Introduction to Optimal Transportation and its applications, part I (prel. version)

Luigi Ambrosio

Scuola Normale Superiore, Pisa
http://cvgmt.sns.it
l.ambrosio@sns.it

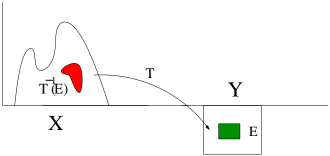
SCUOLA NORMALE SUPERIORE

Outline

SCUOLA NORMALE SUPERIORE

Monge's optimal transport problem

We consider:
– X, Y complete and separable metric spaces;
– $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$;
– a cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$.



We minimize

$$T \mapsto \int_X c(x, T(x))\, d\mu(x)$$

among all *transport maps* T from μ to ν:

$$\mu(T^{-1}(E)) = \nu(E) \qquad \forall E \in \mathscr{B}(Y).$$

Notation: $T : X \to Y$, $T_\sharp : \mathscr{P}(X) \to \mathscr{P}(Y)$,

$$T_\sharp \mu(E) := \mu\big(T^{-1}(E)\big) \qquad \forall E \in \mathscr{B}(Y).$$

$$(S \circ T)_\sharp = S_\sharp \circ T_\sharp, \qquad \int_Y \phi \, dT_\sharp \mu = \int_X \phi \circ T \, d\mu.$$

Monge's problem can be ill-posed because:

1) No admissible T exists: $\mu = \delta_0$, $\nu = (\delta_{-1} + \delta_1)/2$;

2) the infimum is not attained;

3) the constraint $T_\sharp \mu = \nu$ is not weakly sequentially closed.

Lemma 1. *Let $T : X = R^n \to Y = \mathbb{R}^n$ be injective and differentiable out of a $\mathscr{L}^n$-negligible set with $\det \nabla T \neq 0$ a.e. Then*

$$T_\sharp(\rho_1 \mathscr{L}^n) = \rho_2 \mathscr{L}^n \quad \Leftrightarrow \quad \rho_2(T(x))|\det \nabla T(x)| = \rho_1(x) \text{ a.e.}$$

Proof. By a Lusin approximation, the change of variables formula still holds for T, hence

$$\int_X \phi \circ T \rho_1 dx = \int_Y \phi \frac{\rho_1}{|\det \nabla T|} \circ T^{-1} dy.$$

This implies that $\rho_2 = (\rho_1 / |\det \nabla T|) \circ T^{-1}$. $\square$

Remark. If $\det \nabla T = 0$ in a set of positive $\mathscr{L}^n$-measure, then $T_\sharp(\rho_1 \mathscr{L}^n)$ is singular with respect to $\mathscr{L}^n$, with a singular part having mass $\int_{\{\det \nabla T = 0\}} \rho_1 dx$.

Kantorovich's formulation of optimal transportation

We minimize

$$\pi \mapsto \int_{X\times Y} c(x,y)\, d\pi(x,y)$$

in the set $\Gamma(\mu,\nu)$ of all *transport plans* $\pi \in \mathscr{P}(X \times Y)$ from μ to ν:

$$\pi(A\times Y) = \mu(A) \quad \forall A \in \mathscr{B}(X), \qquad \pi(X\times B) = \nu(B) \quad \forall B \in \mathscr{B}(Y).$$

Equivalently: $(\pi_X)_\sharp \pi = \mu$, $(\pi_Y)_\sharp \pi = \nu$.

SCUOLA NORMALE SUPERIORE

Heuristic meaning of π:

$$\pi(A \times B) = \text{the mass initially at } A \text{ sent at } B$$

Advantages:
1) $\Gamma(\mu, \nu)$ is not empty (it contains $\mu \times \nu$);
2) The set $\Gamma(\mu, \nu)$ is weakly closed in $\mathscr{P}(X \times Y)$, and $\pi \mapsto \int c \, d\pi$ is linear;
3) Solutions always exist under mild assumptions on c.
4) Transport plans "include" transport maps, since $T_\sharp \mu = \nu$ implies that $\pi := (Id \times T)_\sharp \mu$ belongs to $\Gamma(\mu, \nu)$. If, c is real-valued, then

$$\inf \text{(Monge)} = \inf \text{(Kantorovich)}.$$

SCUOLA NORMALE SUPERIORE

Theorem 2. *If c is lower semicontinuous, then (K) has a solution.*
Notation. We shall use the notation $\Gamma_0(\mu,\nu)$ for *optimal* plans.
Proof. $\pi \mapsto \int c\, d\pi$ is sequentially l.s.c. with respect to convergence in duality with $C_b(X \times Y)$, since

$$\exists c_n \in C_b(X \times Y), \qquad c_n \uparrow c.$$

On the other hand, by Ulam theorem, there exist nondecreasing sequences of compact sets $C_n \subset X$ and $K_n \subset Y$ such that $\mu(X \setminus \cup_n C_n) = 0$ and $\nu(Y \setminus \cup_n K_n) = 0$. It turns out that

$$\pi(X \times Y \setminus C_n \times K_n) \leq \mu(X \setminus C_n) + \nu(Y \setminus K_n) \qquad \forall \pi \in \Gamma(\mu,\nu),$$

hence $\Gamma(\mu,\nu)$ is a *tight* family in $\mathscr{P}(X \times Y)$. Prokhorov theorem ensures the sequential relative compactness w.r.t. the weak topology of $\Gamma(\mu,\nu)$. □

SCUOLA NORMALE SUPERIORE

Can we recover an optimal T from an optimal π ?

Why should we look for (optimal) transport maps ? Optimal transport provides a canonical way to “rearrange” a mass distribution into another.

c-monotonicity

Definition. We say that $\Gamma \subset X \times Y$ is *c-monotone* if $(x_i, y_i) \in \Gamma$, $1 \le i \le n$, implies

$$\sum_{i=1}^{n} c(x_i, y_i) \le \sum_{i=1}^{n} c(x_{\sigma(i)}, y_i) \qquad \text{for all permutations } \sigma.$$

If $X = Y = H$ and $c(x, y) = |x - y|^2/2$, this concept reduces to the classical *cyclical monotonicity*. Indeed, expanding the squares, one obtains

$$\sum_{i=1}^{n} \langle y_i, x_{\sigma(i)} - x_i \rangle \le 0.$$

Remark. For general cost functions the c-monotonicity is much harder, if not impossible, to characterize, and it should be taken as it is. It is much better to extend the concepts of duality, subdifferential, etc. to general cost c.

For $F : H \to \mathbb{R} \cup \{+\infty\}$ and $x \in D(F) = \{F < +\infty\}$ the *subdifferential* $\partial F(x)$ is defined by

$$\partial F(x) := \{v \in H : \ F(y) \geq F(x) + \langle v, y - x\rangle \ \forall y \in H\}.$$

Theorem 3. (Rockafellar) *$\Gamma \subset H \times H$ is cyclically monotone iff it is contained in the graph of ∂F for some l.s.c. proper $F : H \to \mathbb{R} \cup \{+\infty\}$.*

The easy implication: add the inequalities

$$\langle y_i, x_{\sigma(i)} - x_i\rangle \leq F(x_{\sigma(i)}) - F(x_i)$$

to get

$$\sum_{i=1}^{n} \langle y_i, x_{\sigma(i)} - x_i\rangle \leq 0.$$

The converse implication requires a construction that we will see later on, in a more general context.

Theorem 4. *Assume c l.s.c., π optimal and $\int c d\pi$ finite. Then π is concentrated on a Borel c-monotone set. The converse holds if*

$$(*) \qquad c(x,y) \le a(x) + b(y), \qquad a \in L^1(\mu),\ b \in L^1(\nu).$$

Remark. Condition (*) is natural in some, but not all, problems. For instance problems with constraints or in Wiener spaces include $+\infty$-valued costs.
Since Theorem 4 is easy to show for discrete measures, a possibility is to work by approximation. But, this provides only the existence of *some* optimal π concentrated on a c-monotone set (this would suffices for the proof of duality).
Following Gangbo-McCann, we follow a slightly different strategy.

SCUOLA NORMALE SUPERIORE

Proof. (only necessity, in the case $c \in C_b(X \times Y)$). Assume, by contradiction, that the cyclical monotonicity condition fails for some $\{(x_i, y_i)\}_{1 \le i \le n} \subset \operatorname{supp} \pi$ and some permutation σ. By continuity we can find neighbourhoods $U_i \ni x_i$, $V_i \ni y_i$ with

$$\sum_{i=1}^{n} c(u_i, v_{\sigma(i)}) - c(u_i, v_i) < 0 \qquad \forall (u_i, v_i) \in U_i \times V_i,\ 1 \le i \le n.$$

Our goal is to build a "variation" $\hat{\pi} = \pi + \sigma$ of π in such a way that minimality of π is violated.To this aim, we need a *signed* measure σ with:

(i) $\sigma^- \le \pi$ (so that $\hat{\pi}$ is nonnegative);
(ii) null first and second marginal (so that $\hat{\pi} \in \Gamma(\mu, \nu)$);
(iii) $\int c \, d\sigma < 0$ (so that π is not optimal).

SCUOLA NORMALE SUPERIORE

If we are given measures $\lambda_1, \dots, \lambda_n$ in a measurable space Z, we know the existence of a probability space (Ω, P) and measurable maps $h_i : \Omega \to Z$ with $h_{i\sharp} P = \lambda_i$ (for instance one may take $\Omega = Z^n$ with the product measure).
Apply this construction to the measures $\lambda_i = \frac{1}{m_i} \chi_{U_i \times V_i} \pi$, with $m_i = \pi(U_i \times V_i)$, to find maps $h_i = (f_i, g_i) : \Omega \to U_i \times V_i$ such that

$$(f_i, g_i)_\sharp P = \lambda_i \le \frac{1}{m_i} \pi.$$

If we define

$$\sigma := \frac{\min_j m_j}{n} \sum_{i=1}^{n} \big[(f_i, g_{\sigma(i)})_\sharp P - (f_i, g_i)_\sharp P\big],$$

obviously (i) and (ii) above are fulfilled, and

$$\int c \, d\sigma = \frac{\min_j m_j}{n} \int_\Omega \sum_{i=1}^{n} [c(f_i, g_{\sigma(i)}) - c(f_i, g_i)] \, dP < 0.$$

Theorem 4 leads in a natural way to the analysis of the properties of *c*-monotone sets, to see how far are they from being graphs. Indeed:

Lemma 5. *Assume that* Γ *is a* π*-measurable graph and that* $\pi \in \Gamma(\mu, \nu)$ *is concentrated on* Γ*. Then* π *is induced by a transport* T*, i.e.*

$$\pi = (Id \times T)_\sharp \mu.$$

Proof. Let $T : \pi_X(\Gamma) \to Y$ be the map whose graph is Γ and let $K_n \subset \Gamma$ compact with $\pi(\Gamma \setminus K_n) \to 0$, $F_n := \pi(K_n)$.
Since $\mu(F_n) = \gamma(F_n \times Y) \geq \gamma(K_n) \to 1$, the union F of F_n covers μ-almost all of X. On the other hand, $T : F_n \to Y$ is continuous (since its graph K_n is closed), hence T is μ-measurable.
Since $y = T(x)$ π-a.e. in $X \times Y$ we conclude that

$$\int \phi(x,y)\, d\pi(x,y) = \int \phi(x, T(x))\, d\pi(x,y) = \int \phi(x, T(x))\, d\mu(x),$$

so that $\pi = (Id \times T)_\sharp \mu$. $\square$

The dual formulation

Developed by Kantorovich, Kellerer, Levin, Rüschendorf,...

Theorem 6. *Assume that $c : X \times Y \to [0, +\infty]$ is l.s.c. Then*

$$\min(K) = \sup\left\{\int_X \varphi\, d\mu + \int_Y \psi\, d\nu :\ (\varphi,\psi) \in L^1_\mu \times L^1_\nu,\ \varphi + \psi \le c\right\}.$$

The inequality $\ge$ is trivial:

$$\int_X \varphi\, d\mu + \int_Y \psi\, d\nu = \int_{X\times Y} (\varphi+\psi)\, d\pi \le \int_{X\times Y} c\, d\pi \qquad \forall \pi \in \Gamma(\mu,\nu).$$

We prove the opposite inequality in the simplest case, namely when $c \in C_b(X \times Y)$. Fix $\pi \in \Gamma(\mu,\nu)$ optimal and a c-monotone set Γ on which π is concentrated. For $\phi : X \to \mathbb{R} \cup \{-\infty\}$ we define the *c-transform of ϕ* by

$$\phi^c(y) := \inf_{x\in X} c(x,y) - \phi(x).$$

SCUOLA NORMALE SUPERIORE

Obviously $\varphi + \varphi^c \le c$ (φ^c is the largest with this property) and we conclude that duality holds, with $\psi := \varphi^c$, if we are able to build φ in such a way that

$$\varphi + \varphi^c = c \quad \text{on } \Gamma \text{ (and, in particular, } \pi\text{-a.e.)}$$

To do this, we adapt the construction of Rockafellar, defining

$$\varphi(x) := \inf\Big\{ c(x, y_n) - c(x_n, y_n) + c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1}) + \cdots + c(x_1, y_0) - c(x_0, y_0) \Big\},$$

where $(x_0, y_0) \in \Gamma$ is fixed and the infimum runs among all finite families $\{(x_i, y_i)\}_{1 \le i \le n} \subset \Gamma$. The definition of φ immediately gives

$$(*) \quad \varphi(x) - c(x, y) \le \varphi(x') - c(x', y) \qquad \forall (x, y) \in \Gamma,\ x' \in X$$

hence $\varphi^c(x) = c(x, y) - \varphi(x)$ on Γ. However, we need to check that both φ and φ^c are integrable! Monotonicity of Γ implies that $\varphi(x_0) = 0$ and this, in combination with (*), yields $\varphi \in L^\infty(X)$, $\varphi^c \in L^\infty(Y)$. $\square$

Regularity of the optimal φ

Remark. We have seen that the "optimal" potential φ is *c-concave*, namely it can be represented as an inf-convolution:

$$\varphi(x) = \inf_{i \in I} c(x, y_i) - m_i \qquad \forall x \in X.$$

In the case $c(x, y) = |x - y|^2/2$, c-concavity is equivalent to semiconcavity, more precisely

$$\varphi(x) - \frac{|x|^2}{2} \qquad \text{is concave.}$$

In the case $c(x, y) = \mathrm{distance}$, c-concavity is equivalent to 1-Lipschitz continuity with respect to c.

SCUOLA NORMALE SUPERIORE

Back to transport maps

If (φ,ψ) is a maximizing pair and $\pi \in \Gamma_0(\mu,\nu)$ has finite cost, "min = max" leads to the *optimality condition*:

$$\varphi(x)+\psi(y)=c(x,y) \qquad \pi\text{-a.e. in } X\times Y.$$

If $X=Y=H$, $c(x,y)=|x-y|^2/2$, and $(x,y)\in \operatorname{supp}\pi$, then

$$x'\mapsto \frac{1}{2}|x'-y|^2-\varphi(x')$$

attains its minimum, equal to $\psi(y)$, at $x'=x$. As a consequence

$$y=x-\nabla\varphi(x)$$

provided φ is differentiable at x.

Key point: π is concentrated on a graph if we have differentiability, at least μ-a.e., of the maximizing dual Kantorovich potential φ. If $H=\mathbb{R}^n$, since φ is semiconcave, a result by Zajicek shows that φ is differentiable out of the union of countably many Lipschitz hypersurfaces (and in particular $\mathscr{L}^n$-a.e.).

SCUOLA NORMALE SUPERIORE

Notation. $\mathscr{P}_2(H) := \{\mu \in \mathscr{P}(H) : \int_H |x|^2\,d\mu < \infty\}$.

Theorem 7. (Brenier, Knott-Smith, Rüschendorf) *Assume $X = Y = \mathbb{R}^n$, $c(x,y) = |x-y|^2/2$, $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^n)$. If $\mu \ll \mathscr{L}^n$ or, more generally, μ vanishes on Lipschitz hypersurfaces, then:*

(i) *there exists a unique optimal transport map T;*

(ii) *the map is the gradient of a convex function φ.*

Proof. The assumption $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^n)$ ensures finiteness of the minimal cost. We have seen that any optimal π is concentrated on the graph of the function $x \mapsto T(x) := x - \nabla\varphi(x)$; it is the gradient of the convex map $\frac{1}{2}|x|^2 - \varphi(x)$. This proves that $\pi = (Id \times T)_\sharp\mu$ is uniquely determined by φ, so uniqueness holds even in the larger class of transport plans. □

Writing $\mu = \rho_1\mathscr{L}^n$, $\nu = \rho_2\mathscr{L}^n$, the transport condition $\rho_2(T(x))|\det\nabla T|(x) = \rho_1(x)$ gives that φ is a *pointwise* solution to the Monge-Ampere equation

$$\det\nabla^2\varphi(x) = \frac{\rho_1(x)}{\rho_2(\nabla\varphi(x))}.$$

SCUOLA NORMALE SUPERIORE

Regularity of the optimal map

Alexandrov theorem. (A-Alberti, '99) *Let* $\Gamma : \mathbb{R}^n \to P(\mathbf{R}^n)$ *be a monotone operator. Then*

(i) *for* $\mathscr{L}^n$*-a.e.* $x \in D(\Gamma)$, $\Gamma(x) = \{p(x)\}$ *is single-valued;*

(ii) *for* $\mathscr{L}^n$*-a.e.* $x \in D(\Gamma)$ *there exists* $A(x) \in \mathbf{R}^{n \times n}$ *with*

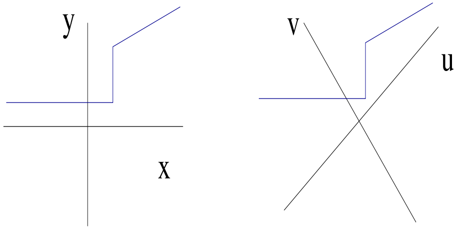
$$\lim_{q \in \Gamma(y),\, y \to x} \frac{|q - p(x) - A(x)(y - x)|}{|y - x|} = 0.$$

If $\Gamma(x) = \partial\varphi(x)$*, with* $\varphi(x)$ *convex and l.s.c., then* $A \in \mathrm{Sym}_+^{n \times n}$ $\mathscr{L}^n$*-a.e.*

It follows by a canonical 1-1 correspondence between monotone operators and 1-Lipschitz maps. Under this translation Alexandrov theorem corresponds to Rademacher differentiation theorem of Lipschitz maps.

SCUOLA NORMALE SUPERIORE

Minty correspondence



$$\begin{cases} u = \frac{y+x}{\sqrt{2}} \\ v = \frac{y-x}{\sqrt{2}} \end{cases}$$

SCUOLA NORMALE SUPERIORE

Regularity the optimal map

In general the optimal transport map is only BV (and discontinuities can occur), unless geometric restrictions are imposed on $\operatorname{supp}\nu$.

Regularity theorem. (Caffarelli) *Assume V convex, $\ln \rho_1, \ln \rho_2 \in L^\infty(V)$. Then $\varphi \in C^{1,\alpha}(V)$ for all $\alpha < 1$ and*

$$\rho_1,\ \rho_2 \in C^{0,\alpha}(V) \qquad \Longrightarrow \qquad \varphi \in C^{2,\alpha}(V).$$

The convexity assumption on V is needed to show that φ is a *viscosity* solution to (MA), and then the regularity theory for (MA), developed by Caffarelli and Urbas, applies.

Open problem. Under the assumption of the regularity theorem, we know $T = \nabla\varphi$ is Hölder continuous and BV. Can we say that $T \in W^{1,1}(V;\mathbb{R}^n)$?

A positive solution would give existence of solutions, in the physical (velocity) variables to a semi-geostrophic PDE in fluid mechanics (Brenier, Cullen-Gangbo, Cullen-Feldman).

Application: polar factorization

Let $D \subset \mathbb{R}^n$ be a bounded domain, denote by μ_D the normalized Lebesgue measure on D and consider the space

$$S(D) := \{s : D \to D : \ s_\sharp \mu_D = \mu_D\}.$$

The following result provides a kind of (nonlinear) projection on the (nonconvex) space $S(D)$.

Polar factorization theorem. *Let $f \in L^2(\mu_D; \mathbb{R}^n)$ be satisfying the non-degeneracy condition $f_\sharp \mu_D \ll \mathscr{L}^n$. Then there exist $s \in S(D)$ and φ convex such that $f = (\nabla\varphi) \circ s$.*

Remark. Compare with the Helmholtz projection $f = \nabla\varphi + g$, g divergence-free.

Proof. Let T be the optimal transport map from $f_\sharp \mu_D \in \mathscr{P}_2(\mathbb{R}^n)$ to μ_D. Then $s = T \circ f$ is measure-preserving and

$$f = T^{-1} \circ s = (\nabla\varphi) \circ s.$$

SCUOLA NORMALE SUPERIORE

Extensions of Brenier's result

– Infinite-dimensional Hilbert spaces, A-Gigli-Savaré.
– compact Riemannian manifolds (M, g), $c = d_M^2/2$: McCann.
– cost functions induced by Lagrangians Bernard-Buffoni, namely

$$c(x, y) := \inf\left\{\int_0^1 L(t, \gamma(t), \dot\gamma(t))\,dt:\ \gamma(0) = x,\ \gamma(1) = y\right\};$$

– Carnot groups and sub-Riemannian manifolds, $c = d_{CC}^2/2$: A-Rigot, Figalli-Rifford;
– cost functions induced by sub-Riemannian Lagrangians Agrachev-Lee.

A common framework is to consider measures $\boldsymbol{\eta}$ in the space Ω of paths in M, minimizing an action

$$\int_\Omega \mathcal{A}(\omega)\,d\boldsymbol{\eta}(\omega)$$

with the constraint $(e_0)_\sharp\boldsymbol{\eta} = \mu$, $(e_1)_\sharp\boldsymbol{\eta} = \nu$, where $e_t : \Omega \to M$ are the evaluation maps, $e_t(\omega) = \omega(t)$.

SCUOLA NORMALE SUPERIORE

Extensions of Brenier's result

– Wiener spaces (E, H, γ), Feyel-Üstünel. Here E is a Banach space, $\gamma \in \mathscr{P}(E)$ is Gaussian and H is its Cameron-Martin space, namely

$$H := \{h \in E : \ (\tau_h)_\sharp \gamma \ll \gamma\} .$$

In this case

$$c(x, y) := \begin{cases} \dfrac{|x-y|_H^2}{2} & \text{if } x - y \in H; \\ +\infty & \text{otherwise.} \end{cases}$$

This model is extremely interesting: $c = +\infty$ for "most" pairs (x, y) (since $\gamma(H) = 0$), nevertheless we have the Talagrand transport inequality

$$\inf \left\{ \int_E c(x, T(x))\, d\gamma(x) : \ T_\sharp \gamma = \rho\gamma \right\} \leq \int \rho \ln \rho \, d\gamma.$$

SCUOLA NORMALE SUPERIORE

Optimal maps on manifolds

Theorem. *M compact Riemannian manifold with no boundary. $c = \frac{1}{2}d^2$. Then, for $\mu, \nu \in \mathscr{P}(M)$ with $\mu \ll \mathrm{vol}_M$ there exists a unique optimal transport map.*

Here, the extra difficulty is that $d^2(x, y)$ need not be differentiable. However, a little bit of nonsmooth analysis solves the problem: first, the differentiation argument gives

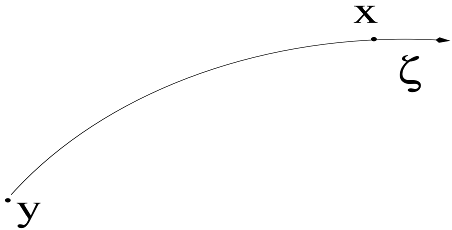
$$\nabla\varphi(x) \in \partial_F \frac{1}{2}d^2(\cdot, y)(x) = d(x, y)\partial_F d(\cdot, y)(x).$$

Here ∂_F is the Frechet subdifferential), namely

$$\partial\phi(x) := \{v \in T_x M : \ \phi(\exp_x w) \geq \phi(x) + g_x(v, w) + o(|w|)\} .$$

SCUOLA NORMALE SUPERIORE

Optimal maps on manifolds



On the other hand, if $\xi \in T_x M$ is the final velocity of any unit speed *minimizing* geodesic from y to x, it is easy to check that

$$-\xi \in \partial_F^+ d(\cdot, y)(x)$$

(here ∂_F^+ is the Frechet superdifferential).

Optimal maps on manifolds

Since $d(\cdot, y)$ is both sub- and super- differentiable at x, it is differentiable and $-\xi d(x, y) = \nabla\varphi(x)$. We conclude that

$$y = \exp_x(-\xi d(x, y)) = \exp_x\big(\nabla\varphi(x)\big) =: T(x)$$

is uniquely determined by x and we recover the optimal map T.

We know actually more: there exists a *unique* geodesic from x to y, so that the optimal transport map does not go beyond the cut locus!

Regularity of the optimal transport map, cut and focal locus: Ma-Trudinger-Wang, Loeper, Loeper-Villani, Figalli-Rifford-Villani.

Cost=distance

In the case cost=Euclidean distance, by differentiating $|x' - y| - \varphi(x')$ at $x' = x$, we get

$$\nabla\varphi(x) = \frac{x - y}{|x - y|}.$$

Hence, the information is only on the direction of transportation, *not* on transportation length.
If $c(x, y) = \|x - y\|$ with $\|\cdot\|$ not strictly convex, even this information is (partially) lost:

$$x - y \in (d_x\varphi)^* := \{v \in \mathbb{R}^n : \ |d_x\varphi(v)| = \|v\|\}.$$

In order to attack this problem one has to get the missing pieces of information by a perturbation argument (for instance, in the case cost=Euclidean distance, considering the costs

$$c_\epsilon(x, y) := |x - y|^{1+\epsilon}, \qquad \epsilon \downarrow 0).$$

SCUOLA NORMALE SUPERIORE

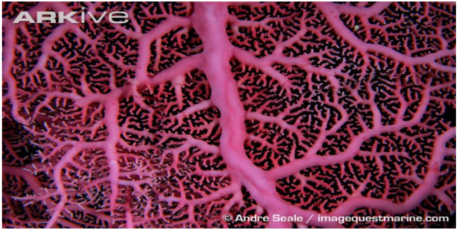
Cost=distance

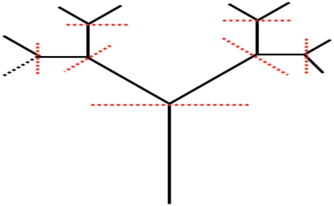
The analysis of this problem is related to deep questions in Real Analysis and Probability (regularity of disintegrations, Nikodym sets,..).

– '78 Sudakov, any norm (?);
– '96 Evans-Gangbo, Euclidean norm;
– '00 Caffarelli-Feldman-McCann, A, Trudinger-Wang, C^2 and uniformly round norms;
– '03 A-Kirchheim-Pratelli, crystalline norms;
– '08 Champion-DePascale, Caravenna strictly convex norms;
– '09 General norms: Champion-DePascale, Caravenna, Caravenna-Daneri.

Branched optimal transportation and irrigation models

The optimal transportation problem, in its classical formulation, is not realistic for some real life problems where networks, i.e. branched transportation structures, are expected.





SCUOLA NORMALE SUPERIORE

E.N.Gilbert: *Minimum cost communication networks.* Bell System Tech. J., 1967.

Q.Xia: *Optimal paths related to transport problems.* Comm. Contemp. Math., 2003.

M.Bernot-V.Caselles-J.M.Morel: *Optimal transportation networks – models and theory.* Springer, LNM **1955**, 2009.

SCUOLA NORMALE SUPERIORE

In shipping two items from nearby cities to the same far away city, it may be less expensive to first bring them into a common location and put them on a single truck for most of the transport. In this way a "Y shaped" path is preferable to a "V shaped" path.

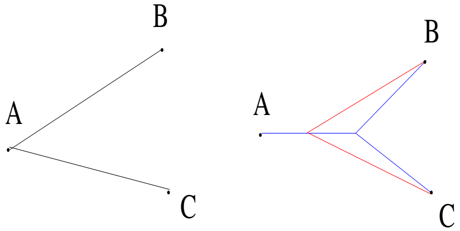


Figure: Transport from δ_A to $(\delta_B + \delta_C)/2$: MK versus Gilbert solution

SCUOLA NORMALE SUPERIORE

Gilbert's model

We consider an oriented finite tree $T \subset \mathbb{R}^n$, with initial nodes N_i, final nodes N_f and a nonnegative multiplicity function $w(e)$ defined on edges satisfying Kirchoff's rule

$$\text{(Kir)} \qquad \sum_{e \text{ incoming at } x} w(e) = \sum_{e \text{ outgoing from } x} w(e)$$

at any "internal" node x of T.
The measures

$$\begin{cases} \mu(A) := \sum_{x \in A \cap N_i} \sum_{e \text{ outgoing from } x} w(e) \\ \nu(B) := \sum_{y \in B \cap N_f} \sum_{e \text{ incoming at } y} w(e) \end{cases}$$

play the role of "initial" and "final" measures and, by (Kir), they have the same total mass.

Then, for $\alpha \in [0, 1]$, one minimizes the cost function

$$\mathcal{E}_\alpha(T) := \sum_{e \in T} w^\alpha(e)\mathrm{length}(e).$$

among all trees starting from μ to ν.

Limit cases. $\alpha = 0$: Steiner's problem.
$\alpha = 1$: Monge-Kantorovich problem, cost=distance.

As soon as $\alpha < 1$ a branched transportation structure does appear, and another critical parameter is $\alpha = 1 - 1/n$.

SCUOLA NORMALE SUPERIORE

Continuous formulation

We consider "continuous" trees, i.e. countably $\mathscr{H}^1$-rectifiable sets Γ, oriented by a tangent vector field $\tau : \Gamma \to \mathbf{S}^{n-1}$ and an integrable multiplicity function $w : \Gamma \to [0, +\infty)$.

Canonically associated with (Γ, τ, w) is the $\mathbb{R}^n$-valued measure

$$J_{\Gamma,\tau,w} := w\tau\mathcal{H}^1 \llcorner \Gamma \quad \left(\text{i.e.} \int \phi \, dJ_{\Gamma,\tau,w} = \int_\Gamma w\phi\tau \, d\mathscr{H}^1\right).$$

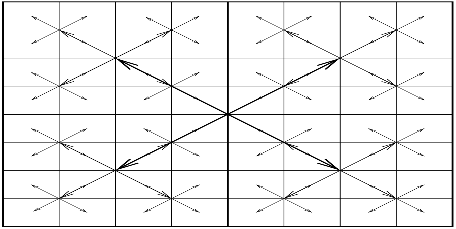
Both the initial/final condition and (Kir) become $\nabla \cdot J = \nu - \mu$, and

$$\mathcal{E}_\alpha(J) := \int_\Gamma w^\alpha \, d\mathscr{H}^1.$$

SCUOLA NORMALE SUPERIORE

Existence and regularity results

Theorem. (Xia, Maddalena-Solimini, Bianchini-Brancolini) *For all $\alpha \in [0,1)$ there exists a continuous tree with minimal cost among those connecting μ to ν. If $\mu = \delta_z$ and $\nu = \mathscr{L}^n \llcorner [0,1]^n$, the minimal cost is finite if and only if $\alpha > 1 - 1/n$.*



Remark. If $\alpha > 1 - 1/n$, an "irrigation" T with finite $\mathcal{E}_\alpha$ cost can be achieved by a dyadic splitting procedure, i.e. first reaching the center of $[0,1]^n$, then the centers of the 2^n subcubes, and so on.

$$\mathcal{E}_\alpha(T) \sim \sum_{k=1}^{\infty} 2^{nk} \cdot 2^{-\alpha nk} \cdot 2^{-k} = \sum_{k=1}^{\infty} 2^{k(n-\alpha n-1)}.$$

Existence and regularity results

Theorem. (Xia, Bernot-Caselles-Morel) *Let* $\alpha \in [0, 1)$ *and let* (Γ, τ, w) *be a continuous tree with minimal* $\mathcal{E}_\alpha$ *cost between* μ *and* ν*. Then* Γ *is locally a finite tree away from* $\operatorname{supp}\mu \cup \operatorname{supp}\nu$*.*

Finally, a Lagrangian formulation involves the minimization of the *nonlocal* action

$$\mathscr{A}(\boldsymbol{\eta}) := \int_\Omega \int_0^{T(\gamma)} [\omega(t)]_{\boldsymbol{\eta}}^{\alpha-1} |\dot{\omega}(t)| \, dt \, d\boldsymbol{\eta}(\gamma),$$

where

$$[x]_{\boldsymbol{\eta}} := \boldsymbol{\eta} \{\omega : \ x \in \omega([0, +\infty[)\} .$$

SCUOLA NORMALE SUPERIORE

Variational models in incompressible Euler equations

We consider an incompressible fluid moving inside a d-dimensional region D with velocity $\boldsymbol{u}$. The Euler equations for $\boldsymbol{u}$ are

$$\begin{cases} \partial_t \boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\nabla p & \text{in } [0,T]\times D, \\ \nabla\cdot\boldsymbol{u} = 0 & \text{in } [0,T]\times D, \\ \boldsymbol{u}\cdot n = 0 & \text{on } [0,T]\times \partial D, \end{cases}$$

where p, the pressure field, is a Lagrange multiplier for the divergence-free constraint.

If $\boldsymbol{u}$ is smooth, it produces a unique flow map g, given by

$$\begin{cases} \dot{g}(t,a) = \boldsymbol{u}(t, g(t,a)), \\ g(0,a) = a. \end{cases}$$

By the incompressibility condition, we get that $g(t,\cdot) : D \to D$ is a measure-preserving diffeomorphism of D, $g(t,\cdot)_\sharp \mu_D = \mu_D$.

Writing Euler's equations in terms of g, we get

$$\begin{cases} \ddot{g}(t,a) = -\nabla p\,(t, g(t,a)) & (t,a) \in [0,T] \times D, \\ g(0,a) = a & a \in D, \\ g(t,\cdot) \in \mathrm{SDiff}(D) & t \in [0,T]. \end{cases}$$

We can (formally) view the space $\mathrm{SDiff}(D)$ of measure-preserving diffeomorphisms of D as an infinite-dimensional manifold with the metric inherited from the embedding in $L^2(D)$, and with tangent space made by the divergence-free vector fields.

SCUOLA NORMALE SUPERIORE

Arnold's geodesic interpretation

Using this viewpoint, Arnold interpreted the previous ODE, and therefore Euler's equations, as a *geodesic* equation on $\mathrm{SDiff}(D)$. Therefore one can look for solutions of Euler's equations by minimizing

$$\int_0^1 \int_D \frac{1}{2} |\dot{g}(t,x)|^2 \, d\mu_D(x) \, dt$$

among all paths $g(t,\cdot) : [0,1] \to \mathrm{SDiff}(D)$ with $g(0,\cdot) = f$ and $g(1,\cdot) = h$ prescribed (typically, by right invariance, f is taken as the identity map $\boldsymbol{i}$).

Existence: Ebin-Marsden (1970), $g \circ f^{-1} \sim \boldsymbol{i}$.

In general, as pointed out by Shnirelman, geodesics (and even curves with finite length if $d = 2$) need not exist.

SCUOLA NORMALE SUPERIORE

Relaxed formulation

These results led in 1989 Brenier to the following model: we minimize the action functional

$$\mathscr{A}(\boldsymbol{\eta}) := \int_{\Omega(D)} \frac{1}{2} \int_0^1 |\dot{\omega}|^2 \, dt \, d\boldsymbol{\eta}(\omega), \qquad \boldsymbol{\eta} \in \mathscr{P}(\Omega(D))$$

with the constraints

$$(e_0, e_1)_\sharp \boldsymbol{\eta} = (\boldsymbol{i} \times h)_\sharp \mu_D, \qquad (e_t)_\sharp \boldsymbol{\eta} = \mu_D \ \ \forall t \in [0, T].$$

Classical flows $g(t, a)$ induce generalized ones via the relation $\boldsymbol{\eta} = (\Phi_g)_\sharp \mu_D$, with

$$\Phi_g : D \to \Omega(D), \qquad \Phi_g(a) := g(\cdot, a).$$

Advantages: existence of solutions, uniqueness of the pressure field (identified through a suitable dual formulation). An adaptation of this approach produces a complete length distance in the space $\Gamma(\mu_D, \mu_D)$ of *measure-preserving plans*. First variation also leads to a weak formulation of Euler's equations

$$\partial_t \overline{\boldsymbol{v}}_t(x) + \nabla \cdot (\overline{\boldsymbol{v} \otimes \boldsymbol{v}}_t(x)) + \nabla_x p(t,x) = 0.$$

Here

$$\overline{\boldsymbol{v}}_t \mu_D = (e_t)_\sharp(\dot\omega(t)\boldsymbol{\eta}), \qquad \overline{\boldsymbol{v} \otimes \boldsymbol{v}}_t \mu_D = (e_t)_\sharp(\dot\omega(t) \otimes \dot\omega(t)\boldsymbol{\eta}).$$

In general, however, $\overline{\boldsymbol{v} \otimes \boldsymbol{v}}_t \neq \overline{\boldsymbol{v}}_t \otimes \overline{\boldsymbol{v}}_t$ (due to branching and multiple velocities), and this precisely marks the difference between genuine distributional solutions to Euler's equation and "generalized" ones (in analogy with the DiPerna-Majda weak solutions).

Connection with optimal transport models

The following result provides a regularity condition on the pressure field (improving Brenier's condition $\nabla p \in \mathcal{M}_{\mathrm{loc}}\big((0,T)\times D\big)$) and *necessary* optimality conditions.

Theorem (A-Figalli) *$\nabla p \in L^2_{\mathrm{loc}}\big((0,T);\mathcal{M}_{\mathrm{loc}}(D)\big)$, so that $p(t,\cdot)$ is $BV_{\mathrm{loc}}(D)$, and any minimizer $\boldsymbol{\eta}$ is concentrated on curves locally minimizing in $(0,T)$ the action*

$$\mathcal{L}_p^{st}(\gamma) := \frac{1}{2}\int_s^t |\dot\omega(r)|^2 - p(r,\gamma(r))\,dr.$$

We also provide *necessary and sufficient* optimality conditions, whose formulation is however more involved.

SCUOLA NORMALE SUPERIORE