Introduction to Optimal Transportation and its applications, part II (prelim. version)

Luigi Ambrosio

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

Outline

1. The metric side of optimal transportation

2. Convex functionals in $\mathscr{P}_2(H)$

3. The differentiable side of optimal transportation

SCUOLA NORMALE SUPERIORE

The Kantorovich-Rubinstein-Wasserstein distance

(X, d) metric space.

$$\mathscr{P}_2(X) := \left\{ \mu \in \mathscr{P}(X) : \int_X d^2(x_0, x)\, d\mu(x) < \infty \quad \forall x_0 \in X \right\}.$$

Set

$$W_2^2(\mu, \nu) := \min \left\{ \sqrt{\int_{X \times X} d^2(x, y)\, d\pi} : \pi \in \Gamma(\mu, \nu) \right\}.$$

Let us show, first, that W_2 is a distance. A “formal” proof using transport maps is

$$W_2^2(\mu, \nu) \sim \int_X d^2(x, T(x))\, d\mu(x), \qquad W_2^2(\nu, \sigma) \sim \int_X d^2(y, S(y))\, d\nu(y)$$

with $T_\sharp \mu = \nu$, $S_\sharp \nu = \sigma$. Then $(S \circ T)_\sharp \mu = \sigma$ and

SCUOLA NORMALE SUPERIORE

$$
\begin{aligned}
W_2(\mu,\sigma) &\le \|d(S\circ T, Id)\|_{L^2(\mu)} \\
&\le \|d(S\circ T, T)\|_{L^2(\mu)} + \|d(T, Id)\|_{L^2(\mu)} \\
&= \|d(S, Id)\|_{L^2(\nu)} + \|d(T, Id)\|_{L^2(\mu)} \\
&\sim W_2(\nu,\sigma) + W_2(\mu,\nu).
\end{aligned}
$$

To make this rigorous, one can use the result $\min(K) = \inf(M)$. Alternatively, the theory of disintegration of measures provides a (non canonical) “composition map” between plans, mapping in the right way $\Gamma(\mu,\nu)\times\Gamma(\nu,\sigma)$ into $\Gamma(\mu,\sigma)$: the decompositions

$$
d\pi(x,y) = d\pi_y(x)d\nu(y), \qquad d\pi'(y,z) = d\pi'_y(z)d\nu(y)
$$

yield a plan $d(\pi'\circ\pi)(x,z) := \int d(\pi_y\times\pi'_y)(x,z)\,d\nu(y) \in \Gamma(\mu,\sigma)$.

SCUOLA NORMALE SUPERIORE

$\big(\mathscr{P}_2(X), W_2\big)$ inherits many properties from X:

- it is complete if X is complete;
- it is compact if X is compact;
- it is a length space if X is a length space;
- it is a Positively Curved (PC) space if X is PC.

Furthermore, X canonically and isometrically embeds into $\mathscr{P}_2(X)$ via the map $x \mapsto \delta_x$.

SCUOLA NORMALE SUPERIORE

Convergence in $\mathscr{P}_2(X)$

Theorem 1. (X, d) *complete. Then* $(\mu_n) \subset \mathscr{P}_2(X)$ *converges to* $\mu \in \mathscr{P}_2(X)$ *iff*

$$\mu_n \xrightarrow{C_b} \mu \quad \textit{and} \quad \lim_{n\to\infty} \int_X d^2(x, x_0)\, d\mu_n(x) = \int_X d^2(x, x_0)\, d\mu(x).$$

We prove just one implication, from W_2 convergence to weak C_b convergence. By a monotone approximation argument, suffices to consider $\phi \in \mathrm{Lip}_b(X)$; then, with $\pi_n \in \Gamma_0(\mu, \mu_n)$, we have

$$\begin{aligned} \int_X \phi\, d\mu_n &= \int_{X\times X} \phi(y)\, d\pi_n(x, y) \\ &= \int_{X\times X} \phi(x)\, d\pi_n(x, y) + O\big(W_2(\mu_n, \mu)\big) \\ &= \int_X \phi\, d\mu + O\big(W_2(\mu_n, \mu)\big) \end{aligned}$$

because $\int d(x, y)\, d\pi_n \le W_2(\mu_n, \mu)$. $\square$

SCUOLA NORMALE SUPERIORE

Geodesics in $\mathscr{P}_2(H)$

Theorem 2. *$X = H$ Hilbert. If $\pi \in \Gamma_0(\mu,\nu)$, $z_t(x,y) = (1-t)x + ty$, then*

$$(*) \qquad \mu_t := (z_t)_\sharp \pi, \qquad t \in [0,1]$$

is a constant speed geodesic from μ to ν. Conversely, any constant speed geodesic from μ to ν has this representation for some unique π.

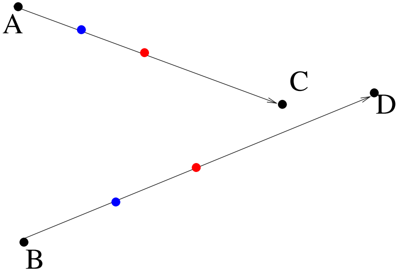

Figure: Geodesic from $(\delta_A + \delta_B)/2$ to $(\delta_C + \delta_D)/2$, $t = 1/4$, $t = 1/2$

μ_t geodesic

$$\Updownarrow$$

$$\mu_t = (z_t)_\sharp \pi,\ \pi \in \Gamma_0(\mu_0, \mu_1).$$

Remarks. (1) When π is induced by a transport T, (*) reduces to a linear interpolation between Id and T:

$$\mu_t = \big((1-t)Id + tT\big)_\sharp \mu \qquad t \in [0,1].$$

(2) $T - Id$ can be thought as the "initial velocity" of the geodesic, and we will see how the concept of velocity in $\mathscr{P}_2(H)$ can be made rigorous.

(2) Notice that the "conventional" interpolations

$$\tilde{\mu}_t := (1-t)\mu_0 + t\mu_1$$

are *unnatural* in this context, since they have an infinite length. The same is true for the OTT distances W_p, $p > 1$.

(4) Similar results hold for length spaces, even when geodesics are not unique: mass is moved with constant speed along a family of geodesics induced by π (which should be thought as a measure in the space of constant speed geodesics, rather than a measure in $X \times X$).

SCUOLA NORMALE SUPERIORE

Proof. Let us prove that μ_t defined in (*) are a constant speed geodesic, namely $W_2(\mu_t, \mu_s) = |t-s| W_2(\mu_0, \mu_1)$. By the triangle inequality it suffices to show that

$$W_2(\mu_s, \mu_t) \le (t-s) W_2(\mu_0, \mu_1) \qquad 0 \le s \le t \le 1.$$

This follows by considering $\pi_{st} := \big((1-s)x + sy, (1-t)x + ty\big)_\sharp \in \Gamma(\mu_s, \mu_t)$, so that

$$\begin{aligned} W_2^2(\mu_s, \mu_t) &\le \int |z-w|^2 \, d\pi_{st} = (t-s)^2 \int |x-y|^2 \, d\pi \\ &= (t-s)^2 W_2^2(\mu_0, \mu_1). \ \Box \end{aligned}$$

SCUOLA NORMALE SUPERIORE

Remark. The proof of the converse implication depends on the fact that geodesics are "very regular from the inside", namely $\Gamma_0(\mu_t, \mu_1)$ and $\Gamma_0(\mu_t, \mu_0)$, $t \in (0, 1)$, are singletons, induced by transport maps T_1, T_0, with Lipschitz constants less than $1/t$ and $1/(1-t)$ respectively. In addition

$$\pi = (T_0, T_1)_\sharp \mu_t,$$

where π is precisely the plan inducing the geodesic.

This ultimately depends on the monotonicity of the support of π. In the simpler case when π is induced by a transport T, $(1-t)Id + tT$ mapping μ_0 to μ_t is a monotone operator larger than $(1-t)Id$, hence its inverse is a monotone operator mapping μ_t to μ_0 with Lipschitz constant less than $1/(1-t)$.

We can now consider the subspace

$$\mathscr{P}_2^a(\mathbb{R}^n) := \{\mu \in \mathscr{P}_2(\mathbb{R}^n) : \ \mu \ll \mathscr{L}^n\}$$

and analyze its closure with respect to geodesic interpolation.

SCUOLA NORMALE SUPERIORE

Theorem 3. *Let $\mu = \rho\mathscr{L}^n \in \mathscr{P}_2^a(\mathbb{R}^n)$, $\nu \in \mathscr{P}_2(\mathbb{R}^n)$ and let T be the optimal map. Then $\mu_t \ll \mathscr{L}^n$ for all $t \in [0,1)$ and its density ρ_t is given by*

$$\rho_t = \frac{\rho}{\det \nabla T_t} \circ T_t^{-1} \qquad \text{on } \mathbb{R}^n \setminus T_t(\Sigma).$$

Here $T_t = (1-t)Id + tT$ and Σ is the set where T is not differentiable.

Proof. The formula for the density follows by the change of variables formula. The proof of absolute continuity of μ_t is easy:

$$\mathscr{L}^n(A) = 0 \Rightarrow \mathscr{L}^n\big(T_t^{-1}(A)\big) = 0 \Rightarrow \mu\big(T_t^{-1}(A)\big) = 0 \Rightarrow \mu_t \ll \mathscr{L}^n.$$

Here the first implication depends on the Lipschitz property of T_t^{-1}, and the second one by the fact that $\mu \ll \mathscr{L}^n$.

Geodesics and Hamilton-Jacobi equations

Let $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^n)$ and let us look at the optimal Kantorovich potential φ_t, ψ_t in the optimal transport problem from $\mu_0 \ll \mathscr{L}^n$ to μ_t.

Since $T_t(x) = (1-t)x + tT(x)$ and $T(x) = x - \nabla\varphi_1(x)$ we get

$$\nabla\varphi_t(x) = x - T_t(x) = t\nabla\varphi_1(x),$$

hence we may take $\varphi_t = t\varphi_1$ and $\psi_t = (t\varphi_1)^c$, namely

$$\psi_t(y) = \inf\left\{\frac{|x-y|^2}{2} - t\varphi_1(x)\right\}.$$

By the Hopf-Lax formula, we recognize that $\psi_t = tu_t$, where u_t solves

$$\text{(HJ)} \qquad \partial_t u_t + \frac{1}{2}|\nabla u_t|^2 = 0$$

and starts from $-\varphi_1$ at $t = 0$.

Geodesics and Hamilton-Jacobi equations

Since $T_t^{-1}(y) = y - \nabla\psi_t(y)$ is the optimal transport map from μ_t to μ_0, $t^{-1}\nabla\psi_t = \nabla u_t$ can also be thought as the “velocity” of the curve μ_t, and indeed $(\mu_t, \nabla u_t)$ solve the *continuity* equation

$$\frac{d}{dt} + \nabla \cdot \big((\nabla u_t)\mu_t\big) = 0.$$

This picture remains true in more general contexts, where an Eulerian description of geodesics can be achieved by a continuity equation with a gradient velocity field, coupled with an Hamilton-Jacobi equation:

$$\begin{cases} \frac{d}{dt}\mu_t + \nabla \cdot \big((\nabla u_t)\mu_t\big) = 0 \\ \partial_t u_t + \frac{1}{2}|\nabla u_t|^2 = 0. \end{cases}$$

SCUOLA NORMALE SUPERIORE

A more general class of interpolating curves

Given $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^n)$ and $\sigma \in \mathscr{P}_2^a(\mathbb{R}^n)$ we define the *interpolating curve with base* σ as follows:

$$\mu_t := \big((1-t)T_0 + tT_1\big)_\sharp \sigma,$$

where T_i, $i = 0, 1$, are the optimal maps from σ to μ_i.

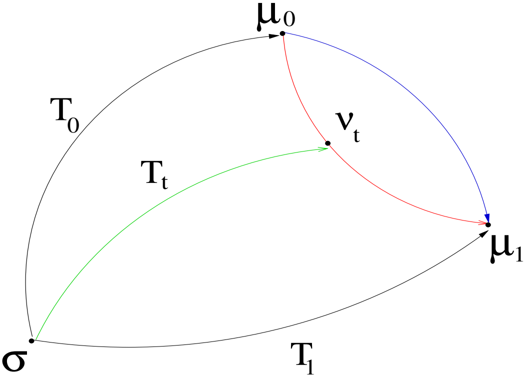

Figure: Interpolating curve with base σ

A more general class of interpolating curves

In general, for $\pi_0 \in \Gamma_0(\sigma, \mu_0)$, $\pi_1 \in \Gamma_0(\sigma, \mu_1)$, we find

$$\eta \in \mathscr{P}(H \times H \times H) \quad \text{with} \quad \pi_0 = (\pi_1, \pi_2)_\sharp \eta,\ \pi_1 = (\pi_1, \pi_3)_\sharp \eta$$

and set

$$\mu_t := \big((1-t)x_2 + tx_3\big)_\sharp \sigma.$$

An explicit, not canonical, formula for η is

$$d\eta(x_1, x_2, x_3) = d\big((\pi_0)_{x_1} \times (\pi_1)_{x_1}\big)(x_2, x_3) d\sigma(x_1),$$

where

$$d\pi_0(x_1, x_2) = d(\pi_0)_{x_1}(x_2) d\sigma(x_1), \quad d\pi_1(x_1, x_3) = d(\pi_1)_{x_1}(x_3) d\sigma(x_1).$$

SCUOLA NORMALE SUPERIORE

Convex functionals in $\mathscr{P}_2(H)$

We start from the simplest example, the *potential energy*:

$$\mathcal{V}(\mu) := \int V \, d\mu.$$

We assume V bounded from below, although for some applications this is too restrictive.

Lemma 4. *V is λ-convex iff $\mathcal{V}$ is λ-convex along geodesics. If this happens, then $\mathcal{V}$ is λ-convex along all interpolating curves.*

Proof. One implication is trivial (just consider Dirac masses). To prove the other one, notice that $\mu_t = \big((1-t)I + tT\big)_\sharp \mu_0$ yields

$$\begin{aligned} \mathcal{V}(\mu_t) &= \int V\big((1-t)I + tT\big) \, d\mu_0 \leq (1-t)\mathcal{V}(\mu_0) + t\mathcal{V}(\mu_1) \\ &- \frac{\lambda}{2} t(1-t) \int |T - Id|^2 \, d\mu_0. \end{aligned}$$

SCUOLA NORMALE SUPERIORE

The proof of the final statement about convexity along all interpolating curves is analogous:

$$\begin{aligned}\mathcal{V}(\mu_t) &= \int V\big((1-t)T_0 + tT_1\big)\, d\sigma \le (1-t)\mathcal{V}(\mu_0) + t\mathcal{V}(\mu_1) \\ &- \frac{\lambda}{2} t(1-t) \int |T_0 - T_1|^2\, d\sigma.\end{aligned}$$

Here we take into account that $\lambda \ge 0$ and that

$$\int |T_0 - T_1|^2\, d\sigma \ge W_2^2(\mu_0, \mu_1),$$

because $(T_0, T_1)_\sharp \sigma \in \Gamma(\mu_0, \mu_1)$.

SCUOLA NORMALE SUPERIORE

Interaction energy

Now we consider the *interaction energy*

$$\mathcal{W}(\mu) := \int W(x_1, \ldots, x_k)\, d\mu \times \cdots \times d\mu.$$

Lemma 5. *Let* $W : H^k \to \mathbb{R} \cup \{+\infty\}$ *be* λ *convex, with* $\lambda \geq 0$*. Then* $\mathcal{W}$ *is* $k\lambda$*-convex.*

The proof is entirely similar to the proof of Lemma 4.

SCUOLA NORMALE SUPERIORE

Internal energy

Now we come to the most important example, the *internal energy*:

$$\Phi(\rho) := \int_{\mathbb{R}^n} U(\rho)\, dx \qquad \mu = \rho\mathscr{L}^n \in \mathscr{P}_2(\mathbb{R}^n)$$

with $U(0) = 0$, satisfying Mc Cann's displacement convexity condition

$$(\mathrm{Mc}_n) \qquad s \mapsto s^n U(s^{-n}) \quad \text{convex and nonincreasing.}$$

Main examples. $U(z) = z^\alpha$, $\alpha \geq 1$ $\quad (s^n U(s^{-n}) = s^{n(1-\alpha)})$;
$U(z) = z^\alpha/(\alpha - 1)$, $\alpha \geq 1 - 1/n$ $\quad (s^n U(s^{-n}) = t^{n(1-\alpha)}/(1-\alpha)$;
$U(z) = z \ln z$ $\quad (s^n U(s^{-n}) = -n \ln s)$.

SCUOLA NORMALE SUPERIORE

Convexity of internal energy

Theorem 6. *If* (Mc_n) *holds, then* Φ *is convex along all interpolating curves.*

Proof. Let T_0, T_1 be optimal transport maps from σ to μ_0, μ_1 respectively, $T_t := (1-t)T_0 + tT_1$. The proof relies on the following two facts:

• T_t fulfils the regularity assumptions in the change of variables formula (differentiability μ_t-a.e., $\det \nabla T_t \neq 0$ μ_t-a.e.);

• ∇T_t is symmetric, nonnegative and $A \mapsto \left[\det A\right]^{1/n}$ is concave in the space $\mathrm{Sym}_+^{n\times n}$ of symmetric and nonnegative operators.

SCUOLA NORMALE SUPERIORE

Convexity of internal energy

These two facts imply that $\mu_t = \rho_t \mathscr{L}^n$ with

$$\rho_t = \frac{\rho}{\det \nabla T_t} \circ T_t^{-1},$$

hence

$$\Phi(\rho_t) \overset{(y=T_t(x))}{=} \int \det \nabla T_t(x) U\Big(\frac{\rho(x)}{\det \nabla T_t(x)}\Big)\, dx.$$

We conclude noticing that $A \mapsto \det A\, U\big(\rho(x)/\det A\big)$ is convex, being the composition of the concave map $A \mapsto [\det A]^{1/n}$ and the convex nonincreasing map $z \mapsto z^n U(\rho(x)/z^n)$. $\square$

SCUOLA NORMALE SUPERIORE

Concavity of $A \mapsto [\det A]^{1/n}$

By homogeneity, suffices to show

$$[\det(A+B)]^{1/n} \geq [\det(A)]^{1/n} + [\det(B)]^{1/n}.$$

Up to a rotation we can assume $B_{ij} = \lambda_i \delta_{ij}$ diagonal.

By approximation we can assume $\lambda_i > 0$ for all i. Dividing the ij-th entries of A and B by $\sqrt{\lambda_i \lambda_j}$ reduces to the case when $B = I$.

A further rotation allows a reduction to the case when $A_{ij} = \eta_i \delta_{ij}$ is diagonal as well. The geometric mean/arithmetic mean inequality then gives

$$\left(\prod_{i=1}^{n} \frac{\eta_i}{1+\eta_i}\right)^{1/n} + \left(\prod_{i=1}^{n} \frac{1}{1+\eta_i}\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^{n} \frac{\eta_i}{1+\eta_i} + \frac{1}{1+\eta_i} = 1.$$

SCUOLA NORMALE SUPERIORE

$\det \nabla T_t > 0$ σ-a.e.

Since $\mu_0 \ll \mathscr{L}^n$ we have that $\det \nabla T_0 > 0$ σ-a.e.; concavity then gives

$$[\det \nabla T_t]^{1/n} \geq (1-t)[\det \nabla T_0]^{1/n} + t[\det \nabla T_1]^{1/n} > 0 \quad \sigma\text{-a.e.}$$

for all $t \in (0, 1)$.

SCUOLA NORMALE SUPERIORE

Application: the Brunn-Minkowski inequality

A direct proof via optimal transportation of the Brunn-Minkowski inequality, in the scaled version

$$\mathrm{Vol}^{1/n}\left(\frac{A+B}{2}\right) \geq \frac{1}{2}\mathrm{Vol}^{1/n}(A) + \frac{1}{2}\mathrm{Vol}^{1/n}(B),$$

can be achieved as follows. First, we know that the energy $\mathcal{E}(\rho) := \int \rho^{1-\frac{1}{n}}\, dx$ is *concave* along Wasserstein geodesics. Then, set

$$\rho_A(x) := \begin{cases} \frac{1}{\mathrm{Vol}(A)} & \text{if } x \in A \\ 0 & \text{if } x \notin A, \end{cases} \qquad \rho_B(x) := \begin{cases} \frac{1}{\mathrm{Vol}(B)} & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$$

and denote by $\{\rho_t\}_{t\in[0,1]}$ the constant speed geodesic between ρ_A and ρ_B. Then, the conclusion follows by

$$\mathcal{E}(\rho_0) = \mathrm{Vol}^{1/n}(A), \quad \mathcal{E}(\rho_1) = \mathrm{Vol}^{1/n}(B), \quad \mathcal{E}(\rho_{1/2}) \leq \mathrm{Vol}^{1/n}\left(\frac{A+B}{2}\right).$$

The latter inequality is implied by Jensen's inequality and the fact that $\rho_{1/2}$ is concentrated on $(A+B)/2$.

SCUOLA NORMALE SUPERIORE

Relative Entropy

The previous examples of convex functionals can be combined and generalized in several ways. Particularly interesting is:

$$\mathcal{F}(\rho) := \int \rho \ln \rho \, dx + \int V \rho \, dx.$$

Then $\mathcal{F}$ is λ-convex along all interpolating curves iff V is λ-convex.

In view of some extensions to infinite-dimensional spaces it is convenient to read $\mathcal{F}$ and its properties more intrinsically, in terms of the new reference measure $\gamma = e^{-V} \mathscr{L}^n$. Setting

$$\mu = \rho \mathscr{L}^n = u\gamma \qquad \text{(so that } u = \rho e^V\text{)}$$

we obtain that $\mathcal{F}$ is the *Relative Entropy* of μ with respect to γ:

$$\mathcal{F}(\rho) = \int u \ln u \, d\gamma =: \mathcal{H}(\mu|\gamma).$$

We adopt the convention $\mathcal{H}(\mu|\gamma) = +\infty$ if $\mu \ll \gamma$ does not hold.

Log-concavity and convexity along geodesics

Log-concavity. $\gamma \in \mathscr{P}(H)$ is log-concave if

$$\ln \gamma\big((1-t)A + tB\big) \geq (1-t)\ln \gamma(A) + t \ln \gamma(B) \qquad \forall t \in (0,1).$$

(Borell) $\gamma \in \mathscr{P}(\mathbb{R}^n)$, non degenerate, is log-concave iff $\gamma = e^{-V}\mathscr{L}^n$, with $V : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convex.
Then, in [AGS] it is proved that:

$$\gamma \text{ log-concave} \qquad \Longleftrightarrow \qquad \mathcal{H}(\cdot|\gamma) \text{ convex in } \mathscr{P}_2(H).$$

This opens the possibility to study the gradient flow of $\mathcal{H}(\cdot|\gamma)$ by optimal transportation methods.

SCUOLA NORMALE SUPERIORE

The differentiable side of optimal transportation

We start with some heuristics, suggested by Otto, and then we proceed with some rigorous results.
We start from the continuity equation

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0$$

suggesting $\delta\mu = -\nabla \cdot (v\mu)$, so that

$$T_\mu \mathscr{P}_2(\mathbb{R}^n) \sim \left\{ -\nabla \cdot (v\mu) : \ v \in L^2(\mu; \mathbb{R}^n) \right\}$$

with the metric

$$\langle -\nabla \cdot (v\mu), -\nabla \cdot (w\mu) \rangle := \int_{\mathbb{R}^n} \langle v, w \rangle d\mu.$$

Benamou-Brenier formula

According to this interpretation we might consider the induced Riemannian distance

$$\tilde{W}(\mu,\nu) \quad := \inf\Big\{ \int_0^1 \int |v_t|^2\, d\mu_t dt : \ \mu_0 = \mu,\ \mu_1 = \nu, \\ \frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \Big\}.$$

Theorem 7. (Benamou–Brenier) *$W_2(\mu,\nu) = \tilde{W}_2(\mu,\nu)$.*

We start with the proof of the inequality $\tilde{W}_2 \le W_2$. To this aim, we just estimate the action of the geodesic path in $\mathscr{P}_2(\mathbb{R}^n)$.

SCUOLA NORMALE SUPERIORE

Lemma 8. *$f \in L^2(\sigma, \mathbb{R}^n)$ implies $T_\sharp(f\sigma) \ll T_\sharp\sigma$ (componentwise) and its density h satisfies*

$$(*) \qquad \|h\|_{L^2(T_\sharp\sigma)} \le \|f\|_{L^2(\sigma)}.$$

Indeed,

$$\begin{aligned} \langle T_\sharp(f\sigma), g\rangle &= \langle g \circ T, f\sigma\rangle \le \|f\|_{L^2(\sigma)}\|g \circ T\|_{L^2(\sigma)} \\ &= \|f\|_{L^2(\sigma)}\|g\|_{L^2(T_\sharp\sigma)}. \ \square \end{aligned}$$

By Riesz theorem the inequality (*) follows. Returning to the proof of the inequality, a velocity field v_t compatible with μ_t is (implicitly) defined by

$$v_t\mu_t = (T_t)_\sharp\big((T - Id)\mu\big).$$

$\tilde{W}_2 \leq W_2$

Indeed,

$$\frac{d}{dt}\int \phi \, d\mu_t = \frac{d}{dt}\int \phi \circ T_t \, d\mu = \int \langle \nabla\phi(T_t), T - Id \rangle \, d\mu$$
$$= \langle \nabla\phi, (T_t)_\sharp((T - Id)\mu) \rangle = \int \langle \nabla\phi, v_t \rangle \, d\mu_t.$$

Lemma 8 gives

$$\tilde{W}_2(\mu,\nu) \leq \int_0^1 \|v_t\|^2_{L^2(\mu_t)} dt \leq \int_0^1 \|T - Id\|^2_{L^2(\mu)} \, dt = W_2^2(\mu,\nu). \ \square$$

SCUOLA NORMALE SUPERIORE

$\tilde{W}_2 \geq W_2$

We now give a sketch of proof of the opposite inequality along the lines of [AGS]. First we provide a lower bound on the action under suitable regularity assumptions on v_t, then we use a smoothing argument.
Assume that v_t is sufficiently regular to have:
1) a flow map $X(t,x)$, i.e. a solution to the ODE $\partial_t X(t,x) = v_t\big(X(t,x)\big)$ with the Cauchy condition $X(0,x) = x$;
2) the unique solution μ_t to the continuity equation with velocity v_t, starting from μ_0, is $X(t,\cdot)_\sharp \mu_0$.
By 1) we have $|X(1,x) - x|^2 \leq \int_0^1 |v_t|^2\big(X(t,x)\big)\,dt$ and an integration with respect to μ_0 gives (using $X(1,\cdot)$ as a transport from μ_0 to μ_1)

$$W_2^2(\mu_0,\mu_1) \leq \int_0^1 \int |v_t|^2\big(X(t,x)\big)d\mu_0(x)dt = \int_0^1 \int |v_t|^2 d\mu_t dt.$$

SCUOLA NORMALE SUPERIORE

In general we mollify both sides of the continuity equation with the heat kernel ρ_ϵ to get

$$\frac{d}{dt}\mu_t * \rho_\epsilon + \nabla \cdot [(v_t \mu_t) * \rho_\epsilon] = 0.$$

Defining $f_t^\epsilon := \mu_t * \rho_\epsilon$ and

$$v_t^\epsilon := \frac{(v_t \mu_t) * \rho_\epsilon}{\mu_t * \rho_\epsilon}$$

we obtain that:
– $\frac{d}{dt} f_t^\epsilon + \nabla \cdot (v_t^\epsilon f_t^\epsilon) = 0$;
– the convexity of the map $(J, t) \to t|J/t|^2 = |J|^2/t$ ($J = vf$), together with Jensen's inequality, give that the action of the mollified curve does not increase:

$$(*) \qquad \int |v_t^\epsilon|^2 f_t^\epsilon dx \le \int |v_t|^2 \, d\mu_t.$$

SCUOLA NORMALE SUPERIORE

The inequality (*), together with the local (in space) Lipschitz condition

$$\|v_t^\epsilon\|_{W^{1,\infty}(B_R)} \in L^\infty(0,1) \qquad \forall \epsilon > 0,\ R > 0$$

can be used to show that the maximal solution to the ODE $\partial_t X(t,x) = v_t^\epsilon\big(X(t,x)\big)$ is defined up to $t = 1$ for $f_0^\epsilon \mathscr{L}^n$-a.e. x. Hence, the previous step gives

$$W_2^2(f_0^\epsilon \mathscr{L}^n, f_1^\epsilon \mathscr{L}^n) \le \int_0^1 \int |v_t^\epsilon|^2 f_t^\epsilon dx \le \int_0^1 \int |v_t|^2 \, d\mu_t.$$

Letting $\epsilon \downarrow 0$ the inequality is achieved. $\square$

SCUOLA NORMALE SUPERIORE

Extended Monge-Kantorovich distances

The BB formula opens the possibility to define extended Monge-Kantorovich distances simply changing the action functional.

Dolbeault-Nazaret-Savaré recently investigated the distances

$$W_h(\rho_0\mathscr{L}^n, \rho_1\mathscr{L}^n) := \inf\left\{\int_0^1 \int h(\rho_t)|v_t|^2\,dxdt\right\}$$

where the infimum is made among all solutions ρ_t to the *nonlinear continuity equation*

$$\frac{d}{dt}\rho_t + \nabla\cdot\big(h(\rho_t)v_t\big) = 0.$$

Passing to the new velocity $\bar{v} := vh(\rho)/\rho$ we recover a linear continuity equation with velocity $\bar{v}_t$, but the action is

$$\int_0^1 \int \rho_t f(\rho_t)|\bar{v}_t|^2\,dxdt$$

with $f(\rho) = \rho/h(\rho)$, so it depends on ρ in a nonlinear way.

SCUOLA NORMALE SUPERIORE

Existence of curves with minimal action can be proved if h is concave, since this results in the joint convexity of ($J = v\rho$)

$$(\rho, J) \mapsto h(\rho)\left|\frac{J}{h(\rho)}\right|^2.$$

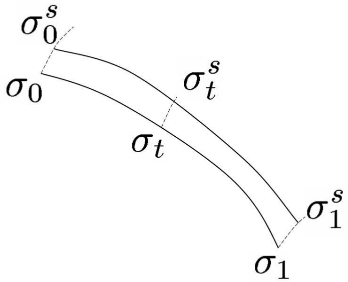

Figure: The action of the semigroup does not increase **E**

Monotonicity of **E**

$\Updownarrow$

Contractivity

(EVI) $\Longrightarrow$ Convexity

Convergence of the wave front tracking method and L^1-like contractive distances: Bressan, Bressan-Crasta-Piccoli.

Absolutely continuous curves and metric derivative

Now, we are going to extend the Benamou-Brenier analysis to all curves $t \mapsto \mu_t$ of finite length, not necessarily geodesics. Up to a reparameterization, I shall assume the curve to be absolutely continuous.

Definition. (E, d) metric space. We say that $x : [0, 1] \to E$ is *absolutely continuous* if

$$d\big(x(s), x(t)\big) \le \int_s^t g(\tau) d\tau \qquad \forall 0 \le s \le t \le 1$$

for some $g \in L^1(0, 1)$.

It turns out that, for x absolutely continuous, the minimal g (up to $\mathscr{L}^1$-negligible sets) with this property exists and is the *metric derivative*:

$$|x'(t)| := \lim_{h \to 0} \frac{d\big(x(t+h), x(t)\big)}{|h|} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, 1).$$

Theorem 9. [AGS] *Let $\mu_t : [0,1] \to \mathscr{P}_2(H)$ be absolutely continuous. Then there exists a unique $v_t \in L^2(\mu_t; H)$ such that*

(i) *the continuity equation $\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0$ holds;*

(ii) $\|v_t\|_{L^2(\mu_t)} \le |\mu_t'|$ *for $\mathscr{L}^1$-a.e. $t \in (0,1)$.*

Conversely, if (v_t, μ_t) fulfil (1) and $\|v_t\|_{L^2(\mu_t)} \in L^1(0,1)$, then μ_t is absolutely continuous as a $\mathscr{P}_2(H)$-valued map and

$$\text{(ii)}' \qquad \|v_t\|_{L^2(\mu_t)} \ge |\mu_t'| \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,1).$$

We shall call the "optimal" velocity field v_t given by Theorem 9 *tangent* field to μ_t; its $L^2(\mu_t)$ norm gives the rate of change of W_2 along the curve.

SCUOLA NORMALE SUPERIORE

The constructive part of Theorem 9 is based on a duality argument, while the converse part uses a smoothing scheme very much similar to the one used in the proof of the Benamou–Brenier formula.

The duality argument provides also the information that v_t is in the $L^2(\mu_t)$ closure of gradient vector fields (this is not surprising, since optimal transport maps are gradients); this leads to the more precise definition

$$\mathrm{Tan}_\mu\big(\mathscr{P}_2(\mathbb{R}^d)\big) := \overline{\{\nabla\phi:\ \phi \in C_c^\infty(\mathbb{R}^n)\}}^{L^2(\mu)}.$$

Eventually the tangent velocity field v_t to μ_t can be also characterized in these terms:

1) validity of the continuity equation;

2) $v_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ for $\mathscr{L}^1$-a.e. $t \in (0,1)$.

Wasserstein gradient

We have an energy $\mathcal{E} : \mathscr{P}_2(H) \to \mathbb{R} \cup \{+\infty\}$ and we want to compute its "Wasserstein" gradient $\nabla^W \mathcal{E}(\mu)$.

Rule: replace the classical additive variations $\mu \mapsto \mu + \varepsilon\nu$ by "transport" variations

$$\mu \mapsto \mu_\epsilon := (\mathrm{Id} + \epsilon v)_\sharp \mu, \qquad v \in \mathrm{Tan}_\mu \mathscr{P}_2(H).$$

This point of view leads to standard transpositions of the concept of differential, subdifferential, etc. to the Wasserstein space. For example:

$$\partial^W \mathcal{E}(\mu) := \Big\{ \xi \in \mathrm{Tan}_\mu\big(\mathscr{P}_2(H)\big) : \mathcal{E}(\nu) \geq \mathcal{E}(\mu) + \int_H \langle \xi, T_\mu^\nu - Id \rangle \, d\mu \ \ \forall \nu \in \mathscr{P}_2(H) \Big\}.$$

SCUOLA NORMALE SUPERIORE

Wasserstein gradient

Notice that μ_ϵ coincide up to first order with the solution $\tilde{\mu}_\epsilon$ to

$$(*) \qquad \frac{d}{dt}\tilde{\mu}_\epsilon + \nabla \cdot (v\tilde{\mu}_\epsilon) = 0$$

because $\frac{d}{d\epsilon}\mu_\epsilon\big|_{\epsilon=0} + \nabla \cdot (v\mu) = 0$. Indeed,

$$\frac{d}{d\epsilon}\int \phi \, d\mu_\epsilon\Big|_{\epsilon=0} = \frac{d}{d\epsilon}\int \phi(Id + \epsilon v)\, d\mu\Big|_{\epsilon=0} = -\langle \nabla \cdot (v\mu), \phi \rangle.$$

Hence, we may use also (*) in computing (formally) the Wasserstein gradient. As an example, we compute the Wasserstein gradient of the internal energy functional.

SCUOLA NORMALE SUPERIORE

$$\nabla^W \int U(\rho)\, dx = \nabla U'(\rho).$$

Indeed, if $\tilde{\mu}_\epsilon = \tilde{\rho}_\epsilon \mathscr{L}^n$, we have

$$\frac{d}{d\epsilon} \int U(\tilde{\rho}_\epsilon)\, dx \bigg|_{\epsilon=0} = -\int U'(\rho) \nabla \cdot (v\rho)\, dx = \int \langle \nabla U'(\rho), v \rangle \rho\, dx.$$

Remark. This derivation works only under some regularity assumptions on ρ, too restrictive for some applications. Working with transport variations, instead, one finds

$$\nabla^W \int U(\rho)\, dx = \frac{\nabla L_U(\rho)}{\rho}.$$

where $L_U(z) = zU'(z) - U(z)$. Since $L_U'(z) = zU''(z)$, the two are equivalent at a smooth level.

Remark. Analogously,

$$\nabla^W \int V\, d\mu = \nabla V, \qquad \nabla^W \int W\, d\mu \otimes \mu = (\nabla W) * \mu.$$

Transport and log-Sobolev inequalities

The energy inequality

$$\Phi(x) \geq \Phi(x_{\min}) + \frac{\lambda}{2} d^2(x, \bar{x})$$

holds for any λ-convex function Φ in a length metric space (E, d): it suffices to consider the map $t \mapsto \Phi(\gamma(t))$, where $\gamma : [0, 1] \to E$ is a constant speed geodesic from $\bar{x}$ to x.

By applying this to the Relative Entropy functional $\mathcal{H}(\cdot|\gamma)$, with $x_{\min} = \gamma = e^{-V}\mathscr{L}^n$ and V λ-convex, we get the *transport inequality*

$$W_2^2(u\gamma, \gamma) \leq \frac{2}{\lambda} \int u \ln u \, d\gamma,$$

because $\mathcal{H}(\cdot|\gamma)$ is λ-convex.

In the Gaussian case $V(x) = |x|^2/2$ this inequality has been discovered by Talagrand, and proved by a tensorization argument.

SCUOLA NORMALE SUPERIORE

Transport and log-Sobolev inequalities

Analogously, the *energy-energy dissipation inequality*

$$\Phi(x) - \Phi(x_{\min}) \leq \frac{1}{2\lambda}|\partial\Phi|^2(x)$$

holds for any λ-convex function Φ in a length metric space (E, d). Again, we can apply this inequality to $\int \rho \ln \rho \, dx + \int \rho V \, dx$ to obtain

$$\int \rho \ln \rho + \rho V \, dx \leq \frac{1}{2\lambda} \int |\nabla(\ln \rho + V)|^2 \rho \, dx.$$

With the change of variables $\rho = h^2 e^{-V}$ (i.e. $u = h^2$, $\nabla \ln \rho = 2\nabla \ln h + \nabla V$) we get the logarithmic Sobolev inequality

$$\int h^2 \ln h^2 \, d\gamma \leq \frac{2}{\lambda} \int |\nabla h|^2 \, d\gamma \quad \text{with} \quad \int h^2 \, d\gamma = 1,$$

first discovered by Gross in the Gaussian case.