OPTIMAL TRANSPORT IN GEOMETRY

Kyoto, 28 July - 1 August 2008

Cédric Villani

ENS Lyon,
Institut Universitaire de France
& JSPS

MAIN THEME

Some "hard" problems in geometry can be attacked by "soft" tools

Optimal transport is one such tool

References

- Topics in Optimal Transportation [**TOT**] (AMS, 2003): Introduction
- Optimal transport, old and new [oldnew] (Springer, 2008): Reference text, more probabilistic & geometric

Plan of the course

(5 chapters)

- Basic theory
- The Wasserstein space
- Isoperimetric/Sobolev inequalities
- Concentration of measure
- Stability of a 4th order curvature condition

Most of the time statements, sometimes elements of proof

Complement

J.P. Bourguignon's Takagi lectures (4–5 oct. 2008) about optimal transport and Ricci curvature

I. BASIC THEORY OF OPTIMAL TRANSPORT

- The modern core of the Monge–Kantorovich theory, built from the eighties to now
- Simplified statements
- Reference: [oldnew, Chap. 4, 5, 10]

The Kantorovich problem (Kantorovich, 1942)

- \mathcal{X} , \mathcal{Y} two Polish (= metric separable complete) spaces
- $\mu \in P(\mathcal{X}), \nu \in P(\mathcal{Y})$
- $c \in C(\mathcal{X} \times \mathcal{Y}; \mathbb{R}), \qquad c \geq c \in L^1(\mu) + L^1(\nu)$

$$\Pi(\mu, \nu) = \left\{ \pi \in P(\mathcal{X} \times \mathcal{Y}); \text{ marginals of } \pi \text{ are } \mu \text{ and } \nu \right\}$$

$$(\forall h, \eta) = \left\{ (h, h) \in \mathcal{A}, (h, h) \in \mathcal{A},$$

$$\int h(x) \, \pi(dx \, dy) = \int h \, d\mu; \qquad \int h(y) \, \pi(dx \, dy) = \int h \, d\nu$$

(K)
$$\left| \inf_{\pi \in \Pi(\mu, \nu)} \int c(x, y) \, \pi(dx \, dy) \right|$$

Prop: Infimum achieved by compactness of $\Pi(\mu, \nu)$

In the sequel, assume infimum is finite

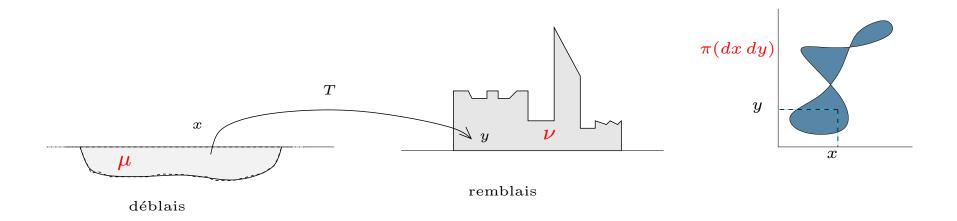
Probabilistic version

X and Y two given random variables (= with given laws)

(K')
$$\inf \mathbb{E} c(X,Y)$$

(Infimum over all couplings of (X, Y))

Engineer's interpretation



Given the initial and final distributions, transport matter at lowest possible cost

Assume
$$\pi = (\mathrm{Id}, T)_{\#} \mu = \mu(dx) \, \delta_{y=T(x)}$$

 \longrightarrow belongs to $\Pi(\mu,\nu)$ iff $T_{\#}\mu = \nu$

i.e.
$$\int (h \circ T) d\mu = \int h d\nu$$
 i.e. $\nu[A] = \mu[T^{-1}(A)]$

⇒ the Kantorovich problem becomes

(M)
$$\left| \inf_{T_{\#}\mu=\nu} \int c(x, T(x)) \, \mu(dx) \right| = \inf \mathbb{E} c(X, T(X))$$

- Interpretation: Don't split mass! Y = T(X)
- No compactness \Longrightarrow not clear if infimum achieved

History of the Monge problem

- Original Monge cost function: c(x,y) = |x-y| in \mathbb{R}^3
- For this cost, existence of a minimizer proven around 1998–2003!! (Ambrosio, Caffarelli, Evans, Feldman, Gangbo, McCann, Sudakov, Trudinger, Wang)
- Easier solution when the cost is "strictly convex" (e.g. $d(x,y)^p$, p > 1) See later.

Kantorovich duality (Kantorovich 1942; still active research area)

(Kdual)

$$\inf_{\pi \in \Pi(\mu,\nu)} \int c(x,y) \, \pi(dx \, dy) = \sup_{(\psi,\phi) \in \Psi_c} \left\{ \int \phi \, d\nu - \int \psi \, d\mu \right\}$$

- $\pi \in \Pi(\mu, \nu)$ if π has marginals μ and ν
- $(\psi, \phi) \in \Psi_c \text{ if } \phi(y) \psi(x) \le c(x, y) \quad (\forall x, y)$
- Economical interpretation: shipper's problem (buys at price $\psi(x)$ at x, sells at price $\phi(y)$ at y)
- Supremum achieved e.g. if $c \leq \overline{c} \in L^1(\mu) + L^1(\nu)$

c-convexity (I)

(Rüschendorf, nineties)

- Fix $c: \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$
- $\psi: \mathcal{X} \to \mathbb{R} \cup \{+\infty\}, \qquad \psi^c(y) := \inf_{x \in \mathcal{X}} [\psi(x) + c(x, y)]$
- $\phi: \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}, \qquad \phi^c(x) := \sup_{y \in \mathcal{Y}} [\phi(y) c(x, y)]$
- ψ is said c-convex if $(\psi^c)^c = \psi$ ϕ is said c-concave if $(\phi^c)^c = \phi$
- Ex: $c(x,y) = -x \cdot y$ in $\mathbb{R}^n \times \mathbb{R}^n$: $\psi^c = -\psi^*$ (Legendre transform); c-convex \iff l.s.c. convex

Rks: (a) many conventions!!

(b) differential criterion for c-convexity?? Yes if the Ma–Trudinger–Wang condition is satisfied, see later.

c-convexity (II)

• If ψ is c-convex, define its c-subdifferential $\partial_c \psi$ by

$$\partial_c \psi(x) = \left\{ y \in \mathcal{Y}; \ \forall z \in \mathcal{X}, \quad \psi(z) + c(z, y) \ge \psi(x) + c(x, y) \right\}$$

• $\Gamma \subset \mathcal{X} \times \mathcal{Y}$ is c-cyclically monotone (c-CM) if

$$\forall N \in \mathbb{N} \quad \forall (x_1, y_1), \dots, (x_N, y_N) \in \Gamma^N,$$

$$\sum_{i=1}^N c(x_i, y_i) \le \sum_{i=1}^N c(x_i, y_{i+1}) \qquad (y_{N+1} = y_1)$$

- $\partial_c \psi$ is c-CM (immediate)
- Ex: $c(x,y) = -x \cdot y \Longrightarrow \partial_c \psi = \partial \psi$

c-convex analysis

(Rockafellar; Rüschendorf)

- $\psi^c(y) \psi(x) \le c(x, y)$ $\forall \psi$
- ψ is c-convex $((\psi^c)^c = \psi)$ iff $\exists \zeta$; $\psi = \zeta^c$
- $\partial_c \psi = \{(x,y); \ \psi^c(y) \psi(x) = c(x,y)\} \text{ is } c\text{-CM}$
- If Γ is c-CM then $\exists \psi$ c-convex s.t. $\Gamma \subset \partial_c \psi$

Pf: Fix
$$(x_0, y_0) \in \Gamma$$
, define $\psi(x) := \sup_{m \in \mathbb{N}} \sup \left\{ \left[c(x_0, y_0) - c(x_1, y_0) \right] + \left[c(x_1, y_1) - c(x_2, y_1) \right] + \cdots + \left[c(x_m, y_m) - c(x, y_m) \right]; \quad (x_i, y_i) \in \Gamma \right\}$

• c-convex functions inherit some regularity from c, e.g.

$$\|\psi\|_{\text{Lip}} \le \sup_{y} \|c(\cdot, y)\|_{\text{Lip}}, \qquad D^2\psi \ge \inf_{x,y} (-D_x^2 c)$$

Saddle point structure $\pi \in \Pi(\mu, \nu), \psi$ c-convex

$$\boxed{ \frac{\operatorname{Spt} \pi \subset \partial_c \psi}{\operatorname{Spt} \pi \subset \partial_c \psi} } \Longrightarrow \left\{ \begin{matrix} \pi \text{ optimal} \\ \psi \text{ optimal} \end{matrix} \right\} \Longrightarrow (\mathbf{Kdual})$$

$$\begin{aligned} \mathbf{Pf:} \quad \forall \widetilde{\pi} \ \forall \widetilde{\psi} \quad & \int c(x,y) \, \widetilde{\pi}(dx \, dy) \geq \int \left[\psi^c(y) - \psi(x) \right] \widetilde{\pi}(dx \, dy) \\ & = \int \psi^c \, d\nu - \int \psi \, d\mu \\ & = \int \left[\psi^c(y) - \psi(x) \right] \pi(dx \, dy) \\ & = \int c(x,y) \, \pi(dx \, dy) \\ & \geq \int \left[\widetilde{\psi}^c(y) - \widetilde{\psi}(x) \right] \pi(dx \, dy) \\ & = \int \widetilde{\psi}^c \, d\nu - \int \widetilde{\psi} \, d\mu \end{aligned} \quad \Box$$

Complements (I)

- Criteria for optimality:
- If π is optimal then (ψ, ψ^c) is optimal iff $\operatorname{Spt} \pi \subset \partial_c \psi$
- If (ψ, ψ^c) is optimal then π is optimal iff $\operatorname{Spt} \pi \subset \partial_c \psi$
- π is optimal iff Spt π is c-CM (Pratelli, Schachermayer—Teichmann 2007-2008)
- This implies stability: If $\pi_k \in \Pi(\mu_k, \nu_k)$ optimal, $\pi_k \longrightarrow \pi \in \Pi(\mu, \nu)$ (weakly), then π is optimal

Complements (II)

• Link with Aubry–Mather theory:

$$c(\mu) := \inf \left\{ C(\mu, \mu); \ \mu \in P(\mathcal{X}) \right\} \implies \exists \mu \text{ minimizer}$$

$$\mathcal{A} := \bigcap_{\psi \text{ opt.}} \partial_c \psi \qquad \qquad \mathcal{M} := \overline{\bigcup_{\pi \text{ opt.}} \operatorname{Spt} \pi}$$

These sets play an important role in dynamical systems theory [Fathi] [oldnew Chap. 8]

• Link with combinatorics: When $c: \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, (**Kdual**) reduces to a "continuous" (measure-theoretic) version of the marriage lemma [TOT Sect. 1.4]

Solution of the Monge problem under a twist condition

(Brenier, Rüschendorf, McCann, Gangbo,)

Let μ, ν, c s.t. the dual Kantorovich problem has a solution ψ

Assume

- (0) \mathcal{X} is a Riemannian manifold
- (1) c(x,y) is (uniformly) Lipschitz in x, uniformly in y
- (2) $[\nabla_x c(x,y) = \nabla_x c(x,y')] \Longrightarrow y = y'$

(Twist: $\nabla_x c$ is a 1-to-1 function of y)

- (3) c is superdifferentiable everywhere
- (4) $\mu(dx) = f(x) \operatorname{vol}(dx)$

<u>Then</u> ∃! solution to the Monge–Kantorovich problem

Structure of the solution

 $\pi(dx\,dy)$ -a.s.

$$\nabla \psi(x) + \nabla_x c(x, y) = 0$$

This determines the transport map:

$$y = T(x) = (\nabla_x c)^{-1} (x, -\nabla \psi(x))$$

Ex: (McCann 1999) $c(x,y) = d(x,y)^2/2$ on a compact Riemannian manifold, then

$$T(x) = \exp_x(\nabla \psi(x)), \text{ where } \psi \text{ is } d^2/2\text{-convex}$$

Proof of solution of Monge problem

- ψ c-convex $\stackrel{(1)}{\Longrightarrow} \psi$ Lipschitz $\stackrel{(0)}{\Longrightarrow} \psi$ differentiable a.e. $\stackrel{(4)}{\Longrightarrow} \psi$ differentiable μ -a.s.
- Spt $\pi \subset \partial_c \psi \implies \pi(dx \, dy) \text{a.s.},$ $\psi(z) + c(z, y)$ is minimum at z = x

$$\implies -\nabla \psi(x) \in \nabla_x^- c(\,\cdot\,,y)$$

$$\stackrel{(3)}{\Longrightarrow} \nabla_x c(\,\cdot\,,y) = -\nabla \psi(x)$$

$$\stackrel{(2)}{\Longrightarrow} y = (\nabla_x c)^{-1} (x, -\nabla \psi(x))$$

II. THE WASSERSTEIN SPACE

- Totally inadequate denomination for the space of probability measures, with the geometry induced by optimal transport
- Starts with Kantorovich, then many contributors
- Modern viewpoint emerges from Otto's work (1999) with many developments
- Still ongoing research

Reference: [oldnew, Chap. 6, 7, 15, 28]

Definition

 (\mathcal{X}, d) a Polish space, $1 \leq p < \infty$

$$P_p(\mathcal{X}) := \left\{ \mu \in P(\mathcal{X}); \int d(x_0, x)^p \, \mu(dx) < \infty \right\}$$

$$W_p(\mu, \nu) = \inf \left\{ \int d(x, y)^p \, \pi(dx \, dy); \, \pi \in \Pi(\mu, \nu) \right\}^{1/p}$$

p-Wasserstein space: $(P_p(\mathcal{X}), W_p)$.

This is a complete, separable metric space.

Rk:
$$P_p(\mathcal{X}) = \lim_{N \to \infty} (\mathcal{X}^N, d_p^{(N)}) / \mathcal{S}_N$$

where
$$d_p^{(N)}(x,y) = \left[\frac{1}{N} \sum_i d(x_i, y_i)^p\right]^{1/p}$$

and S_N = symmetric group

Topology of the Wasserstein space

 $C_p := \text{continuous functions } \mathcal{X} \to \mathbb{R} \text{ growing at most like}$ $d(x_0, x)^p$ at infinity

$$\mu_k \xrightarrow[k \to \infty]{} \mu \quad \text{iff} \quad \left[\int h \, d\mu_k \xrightarrow[k \to \infty]{} \int h \, d\mu \quad \forall h \in C_p \right]$$

Properties of the Wasserstein space

- $\mathcal{X} \subset P_p(\mathcal{X})$ via $x \longmapsto \delta_x$ (isometric embedding)
- $P_1(\mathcal{X})$ is the largest closed convex set generated by \mathcal{X} (in a Banach space)
- $P_p(\mathcal{X})$ has lost all topological information of \mathcal{X} :

M compact connected Riemannian manifold

- $\implies P_p(\mathcal{X})$ simply connected
- But $P_p(\mathcal{X})$ keeps track of metric properties of \mathcal{X}

Continuous dependence

Thm (Lott-V.) $\mathcal{X} \longmapsto P_p(\mathcal{X})$ is continuous (in Gromov-Hausdorff topology)

More precisely: Say p = 1.

Let $f: \mathcal{X} \to \mathcal{Y}$ be an ε -isometry, i.e.

$$\forall x, y \qquad \begin{cases} \left| d(f(x), f(y)) - d(x, y) \right| \le \varepsilon \\ d(f(\mathcal{X}), y) \le \varepsilon \end{cases}$$

Then $f_{\#}: P_1(\mathcal{X}) \to P_1(\mathcal{Y})$ is an ε -isometry

Recall: $\mathcal{X}_k \to \mathcal{X}$ in GH topology means \exists a ε_k -isometry $f_k : \mathcal{X}_k \to \mathcal{X}, \ \varepsilon_k \to 0$

Geodesic structure: Reminders

• $\gamma : [0,1] \to (\mathcal{X}, d)$ has length $\mathcal{L}(\gamma) = \sup \sum_{0=t_0 < t_1 < \dots < t_{N+1}=1} d(\gamma(t_i), \gamma(t_{i+1}))$

- γ is said geodesic if $\mathcal{L}(\gamma) = d(\gamma(0), \gamma(1))$
- (\mathcal{X}, d) is said geodesic if

 $\forall x, y \in \mathcal{X}, \exists \text{ geodesic } \gamma \text{ such that } \gamma(0) = x, \gamma(1) = y$

Geodesic structure of $P_p(\mathcal{X})$

Thm: If (\mathcal{X}, d) is geodesic then so is $P_p(\mathcal{X})$

Pf: Given μ, ν let $\pi(dx dy)$ be optimal. For any (x, y) choose measurably a geodesic $(\gamma_t(x, y))_{0 \le t \le 1}$ joining x to y. Let $\mu_t = (\gamma_t)_{\#}\pi$. Then $(\mu_t)_{0 \le t \le 1}$ is geodesic. \square

Thm (Lott-V): Any geodesic in $P_p(\mathcal{X})$ is generated in this way

"A geodesic in the space of laws is the law of a (random) geodesic"

Riemannian structure of $P_2(M)$

(Otto, 1999)

M a Riemannian manifold, $\mu \in P_2(M)$

 $\partial_t \mu$ an infinitesimal variation of μ in $P_2(M)$

$$\|\partial_t \mu\|_W^2 := \inf \left\{ \int |v|^2 d\mu; \quad \partial_t \mu + \nabla \cdot (\mu v) = 0 \right\}$$

This leads to a (formal) Riemannian calculus on $P_2(M)$

Theme (Otto–V. 2000) Geometric properties of M can be efficiently encoded/recovered by this structure on $P_2(M)$

We shall see illustrations in the sequel

The three levels of optimal transport

 e_t the evaluation at time t: $e_t(\gamma) = \gamma(t)$

Between μ_0 and μ_1 , one can consider:

- an optimal coupling $\pi(dx dy)$
- an interpolation $(\mu_t)_{0 \le t \le 1}$
- a dynamical optimal coupling: Π, probability measure

on geodesics, such that
$$\begin{cases} (e_0, e_1)_{\#}\Pi = \pi \\ (e_t)_{\#}\Pi = \mu_t \end{cases}$$

III. ISOPERIMETRIC-TYPE INEQUALITIES

- Optimal transport is an efficient way to derive/study some inequalities with isoperimetric content (e.g. Sobolev inequalities)
- Ancestors: Knothe, Gromov... ("reparametrization" = transport, not optimal)
- Otto-V (2000): via interpolation in Wasserstein space; many works since then
- **Reference:** [oldnew, Chap. 17, 20, 21]

Intuition

 (\mathcal{X}, d, ν) a Polish space equipped with a reference measure ν

$$U_{\nu}(\mu) = \int U\left(\frac{d\mu}{d\nu}\right) d\nu$$
 a "volumic energy"

$$\mu = \frac{1_A \nu}{\nu[A]} \implies U_{\nu}(\mu) = \nu[A] U\left(\frac{1}{\nu[A]}\right)$$

 $(\mu_t)_{0 \le t \le 1}$ with $\mu_0 = \mu$, $\mu_1 = \nu$ (assume $\nu \in P(\mathcal{X})$ for now)

 \longrightarrow the mass "flows from μ "

 $(d/dt)U_{\nu}(\mu_t)$ can be large only if the "surface of μ " is large \longrightarrow typically, an expression involving derivatives of $d\mu/d\nu$

Theorem

(M,g) a Riemannian manifold, $\nu(dx) = e^{-V(x)} \operatorname{vol}(dx)$

Assume

- $V \in C^2(M)$, Ric + $\nabla^2 V \frac{\nabla V \otimes \nabla V}{N-n} \ge Kg$ for some $N \in (1, +\infty]$, $K \ge 0$ (curvature condition)
- $U \in C(\mathbb{R}_+) \cap C^2((0, +\infty))$, convex, U(0) = 0, $r \longmapsto r^N U(r^{-N})$ convex (thermodynamic condition)
- $\mu_0 = \rho_0 \nu$, $\mu_1 = \rho_1 \nu$, where ρ_0 , ρ_1 decrease fast enough at infinity, ρ_0 Lipschitz

Then

$$U_{\nu}(\mu_{0}) - U_{\nu}(\mu_{1}) \leq \int U''(\rho_{0}(x_{0})) |\nabla \rho_{0}(x_{0})| d(x_{0}, x_{1}) \pi(dx_{0} dx_{1})$$

$$- \frac{K\lambda W_{2}(\mu_{0}, \mu_{1})^{2}}{2 \max(\|\rho_{0}\|_{L^{\infty}}, \|\rho_{1}\|_{L^{\infty}})^{\frac{1}{N}}}$$

where
$$\lambda = \lim_{r \to 0} \left(\frac{r U'(r) - U(r)}{r^{1 - \frac{1}{N}}} \right) \ge 0$$

Remarks

- Several variants (including K < 0)
- This formula extends beyond the Riemannian setting
- Moral: a geometric (curvature) condition together with a thermodynamic condition imply a convexity property of the energy functional (variant of the inequality $\Phi(y) \Phi(x) \ge \langle \nabla \Phi(x), y x \rangle$)
- Otto's calculus is sometimes a precious help for such statements
- Applications: see later

Sketch of proof (in \mathbb{R}^n with K = 0 for simplicity)

Strategy initiated by McCann

Step 1

Introduce a displacement interpolation (geodesic in $P_2(M)$) between μ_0 and μ_1 : here

$$\mu_t = (\mathrm{Id} + t\nabla\psi)_{\#}\mu_0 \qquad \nabla^2\psi \ge -I_n \qquad (0 \le t \le 1)$$

$$\implies f_0(x) = f_t(x + t\nabla\psi(x)) \det(I_n + t\nabla^2\psi(x))$$

$$f_t = d\mu_t/dx$$

$$\Longrightarrow \rho_0(x) = \rho_t(x + t\nabla\psi(x)) \det(I_n + t\nabla^2\psi(x)) e^{V(x) - V(x + t\nabla\psi(x))}$$

$$\rho_t = d\mu_t/d\nu$$

Step 2: Change variables

$$\int U(\rho_t(x)) \nu(dx)
= \int U(\rho_t(x+t\nabla\psi(x))) e^{-V(x+t\nabla\psi(x))} \det(I_n + t\nabla^2\psi(x)) e^{V(x)} \nu(dx)
= \int U\left(\frac{\rho_0(x)}{\mathcal{J}_t(x)}\right) \mathcal{J}_t(x) \nu(dx)$$

Jacobian: $\mathcal{J}_t(x) = e^{V(x) - V(x + t\nabla \psi(x))} \det(I_n + t\nabla^2 \psi(x))$

Step 3: Concavity estimates for Jacobian

Curvature
$$\Longrightarrow \frac{d^2}{dt^2} \mathcal{J}_t(x)^{\frac{1}{N}} \le -\frac{K}{N} \mathcal{J}_t(x)^{\frac{1}{N}}$$

Case V = 0 to simplify: $S \ge -I_n$, then

$$\frac{d}{dt}\det(I+tS) = \det(I+tS) \operatorname{tr}(S(I+tS)^{-1})$$

$$\frac{d^2}{dt^2} \left(\det(I + tS)^{\frac{1}{N}} \right) = \frac{1}{N} (\det(I + tS))^{\frac{1}{N}}$$

$$\left(\frac{1}{N} \left[\operatorname{tr} S(I + tS)^{-1} \right]^2 - \operatorname{tr} \left[S(I + tS)^{-1} S(I + tS)^{-1} \right] \right)$$

This is ≤ 0 by Cauchy–Schwarz

$$(A = S(I + tS)^{-1} \text{ symmetric, so } (\operatorname{tr} A)^2 \le n \operatorname{tr} A^2)$$

Step 4: Convexity estimates for U_{ν}

$$\int U(\rho_t(x)) \nu(dx) = \int U\left(\frac{\rho_0(x)}{\mathcal{J}_t(x)}\right) \mathcal{J}_t(x) \nu(dx)$$

$$= \int U\left(\frac{\rho_0(x)}{\mathcal{J}_t(x)}\right) \frac{\mathcal{J}_t(x)}{\rho_0(x)} \mu_0(dx)$$

$$= \int \Psi\left(\frac{\mathcal{J}_t(x)^{\frac{1}{N}}}{\rho_0(x)^{\frac{1}{N}}}\right) \mu_0(dx)$$

where $\Psi(r) = r^N U(r^{-N})$ convex nonincreasing

$$\begin{cases} t \longmapsto \mathcal{J}_t(x)^{1/N}/\rho_0(x)^{1/N} \text{ concave} \\ r \longmapsto \Psi(r) \text{ convex nonincreasing} \end{cases}$$

$$\implies \Psi(\mathcal{J}_t^{1/N}/\rho_0^{1/N}) \text{ convex function of } t$$

Step 5: Take the tangent

 $t \longmapsto U_{\nu}(\mu_t)$ convex implies

$$U_{\nu}(\mu_1) - U_{\nu}(\mu_0) \ge \frac{d}{dt} \Big|_{t=0} U_{\nu}(\mu_t)$$

 $t \longmapsto U_{\nu}(\mu_t)$ K-convex implies

$$U_{\nu}(\mu_1) - U_{\nu}(\mu_0) \ge \frac{d}{dt} \Big|_{t=0} U_{\nu}(\mu_t) + \frac{K}{2} W_2(\mu_0, \mu_1)^2$$

Step 6: Differentiate $U_{\nu}(\mu_t)$

$$U(\rho_{t}(x)) - U(\rho_{0}(x)) \geq U'(\rho_{0}(x)) \left[\rho_{t}(x) - \rho_{0}(x)\right]$$

$$\Longrightarrow \int U(\rho_{t}) d\nu - \int U(\rho_{0}) d\nu \geq \int U'(\rho_{0}) \rho_{t} d\nu - \int U'(\rho_{0}) \rho_{0} d\nu$$

$$= \int \left[U'(\rho_{0}(y)) - U'(\rho_{0}(x))\right] \pi_{t}(dx dy)$$

$$= \int \left[U'(\rho_{0}(\gamma_{t}(x, y))) - U'(\rho_{0}(\gamma_{0}(x, y)))\right] \Pi(d\gamma)$$

where Π is the dynamical optimal transference plan and $\gamma_t(x,y)$ has constant speed d(x,y)

$$\implies \frac{d}{dt} \bigg|_{t=0} \int U(\rho_t) \, d\nu \ge$$

$$- \int U''(\rho_0(x)) \, |\nabla \rho_0(x)| \, d(x,y) \, \pi(dx \, dy)$$

What has been achieved?

$$U_{\nu}(\mu_{0}) - U_{\nu}(\mu_{1}) \leq \int U''(\rho_{0}(x_{0})) |\nabla \rho_{0}(x_{0})| d(x_{0}, x_{1}) \pi(dx_{0} dx_{1})$$

$$- \frac{K\lambda W_{2}(\mu_{0}, \mu_{1})^{2}}{2 \max(\|\rho_{0}\|_{L^{\infty}}, \|\rho_{1}\|_{L^{\infty}})^{\frac{1}{N}}}$$

First particular case

Assume
$$N = \infty$$
, $\nu \in P_2(M)$, $U(r) = r \log r$, $\mu_1 = \nu$,
 $\mu_0 = \mu$, $\rho = d\mu/d\nu$, get
$$\int \rho \log \rho \, d\nu \le \int \frac{|\nabla \rho(x)|}{\rho(x)} \, d(x,y) \, \pi(dx \, dy) - \frac{K}{2} W_2(\mu,\nu)^2$$

$$\le \sqrt{\int \frac{|\nabla \rho(x)|^2}{\rho(x)^2} \, \pi(dx \, dy)} \, \sqrt{\int d(x,y)^2 \, \pi(dx \, dy)}$$

$$- \frac{K W_2(\mu,\nu)^2}{2}$$

$$= \sqrt{\int \frac{|\nabla \rho|^2}{\rho} \, d\nu \, W_2(\mu,\nu) - \frac{K W_2(\mu,\nu)^2}{2}}$$

This is the HWI inequality of Otto-V

HWI inequality

 $\nu = e^{-V}$ vol reference probability measure on (M, g)

Assume Ric + $\nabla^2 V \ge K g$, $K \in \mathbb{R}$

$$H_{\nu}(\mu) = \int \rho \log \rho \, d\nu, \qquad I_{\nu}(\mu) = \int \frac{|\nabla \rho|^2}{\rho} \, d\nu$$

Then
$$H_{\nu}(\mu) \leq W_2(\mu, \nu) \sqrt{I_{\nu}(\mu)} - \frac{K}{2} W_2(\mu, \nu)^2$$

About the HWI inequality

- A manifestation of the convexity of H_{ν} in $P_2(M)$ under a Ricci curvature condition
- Implies the Bakry-Émery theorem: if K > 0 then $H_{\nu}(\mu) \leq (2K)^{-1} I_{\nu}(\mu)$ (log Sobolev)
- Like log Sobolev, works well in infinite dimension
- In the limit $\mu \simeq \nu$, reduces to $||u||_{L^2} \le C ||\nabla^{-1}u||_{L^2} ||\nabla u||_{L^2}$

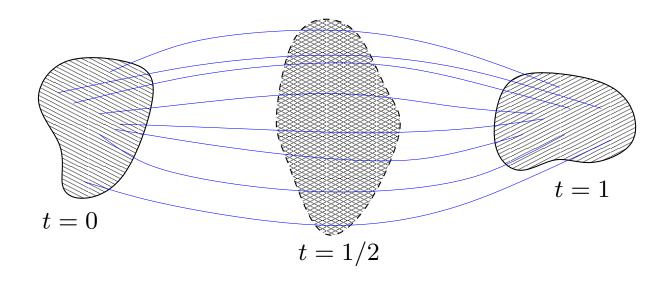
Use of HWI

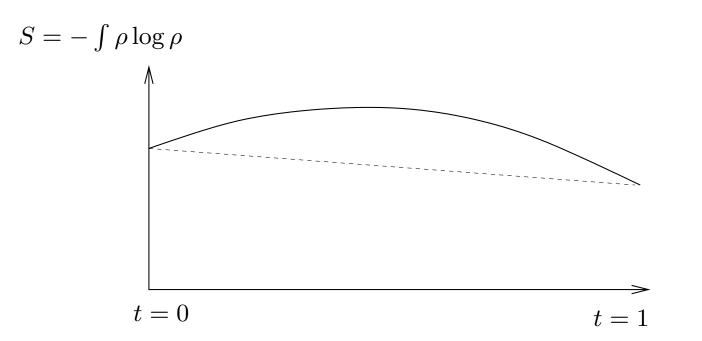
• Used by Gao and Wu to derive uniqueness criteria for certain spin systems; by Carrillo, McCann and V. to get convergence rates for certain nonlinear diffusion equations; by Grünewald, Otto, V. and Reznikoff-Westdickenberg to (re)prove convergence of microscopic entropy to macroscopic entropy in the hydrodynamic limit of Ginzburg-Landau particle system

Robustness

• The HWI inequality (and thus the Bakry-Émery theorem) remain true in a metric-measure space if one imposes the convexity properties of H_{ν} as definition of (Ricci) curvature bound

Lazy gas experiment





Another application of the convexity inequality

$$U_{\nu}(\mu_{0}) - U_{\nu}(\mu_{1}) \leq \int U''(\rho_{0}(x_{0})) |\nabla \rho_{0}(x_{0})| d(x_{0}, x_{1}) \pi(dx_{0} dx_{1})$$

$$- \frac{K\lambda W_{2}(\mu_{0}, \mu_{1})^{2}}{2 \max(\|\rho_{0}\|_{L^{\infty}}, \|\rho_{1}\|_{L^{\infty}})^{\frac{1}{N}}}$$

Choose
$$M = \mathbb{R}^{n}$$
, $\nu = \text{Lebesgue measure}$, $U(r) = -n r^{1-1/n}$, $K = 0$, $N = n$, get
$$n \int \rho_{1}^{1-\frac{1}{n}} - n \int \rho_{0}^{1-\frac{1}{n}}$$

$$\leq \left(1 - \frac{1}{n}\right) \int \rho_{0}(x_{0})^{-(1+\frac{1}{n})} |\nabla \rho_{0}(x_{0})| d(x_{0}, x_{1}) \pi(dx_{0} dx_{1})$$

$$\leq \left(1 - \frac{1}{n}\right) \left(\int \frac{|\nabla \rho_{0}|^{p}}{\rho_{0}^{p(1+\frac{1}{n})}} d\mu_{0}\right)^{\frac{1}{p}} \left(\int d(x_{0}, x_{1})^{p'} \pi(dx_{0} dx_{1})\right)^{\frac{1}{p'}}$$

$$= \left(1 - \frac{1}{n}\right) \left(\int \frac{|\nabla \rho_{0}|^{p}}{\rho^{p(1+\frac{1}{n})}} d\mu_{0}\right)^{\frac{1}{p}} W_{p'}(\mu_{0}, \mu_{1})$$

$$\frac{1}{n} + \frac{1}{n'} = 1, \qquad 1$$

Scaling argument: Replace $\rho_0(x)$ by $\lambda^n \rho_0(\lambda x)$, $\lambda \to \infty$:

$$\bullet \quad \int \rho_0^{1-\frac{1}{n}} = O(\lambda^{-1}) \longrightarrow 0$$

• $\int \frac{|\nabla \rho_0|^p}{\rho_0^{p(1+\frac{1}{n})}} \text{ remains constant}$

•
$$W_{p'}(\mu_0, \mu_1) \longrightarrow W_{p'}(\delta_0, \mu_1) = \left(\int |y|^{p'} \mu_1(dy) \right)^{1/p'}$$

Rewriting

Take the limit $\lambda \to \infty$.

Let $u = \rho_0^{1/p^*}$, $p^* = np/(n-p)$, the inequality becomes

$$1 \le \frac{p(n-1)}{n(n-p)} \left(\frac{\left(\int |y|^{p'} g(y) dy \right)^{1/p'}}{\int g^{1-\frac{1}{n}}} \right) \|\nabla u\|_{L^p}$$

as soon as $\int g = 1$, $\int u^{p^*} = 1$.

This is a Sobolev inequality

$$||u||_{L^{p^*}} \le C(n,p) ||\nabla u||_{L^p}$$

Further comments (in \mathbb{R}^n)

- This method provides optimal constants and works as soon as nonnegative curvature + homogeneity (say cone over a CD(n-2, n-1) manifold). Works for non-Euclidean norms
- The optimal transport approach was used by Cordero-Erausquin, Nazaret and V., then Maggi and V., to study various Sobolev inequalities (including solution of an old problem by Brézis-Lieb about trace Sobolev inequalities)
- Figalli, Maggi and Pratelli use the optimal transport to prove the stability of isoperimetric inequality:

$$|\partial\Omega| \ge \inf_{|B|=|\Omega|} \left\{ |\partial B| \left(1 + \text{const.} \left(\frac{|\Omega \Delta B|}{|\Omega|} \right)^2 \right) \right\}$$

IV. CONCENTRATION OF MEASURE

<u>Idea</u>

 (\mathcal{X}, d, ν) a metric probability space

$$A \subset \mathcal{X} \text{ with } \nu[A] \geq 1/2 \text{ (say)}$$

Then
$$\nu[A^r] \ge 1 - \varepsilon(r)$$
, $\varepsilon(r) \to 0 \text{ as } r \to \infty$

$$A^r = \left\{ x \in \mathcal{X}; \ d(x, A) \le r \right\}$$

(By enlarging a set with positive measure, one **invades** the whole space)

Equivalent formulation: for any (say) 1-Lipschitz function f

$$\nu \left[\left\{ x \in \mathcal{X}; \ f(x) \ge m + r \right\} \right] \le \varepsilon(r)$$

m = median or mean

About the concentration of measure

Many applications (Lévy, Milman, Gromov, Talagrand...)

Reference: [Ledoux] (AMS, 2001)

Two main types of concentration:

- Gaussian $\varepsilon(r) \simeq e^{-cr^2}$
- exponential $\varepsilon(r) \simeq e^{-cr}$ (for large r)

Optimal transport approach

 (\mathcal{X}, d, ν) a Polish probability space

Fact: One can encode concentration principles by transport-energy inequalities

$$\forall \mu \in P(\mathcal{X}), \qquad C(\mu, \nu) \leq \Phi(U_{\nu}(\mu))$$

Intuition

Take e.g. $\mu_A = \frac{1_A \nu}{\nu[A]}$, then the r.h.s. controls the volume of A while the l.h.s. says how easily one can invade the whole space, starting inside A

Pioneers: Marton, Talagrand

Reference: [oldnew, Chap. 22]

T_p inequalities

 (\mathcal{X}, d, ν) satisfies $T_p(K)$ (K > 0) if

(
$$T_p$$
) $\forall \mu \in P(\mathcal{X}), \qquad W_p(\mu, \nu) \leq \sqrt{\frac{2 H_{\nu}(\mu)}{K}}$

$$H_{\nu}(\mu) = \int \rho \log \rho \, d\nu \qquad \rho = \frac{d\mu}{d\nu}$$

$$W_p(\mu, \nu) = \inf \left\{ \int d(x, y)^p \, \pi(dx \, dy); \, \pi \in \Pi(\mu, \nu) \right\}^{1/p}$$

T_p implies Gaussian concentration

$$\nu[A] \ge \frac{1}{2}$$
 $B = \mathcal{X} \setminus (A^r)$

• $W_p(\mu_A, \mu_B) \ge r$

 $W_p(\mu_A, \mu_B) \le W_p(\mu_A, \nu) + W_p(\mu_B, \nu)$ $\le C \left(\sqrt{H_\nu(\mu_A)} + \sqrt{H_\nu(\mu_B)} \right)$ $= C \left(\sqrt{\log \frac{1}{\nu[A]}} + \sqrt{\log \frac{1}{1 - \nu[A^r]}} \right)$

$$\implies \nu[A^r] \ge 1 - e^{-cr^2}$$

About the (T_p) inequalities

- A functional way to encode concentration of measure
- Stronger as p increases
- p = 2 is critical because (T_2) is preserved under tensorization:

$$(\mathcal{X}, d, \nu)$$
 sat. $T_2(K) \Longrightarrow (\mathcal{X}^N, d_2^{(N)}, \nu^{\otimes N})$ also

... So T_2 provides dimension-free concentration

•
$$(T_1) \iff \int e^{a d(x_0, x)^2} \nu(dx) < +\infty$$

• (T_2) is more mysterious. Talagrand proves $T_2(1)$ for the usual Gaussian measure.

Otto-Villani theorem (1999)

$$LSI(K) \Longrightarrow T_2(K)$$

$$H_{\nu}(\mu) = \int \rho \log \rho \, d\nu$$
 $I_{\nu}(\mu) = \int \frac{|\nabla \rho|^2}{\rho} \, d\nu$

Why is it (hopefully) interesting?

• LSI is in terms of "local" quantities \Longrightarrow easy to perturb

Ex: If ν satisfies LSI then so does $e^{-\nu}\nu$, if v is bounded

- All known criteria for LSI apply
- Useful in the study of hydrodynamical limits of particle systems (Grünewald, Otto, Reznikoff-Westdickenberg, V.)

Proofs

Three genuinely different arguments known:

- heat semigroup method (Otto-V): works on a Riemannian manifold with Ricci curvature bounded below
- Hamilton-Jacobi semigroup method (Bobkov-Gentil-Ledoux): works on a geodesic space satisfying local doubling and Poincaré inequalities (Lott-V.)
- Large deviation method (Gozlan): works on any Polish space!

Sketch of Gozlan's proof of the Otto-V theorem

(Reverse things: concentration implies $T_2!$)

Step 1

Well-known (Herbst, Ledoux, Bobkov...)

$$(\mathcal{X}, d, \nu)$$
 sat. $LSI(K) \implies (\mathcal{X}^N, d_2^{(N)}, \nu^{\otimes N})$ sat. $LSI(K)$

$$\implies Gaussian concentration in $\mathcal{X}^N$$$

$$\forall N \in \mathbb{N} \quad \forall f \in \text{Lip}(\mathcal{X}^N, d_2^{(N)}),$$

$$\nu^{\otimes N} \left[\left\{ x \in \mathcal{X}^N; \ f(x) \ge m + r \right\} \right] \le e^{-\frac{K r^2}{2 \|f\|_{\text{Lip}}^2}}$$

Step 2

Let $f_N: (\mathcal{X}^N, d_2^{(N)}) \longrightarrow \mathbb{R}$ be defined by

$$f_N(x) = W_2(\widehat{\mu}_x^N, \nu)$$
 $\widehat{\mu}_x^N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$

$$||f_N||_{\text{Lip}} = \frac{1}{\sqrt{N}} \Longrightarrow \nu^{\otimes N} [f_N \ge m_N + r] \le e^{-\frac{K N r^2}{2}}$$

Step 3

• Start from $\nu^{\otimes N} [f_N \ge m_N + r] \le e^{-\frac{K N r^2}{2}}$

As $N \to \infty$, $\widehat{\mu}_x^N \longrightarrow \nu$ a.s. (Varadarajan), so $m_N \longrightarrow 0$

$$\Longrightarrow \liminf_{N\to\infty} \left(-\frac{1}{N} \log \nu^{\otimes N} \big[W_2(\widehat{\mu}_x^N, \nu) \ge r \big] \right) \ge \frac{K \, r^2}{2}$$

• Compare with Sanov:

$$\limsup_{N \to \infty} \left(-\frac{1}{N} \log \nu^{\otimes N} \left[W_2(\widehat{\mu}_x^N, \nu) \ge r \right] \right)$$

$$\le \inf_{\mu} \left\{ H_{\nu}(\mu); \ W_2(\mu, \nu) > r \right\}$$

• So $[W_2(\mu, \nu) > r] \Longrightarrow [H_{\nu}(\mu) \ge Kr^2/2],$ which is $T_2(K)$.

Conclusion: Chain of functional inequalities

$$CD(K, \infty) \implies (LS) \implies (T_2) \implies (P) \implies (\exp_1)$$
 $\downarrow \downarrow$
 $(T_1) \iff (\exp_2) \implies (\exp_1)$

 $(T_1) \iff (\mathcal{X}, d, \nu)$ Gaussian concentration profile e^{-cr^2}

$$(T_2) \iff (\mathcal{X}^N, d_2^{(N)}, \nu^{\otimes N})$$
 Gaussian profile e^{-cr^2} , for all N

(dimension-free Gaussian concentration)

Rk: Other criteria for exponential concentration, related to the quadratic-linear transport cost $c_{q\ell}(x,y) = \min(d(x,y),d(x,y)^2)$

V. PASSING TO THE LIMIT IN A CURVATURE CONDITION

Frequent situation

A "hard" property is equivalent to a "soft" property, which is well-adapted to pass to a weak limit

This principle will be illustrated on the Ma–Trudinger–Wang condition

Convention: geodesic = constant-speed minimizing geodesic

The Ma-Trudinger-Wang condition

$$(M,g)$$
 compact Riemannian manifold, $c(x,y) = \frac{d(x,y)^2}{2}$
 $\operatorname{cut}(M) = \left\{ (x,y) \in M \times M \text{ where } d \text{ fails to be smooth} \right\}$
 $(x,y) \notin \operatorname{cut}(M); x^1, \dots, x^n, y^1, \dots, y^n \text{ local coordinates}$
 $c_{i_1\dots i_k, j_1\dots j_\ell} := \frac{\partial^{k+\ell} c(x,y)}{\partial x^{i_1} \dots \partial x^{i_k} \partial y^{j_1} \dots \partial y^{j_\ell}} \quad [c^{i,j}] = [c_{i,j}]^{-1}$

For $(\xi, \eta) \in T_x M \times T_y M$, define

$$\mathfrak{S}(x,y) \cdot (\xi,\eta) := \frac{3}{2} \sum_{ijk\ell rs} \left(c_{ij,r} c^{r,s} c_{s,k\ell} - c_{ij,k\ell} \right) \xi^i \xi^j \eta^k \eta^\ell$$

(MTW) $\forall x, y, \xi, \eta$

$$\left[-\sum_{i,j} c_{i,j} \, \xi^i \, \eta^j = 0\right] \Longrightarrow \mathfrak{S}(x,y) \cdot (\xi,\eta) \ge 0$$

The Ma–Trudinger–Wang tensor $\mathfrak S$

- is a fourth-order, nonlocal, nonlinear expression of the Riemannian metric
- is covariant (independent of coordinate change) [Loeper, Kim-McCann]
- generalizes sectional curvature (Loeper): ξ, η two orthogonal unit vectors in T_xM

$$\Longrightarrow$$
 $\mathfrak{S}(x,x)\cdot(\xi,\eta)=\mathrm{Sect}(\{\xi,\eta\})$

So the MTW condition is stronger than (Sect ≥ 0).

It is satisfied e.g. by the sphere \mathbb{S}^n and its quotients.

Influence on regularity theory

(MTW) comes close to be equivalent to the smoothness of optimal transport.

- If it is violated, then $\exists f, g \in C^{\infty}(M)$, positive probability densities, such that the optimal transport map T between $\mu = f$ vol and $\nu = g$ vol, for the cost $c = d^2/2$, is discontinuous.
- If it is satisfied, one can hope that T is C^{∞} . This has been proven under "slightly" stronger assumptions.

References: [oldnew, Chap. 12]

+ recent papers by Delanoë, Figalli, Ge, Kim, Loeper, Ma, McCann, Rifford, Trudinger, V., Wang

Influence on geometry

$$\mathcal{V}(x) = \left\{ v = \dot{\gamma}(0); \ \gamma : [0, 1] \to M \text{ geodesic} \right\}$$

- $\mathcal{V}(x)$ is the manifold M, "written in T_xM "
- boundary of $\mathcal{V}(x) = \text{tangent cut locus of } x$, TCL(x)
- interior of $\mathcal{V}(x) = \text{tangent injectivity locus, } \text{TIL}(x)$

Open problem (Itoh-Tanaka): Is TCL(x) an Alexandrov space?

Conjecture (V.):

(MTW)
$$\Longrightarrow$$
 (CTIL) $\forall x$, TIL(x) is convex

Proven (Loeper-V.) under some additional assumptions ("nonfocalization" + strict MTW)

Stability of MTW

Question:
$$(M_k, g_k) \xrightarrow[k \to \infty]{} (M, g)$$

Assume M_k sat. MTW, does M satisfy MTW??

Natural topology: C^4 convergence of g_k .

Even in this topology, this is quite nontrivial, because of the cut locus analysis (focalization)

Passing to the Gromov–Hausdorff limit (V.)

Thm 1: If
$$(M_k, g_k) \xrightarrow[k \to \infty]{\text{GH}} (M, g)$$

and
$$M_k$$
 sat. (MTW) + (CTIL) + (Sect. \leq const.),

then also M satisfies these properties.

Thm 2: If
$$(M_k, g_k) \xrightarrow[k \to \infty]{\text{GH}} (M, g)$$

and M_k sat. (MTW) + (CTIL),

then also M satisfies (MTW).

Thm 1 is proven by metric geometry.

Thm 2 is proven by metric-measure analysis.

Comment: other examples of weak stability

Other curvature conditions passing to the limit:

- Sect $\geq \kappa$, or Sect $\leq \kappa$, under GH topology (Gromov...)
- Ric $\geq K$ or CD(K, N) under MGH topology (Lott-Sturm-V.)

Always same strategy: reformulate the differential condition into a robust synthetic property (e.g. Cartan—Alexandrov—Toponogov formulation of sectional bounds)

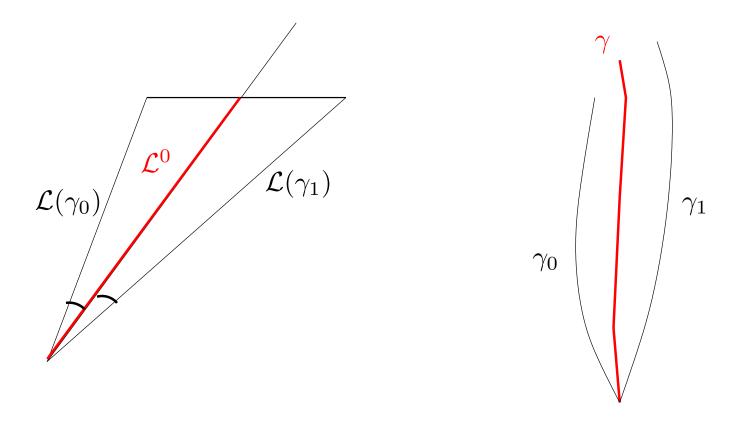
In the case of the Ma–Trudinger–Wang condition, we'll use some geometric properties studied by Loeper, Kim–McCann, V.

Metric reformulation

$$(MTW) + (CTIL)$$
 \iff

 $\forall \overline{x} \in M$, \forall pair of geodesics (γ_0, γ_1) originating from \overline{x} with angle $\theta \in (0, \pi)$, $\exists \gamma$ geodesic forming angles $\theta/2$ with γ_0 and γ_1 , with length $\mathcal{L}(\gamma) = \mathcal{L}^0$, s.t. $\forall x \in M$,

$$d(\overline{x}, \gamma(1))^{2} - d(x, \gamma(1))^{2} \leq \max \left(d(\overline{x}, \gamma_{0}(1))^{2} - d(x, \gamma_{0}(1))^{2}, d(\overline{x}, \gamma_{1}(1))^{2} - d(x, \gamma_{1}(1))^{2} \right)$$



Rk: Upper (and lower) sectional curvature bounds are used to pass to the limit in the angles

c-convexity reformulation

(MTW)
$$\begin{cases} \overset{\text{(CTIL)}}{\Longrightarrow} & \begin{cases} \forall \psi \ c\text{-convex}, \ \forall x \in M, \\ \frac{\partial_c \psi(x) \text{ is } (L\text{-Lipschitz-)connected} \end{cases} \end{cases}$$

Probabilistic reformulation (slightly cheating)

(MTW)
$$\begin{cases} \overset{\text{(CTIL)}}{\Longrightarrow} \\ & \Leftrightarrow \end{cases} \begin{cases} \forall \psi \text{ solution of the} \\ & \text{dual Kantorovich problem,} \\ \forall x \in M, \\ & \partial_c \psi(x) \text{ is } (L\text{-Lipschitz-}) \text{connected} \end{cases}$$

Passing to the limit

- L-Lipschitz-connectedness passes to the GH limit
- The dual Kantorovich problem passes to the GH limit

Gromov-Hausdorff stability of dual Kantorovich pb

- $(\mathcal{X}_k, d_k) \xrightarrow[k \to \infty]{GH} (\mathcal{X}, d) \text{ via } \varepsilon_k\text{-isometries } f_k : \mathcal{X}_k \to \mathcal{X}$
- $c_k(x,y) = d_k(x,y)^2/2$ on $\mathcal{X}_k \times \mathcal{X}_k$
- $\mu_k, \nu_k \in P(\mathcal{X}_k)$ $(f_k)_{\#}\mu_k \xrightarrow[k \to \infty]{} \mu, (f_k)_{\#}\nu_k \xrightarrow[k \to \infty]{} \nu$
- $\psi_k : \mathcal{X}_k \to \mathbb{R} \ c_k$ -convex, $\psi_k^{c_k}(y) = \inf_x [\psi_k(x) + c_k(x, y)],$

achieving
$$\sup \left\{ \int \psi_k^{c_k} d\nu_k - \int \psi_k d\mu_k \right\}$$

Then up to extr. $\exists a_k \in \mathbb{R} \text{ s.t. } (\psi_k - a_k) \circ f'_k \xrightarrow[k \to \infty]{} \psi,$

$$\psi$$
 c-convex achieving sup $\left\{ \int \psi^c d\nu - \int \psi d\mu \right\}$.

Moreover
$$\forall x \in \mathcal{X}$$
, $\limsup_{k \to \infty} f_k \Big(\partial_{c_k} \psi_k(f'_k(x)) \Big) \subset \partial_c \psi(x)$.