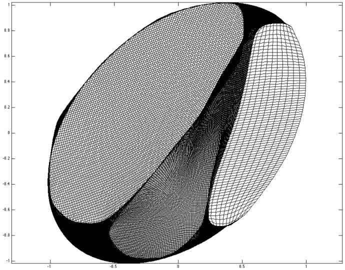
Toy Semi-Geostrophic case



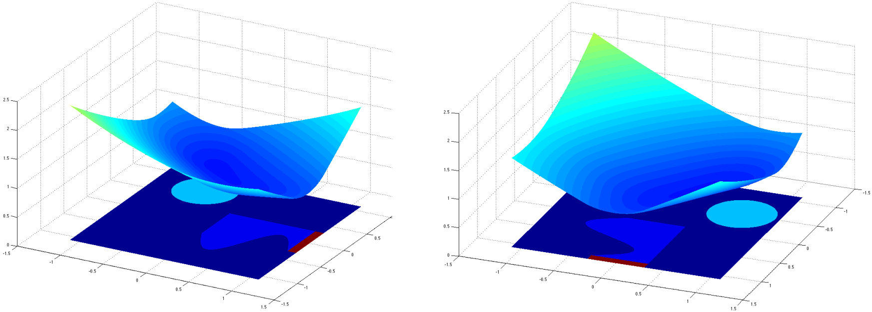


A (hard ?) test case





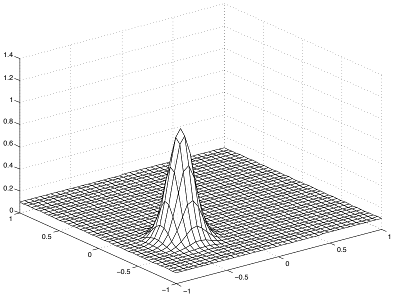
A (hard ?) test case

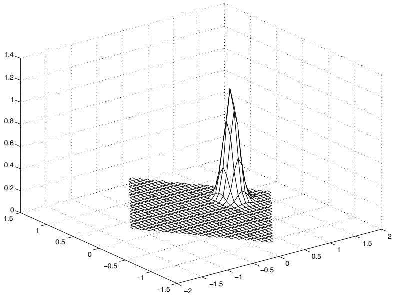


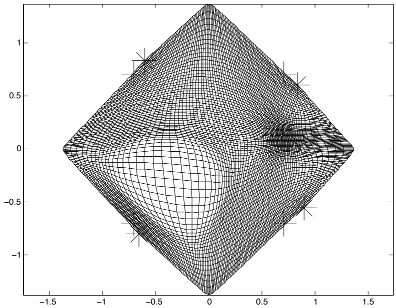
(BV2) numerical state of the art : specific geometries

- The periodic setting : use the ”displacement” change of variable $u \to v = u - \frac{x^2}{2}$.
- Assuming densities with compact support, no easy BCs at infinity (will discuss it again later ...) .
- ”Face to Face” Neuman type BCs : ex. square to square : $u_{x_1}(\pm 1, .) = \pm 1, \quad u_{x_2}(., \pm 1) = \pm 1$.

  This is generalized in Chacon, Delzanno, Finn ...
- BUT it assumes a priori knowledge on the boundary to boundary map : illustration on a square to rhombus map.


1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
0.5
0
−0.5
−1
−1
−0.5
0
0.5
1


1.4
1.2
1
0.8
0.6
0.4
0.2
0
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2


1
0.5
0
−0.5
−1
−1.5
−1
−0.5
0
0.5
1
1.5

Equivalent Hamilton-Jacobi equation on the boundary

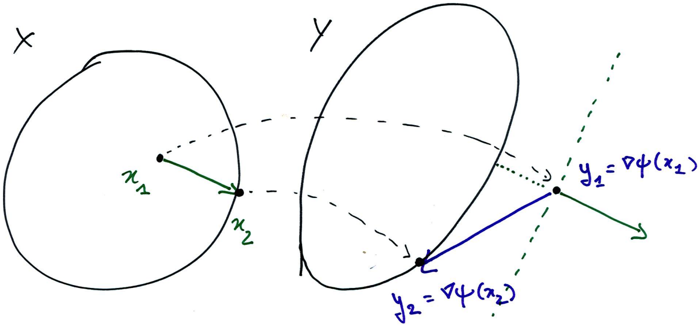
- Use a "*Y* defining function" , (Delanoë, Urbas) : Let $H(y)$ convex such that :

$$\begin{cases} H(y) < 0, & y \in Y, \\ H(y) = 0, & y \in \partial Y, \\ H(y) > 0, & y \in Y^c. \end{cases}$$

- Then $(BV2) \Leftrightarrow (HJ)\ :\ H(\nabla u(x)) = 0,\, x \in \partial X.$
- Proof for Hölder maps probably possible using results by Figalli, Kim, McCann (2011).

Formal proof by contradiction : $H(\nabla u(x)) = 0 \Rightarrow (BV2)$

Y convex + monotonicity



A simple choice for H

- We use the signed Euclidean distance to ∂Y :

$$H(y) = \begin{cases} +\operatorname{dist}(y, \partial Y), & y \in \overline{Y}, \\ -\operatorname{dist}(y, \partial Y), & y \in Y^c. \end{cases}$$

- If $y \in \partial Y$ and n_y the exterior normal at y , then $n_y = \nabla H(y)$.
- Dual formulation of H : using the supporting hyperplane theorem

$$\begin{cases} H(y) = \sup_{||n||=1}\{n \cdot y - H^*(n))\} \\ H^*(n) = \sup_{y_0 \in \partial Y}\{n \cdot y_0\} \end{cases}$$

Obliqueness :

- Let $y = \nabla u(x)$, $x, y \in \partial X, \partial Y$ and n_x, n_y the exterior normals at x, y. We have

$$(OBL) \qquad n_x \cdot n_y \geq 0$$

- Formal proof by contradiction for X convex : If for small $t > 0$, $x + t\, n_y \in X$ then

$$\{\frac{\partial}{\partial t} H(\nabla u(x + t\, ny)\}_{t=0} = (D^2 u(x)\, n_y, \nabla H(y)) > 0$$

(remember $n_y = \nabla H(y)$ and $H(y) \leq 0$ for $y \in \overline{Y}$.)

- On the boundary the sup is attained for n_y the exterior normal to Y at $y = \nabla u(x)$
(Not known a priori) :

$$\begin{aligned} H(\nabla u(x)) &= \sup_{\|n\|=1}\{n \cdot \nabla u(x) - H^*(n))\} \\ &= n_y \cdot \nabla u(x) - H^*(n_y) \end{aligned}$$

- So $(OBL) \Rightarrow$
(n_x is the exterior normal to X au point x)

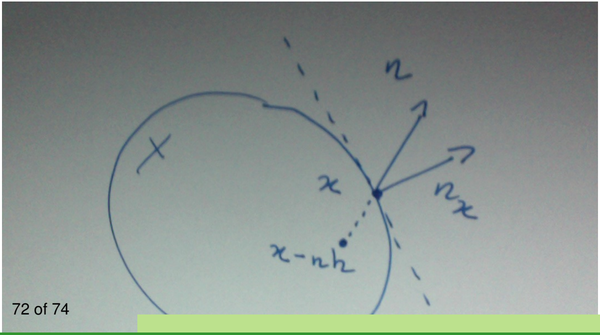
$$H(x, \nabla u(x)) = \sup_{\{\|n\|=1,\, n\cdot n_x>0\}}\{\nabla u(x) \cdot n - H^*(n)\}$$

Good news : Obliqueness $\rightarrow$ monotonicity

$(OBL) \Rightarrow$
(n_x is the exterior normal to X au point x)

$$\begin{aligned} H(x, \nabla u(x)) &= \sup_{\{\|n\|=1,\, n\cdot n_x>0\}}\{\nabla u(x)\cdot n - H^*(n)\} \\ &\approx \sup_{\{\|n\|=1,\, n\cdot n_x>0\}}\{\frac{u(x)-u(x-n\,h)}{h} - H^*(n)\}. \end{aligned}$$

monotone upwind discretization works at the boundary.



Some Refs on numerical methods for MA

JDB, F. Collino and JM Mirebeau, Monotone and Consistent discretization of the Monge-Ampere operator, preprint 2014.

JDB, Brittany D Froese, and A M Oberman. Numerical solution of the Optimal Transportation problem using the MongeAmpre equation. Journal of Computational Physics, 2014.

Xiaobing Feng, Roland Glowinski, and Michael Neilan. Recent Developments in Numerical Methods for Fully Nonlinear Second Order Partial Differential Equations. SIAM Review, 2013.

Susanne Brenner and Michael Neilan. Finite element approximations of the three dimensional Monge-Ampre equation. ESAIM: Mathematical Modelling and Numerical Analysis, 2012.

Brittany D Froese and A M Oberman. Convergent Finite Difference Solvers for Viscosity Solutions of the Elliptic MongeAmpre Equation in Dimensions Two and Higher. SIAM Journal on Numerical Analysis, 2011.

G Loeper and F Rapetti. Numerical solution of the Monge-Ampère equation by a Newtons algorithm. CRAS 2005.

Some Refs on numerical methods

JDB and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numer. Math.,2000.
G. Buttazzo, C. Jimenez, and E. Oudet. An optimization problem for mass transportation with congested dynamics. SIAM J. Control Optim 2009.
Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. Optimal transport with proximal splitting. SIAM Journal on Imaging Sciences, 2014.
G. Carlier, C. Jimenez, and F. Santambrogio. Optimal transportation with traffic congestion and Wardrop equilibria. SIAM J. Control Optim., 2008.
Michel Fortin and Roland Glowinski. Augmented Lagrangian methods, North-Holland Publish- ing Co., Amsterdam, 1983. ...

Augmented Lagrangian methods for Monge's problem, congested transport and Mean-Field Games

Guillaume Carlier [a] **.**

Joint work with Jean-David Benamou (INRIA Rocquencourt),
CRM Applied Math seminar,
Mc Gill, october 15th 2014.

[a]CEREMADE, Université Paris Dauphine

From Benamou-Brenier to MFG

The Benamou-Brenier Formula

Let $c \in C^1(\mathbf{R}^d)$ be a strictly convex transport cost and consider the optimal transport problem

$$W_c(\rho_0, \rho_1) := \inf_{\gamma \in \Pi(\rho_0,\rho_1)} \int_{\overline{\Omega}\times\overline{\Omega}} c(y-x)d\gamma(x,y). \tag{11}$$

Then, at least formally the dynamic formulation of (11) consists in minimizing

$$\int_0^1 \int_{\mathbf{R}^d} c(v_t(x))\rho_t(dx)dt$$

among solutions of the continuity equation

$$\partial_t \rho + \mathrm{div}(\rho v) = 0,\ \rho|_{t=0} = \rho_0,\ \rho|_{t=1} = \rho_1. \tag{12}$$

It is convenient to rewrite this problem in terms of $\sigma(t,x) = (\rho_t(x), m_t(x)) := (\rho_t(x), \rho_t(x)v_t(x)) \in \mathbf{R}^{d+1}$. Indeed, in this case, (12) simply becomes the linear constraint:

$$-\operatorname{div}(\sigma) = f := \delta_1 \otimes \rho_1 - \delta_0 \otimes \rho_0 \tag{13}$$

in the weak sense (and the divergence is of course with respect to t and x). Let us then define

$$E(\sigma) = E(\rho, m) := \begin{cases} c(m/\rho)\rho \text{ if } \rho > 0 \\ 0 \text{ if } \rho = 0 \text{ and } m = 0 \\ +\infty \text{ otherwise.} \end{cases}$$

Note that E is convex, lsc, one homogeneous and incorporates the natural constraints of the transport problem: mass is nonnegative and momentum vanishes where mass does.

The time-dependent formulation of (11) then is:

$$\inf\{\int_0^1 \int_{\mathbf{R}^d} E(\sigma_t(x))dxdt \ : \ -\operatorname{div}(\sigma) = f\}. \tag{14}$$

Benamou-Brenier: quadratic cost

$$E(\rho, \rho v) = \rho|v|^2$$

is the kinetic energy.

Observing that E is the support function of the closed and convex set:

$$K := \{(a,b) \in \mathbf{R} \times \mathbf{R}^d \ : \ a + c^*(b) \leq 0\}$$

problem (14) appears naturally (again, this is slightly formal) as the dual of

$$\inf_{\phi=\phi(t,x)} \{-\langle \phi, f\rangle \ : \ D\phi = (\partial_t \phi, \nabla \phi) \in K\} \tag{15}$$

that is the maximization of
$\int_{\mathbf{R}^d} \phi(1,x) d\rho_1(x) - \int_{\mathbf{R}^d} \phi(0,x) d\rho_0(x)$ among subsolutions of the Hamilton-Jacobi equation with Hamiltonian c^*:

$$\partial_t \phi + c^*(\nabla \phi) \leq 0.$$

Augmented Lagrangian: method and convergence

Finite-dimensional (by finite element) approximation of a model convex variational (static, say) problem lead to

$$\inf_{\phi \in \mathbf{R}^n} J(\phi) := F(\phi) + G(\Lambda\phi) \tag{16}$$

where Λ is a $m \times d$ matrix with real entries, and F: $\mathbf{R}^d \to \mathbf{R} \cup \{+\infty\}$, G: $\mathbf{R}^m \to \mathbf{R} \cup \{+\infty\}$ are two convex lsc and proper functions. We consider the dual of (16):

$$\sup_{\sigma \in \mathbf{R}^m} -F^*(-\Lambda^T\sigma) - G^*(\sigma). \tag{17}$$

A pair $(\overline{\phi}, \overline{\sigma}) \in \mathbf{R}^n \times \mathbf{R}^m$ is said to satisfy the primal-dual extremality relations if:

$$-\Lambda^T \overline{\sigma} \in \partial F(\overline{\phi}), \ \overline{\sigma} \in \partial G(\Lambda \overline{\phi}) \tag{18}$$

which implies that $\overline{\phi}$ solves (16) and that $\overline{\sigma}$ solves (17) as well as the fact that (16) and (17) have the same value (no duality gap).

The primal-dual extremality relations are of course equivalent to finding a saddle-point of the Lagrangian

$$L(\phi, q, \sigma) := F(\phi) + G(q) + \sigma \cdot (\Lambda\phi - q), \quad \forall (\phi, q, \sigma) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \tag{19}$$

in the sense that $(\overline{\phi}, \overline{\sigma}) \in \mathbf{R}^n \times \mathbf{R}^m$ satisfies (18) if and only if $(\overline{\phi}, \overline{q}, \overline{\sigma}) = (\overline{\phi}, \Lambda\overline{\phi}, \overline{\sigma})$ is a saddle-point of L. Now for $r > 0$, we consider the augmented Lagrangian

$$L_r(\phi, q, \sigma) := F(\phi) + G(q) + \sigma \cdot (\Lambda\phi - q) + \frac{r}{2}|\Lambda\phi - q|^2, \tag{20}$$

and we note that being a saddle-point of L is equivalent to being a saddle-point of L_r.

Augmented lagrangian algorithm of Glowinski and Fortin: start with (ϕ^0, q^0, σ^0) and update by:

- **Step 1:** minimization with respect to ϕ:

$$\phi^{k+1} := \text{argmin}_{\phi \in \mathbf{R}^n} \left\{ F(\phi) + \sigma^k \cdot \Lambda\phi + \frac{r}{2} |\Lambda\phi - q^k|^2 \right\} \quad (21)$$

- **Step 2:** minimization with respect to q:

$$q^{k+1} := \text{argmin}_{q \in \mathbf{R}^m} \left\{ G(q) - \sigma^k \cdot q + \frac{r}{2} |\Lambda\phi^{k+1} - q|^2 \right\} \quad (22)$$

- **Step 3:** update the multiplier by the gradient ascent formula

$$\sigma^{k+1} = \sigma^k + r(\Lambda\phi^{k+1} - q^{k+1}). \quad (23)$$

Convergence under very mild conditions:

Theorem 1 *Let $r > 0$, assuming that Λ has full column-rank and that there exists a solution to the primal-dual extremality relations* (18)*, then there exists an* $(\overline{\phi}, \overline{\sigma}) \in \mathbf{R}^n \times \mathbf{R}^m$ *satisfying* (18) *such that the sequence* (ϕ^k, q^k, σ^k) *generated by the ALG2-scheme above satisfies*

$$\phi^k \to \overline{\phi},\ q^k \to \Lambda\overline{\phi},\ \sigma^k \to \overline{\sigma},\quad \text{as } k \to \infty. \tag{24}$$

This theorem is due to Eckstein and Bertsekas (1992) but it follows a long stream of (mainly French) contributions in the 70's: Gabay, Mercier, P.-L. Lions, Glowinski...

Monge's problem

Ω, convex bounded open subset of $\mathbf{R}^d$, two probability measures ρ_0 and ρ_1 on $\overline{\Omega}$, Monge's optimal transport problem (for the euclidean distance) consists in finding the cheapest way to transport ρ_0 to ρ_1 for the euclidean distance. Denoting by $\gamma \in \Pi(\rho_0, \rho_1)$ the set of transport plans between ρ_0 and ρ_1 i.e. the set of probability measures on $\overline{\Omega} \times \overline{\Omega}$ having ρ_0 and ρ_1 as marginals, one thus wishes to solve the Monge-Kantorovich problem:

$$W_1(\rho_0, \rho_1) := \inf_{\gamma \in \Pi(\rho_0, \rho_1)} \int_{\overline{\Omega} \times \overline{\Omega}} |y - x| d\gamma(x, y) \tag{3}$$

whose value $W_1(\rho_0, \rho_1)$ is by definition the 1-Wasserstein distance between ρ_0 and ρ_1.

The well-known Kantorovich duality formula tells us that

$$W_1(\rho_0,\rho_1) = \sup\{\int_{\overline{\Omega}} \phi(\rho_1-\rho_0) \ : \ \phi \text{ 1-Lipschitz } \}. \qquad (4)$$

If one dualizes the constraint in the form $u(x)-u(y) \le |x-y|$, one gets back to (3). If one dualizes the constraint written in a pointwise way as $|\nabla u| \le 1$, (4) also appears as dual to the minimal flow problem:

$$\sup_{\sigma\in L^1(\Omega)} \{-\int_\Omega |\sigma| \ : \ -\operatorname{div}(\sigma) = \rho_1-\rho_0, \sigma\cdot\nu = 0 \text{ on } \partial\Omega \ \} \qquad (5)$$

where the divergence constraint has to be understood in the weak sense:

$$\int_\Omega \nabla u\cdot\sigma = \int_{\overline{\Omega}} u(\rho_1-\rho_0), \forall u \in C^1(\overline{\Omega}).$$

Let ϕ be a 1-Lipschitz potential that solves (4), then by the Kantorovich duality formula a transport plan γ between ρ_0 and ρ_1 is optimal for (3) if and only if

$$\phi(y) - \phi(x) = |x - y| \ \gamma\text{-a.e.}.$$

Which means that the mass at x is transported along a segment on which ϕ grows at maximal rate 1, such rays whose direction is given by the gradient of ϕ are called transport rays and give the direction of optimal transportation in Monge's problem.

Existence of optimal transport maps (i.e. optimal plans given which are induced by a map): Evans-Gangbo, Caffarelli-Feldman-McCann, Ambrosio-Kirchheim, Pratelli, Champion-De Pascale.

An optimal flow field σ for (5) is formally related to an optimal ϕ in (4) by:

$$\nabla\phi = \begin{cases} \frac{\sigma}{|\sigma|} \text{ if } \sigma \neq 0 \\ \text{any vector in the unit ball if } \sigma = 0. \end{cases}$$

so that σ also gives the direction of transport rays, $|\sigma|$ (transport density) measures how much total mass is passing through a given point. One has to be cautious in fact σ is a vector-valued measure (it is L^p when ρ_0 and ρ_1 are: De Pascale-Pratelli, Santambrogio).

Rigorous way to write the optimality conditions in the form of the Monge-Kantorovich PDE system: Bouchitté-Buttazzo.

Applications to static problems

$\Lambda = \nabla$ and ϕ has zero mean.

Monge's problem

Implementing ALG2 for (a finite-element approximation of) Monge's problem is easy, as we shall see. First, the augmented Lagrangian is given by

$$L_r(\phi, q, \sigma) := -\langle f, \phi \rangle + \int_\Omega \chi_B(q) + \langle \sigma, \nabla\phi - q \rangle + \frac{r}{2}\|\nabla\phi - q\|^2_{L^2}$$

where $f = \rho_1 - \rho_0$, B is the closed euclidean unit ball and χ_B its indicator function (0 in B and $+\infty$ outside).

Step 1 of ALG2, then amounts to find first ϕ^{k+1} by solving Laplace's equation

$$-r(\Delta\phi^{k+1} - \mathrm{div}(q^k)) = f + \mathrm{div}(\sigma^k) \text{ in } \Omega \tag{26}$$

together with the Neumann boundary condition

$$r\frac{\partial\phi^{k+1}}{\partial\nu} = rq^k \cdot \nu - \sigma^k \cdot \nu \text{ on } \partial\Omega. \tag{27}$$

Step 2 is explicit since it is a pointwise minimization problem which amounts to set

$$q^{k+1} = p_B\Big(\nabla\phi^{k+1} + \frac{\sigma^k}{r}\Big)$$

where p_B is the projection onto B:

$$p_B(z) = \begin{cases} z \text{ if } |z| \leq 1 \\ \frac{z}{|z|} \text{ otherwise.} \end{cases}$$