Mathematisches Forschungsinstitut Oberwolfach

Report No. 18/2009

DOI: 10.4171/OWR/2009/18

Arbeitsgemeinschaft: Optimal Transport and Geometry

Organised by
Felix Otto, Bonn
Karl-Theodor Sturm, Bonn

March 29th – April 4th, 2009

Abstract. This Arbeitsgemeinschaft was devoted to recent developments in optimal transport with emphasis on links and applications to geometry. The topics reached from the origin of optimal transport as a variational problem, where one minimizes a transportation cost when transporting one density into another, over the introduction of a metric on the space of probability measures, leading to the Wasserstein-space and convex funtionals on it, to the connection between Ricci-curvature and the optimal mass transport problem.

Mathematics Subject Classification (2000): 49-02 (28Axx 37J50 49Q20 53Cxx 58Cxx 82C70).

Introduction by the Organisers

In its origin, optimal transportation is a variational problem where one minimizes a transportation cost when transporting one density into another (Monge). Via its relaxed version (Kantorovich), the solution of this problem (Brenier) connects with convex analysis. Entire classes of inequalities in analysis can be easily proven with this tool. Even for the simplest transportation cost, i. e. the square of the Euclidean distance, the regularity theory for the minimizers is subtle (Caffarelli and others): Its Euler-Lagrange equation is the role model for a fully nonlinear elliptic equation in non-divergence form, the Monge-Ampère equation. The existence and the elements of a theory for more subtle transportation costs, like the Euclidean distance itself or the square of a Riemannian distance, are areas of current research.

Optimal transportation can be used to introduce a metric (distance function) on the space of probability measures which metrizes the weak topology. If the transportation cost is the square of a Euclidean or Riemannian distance, this metric can be seen as induced from a formal, infinite-dimensional Riemannian structure on the space of probability measures (Otto). Loosely speaking this geometry is the "complement" (in the sense of polar decomposition of vector fields) of the one of the space of volume-preserving diffeomorphisms (Arnold), which is motivated from fluid mechanics. Like for the space of volume-preserving diffeomorphisms, the space of probability measures has interesting geometrical properties itself. For instance, in the Aleksandrov sense, this space has non-positive sectional curvature, if the underlying space has this property.

Certain entropy functionals (including the usual entropy) turn out to be convex with respect to this geometry (McCann). Moreover, the convexity properties of these functionals can be used to characterize lower bounds on the Ricci curvature and the dimension of the underlying space — and can be used to define Ricci curvature bounds in the absence of a smooth structure (Sturm, Lott-Villani). This relation between geodesic convexity and Ricci curvature can be assimilated to the longer-known relation between the logarithmic Sobolev inequality and Ricci curvature (Bakry-Emery). Closely related to this property is the fact that the gradient flow (steepest descent) of the entropy functional is a contraction if the Ricci curvature is non-negative. In fact it is always a contraction if the underlying geometry evolves by Ricci flow (McCann-Topping).

This brief tour d'horizon shows that over the past 15 years, many connections between optimal transportation and seemingly unrelated fields have been discovered. Three monographs [1, 2, 3] and several lecture notes address these recent developments.

The Arbeitsgemeinschaft "Optimal transport and geometry", being held from March 29th to April 4th 2009, attracted more than 30 participants from various mathematical fields like partial differential equations, probability theory or geometry, who shared their different views about the topics presented in the 17 well-prepared talks in lively discussions. We would like to thank the staff of Oberwolfach for providing such perfect and pleasant working conditions, resulting in a stimulating atmosphere that gave the highly motivated participants extraordinary possibilities for learning and exchanging ideas.

References

[1] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2005).

[2] C. Villani, *Topics in Mass Transportation*, Graduate Studies in Mathematics, American Mathematical Society (2003).

[3] C. Villani, *Optimal Transport - Old and new*, Grundlehren, Springer, Berlin (2008).

Arbeitsgemeinschaft: Optimal Transport and Geometry

Table of Contents

Abstracts

Kantorovich Duality

Marzena Franek

The optimal transport problem was first formulated by the mathematician Gaspard Monge in the year 1781 and then reformulated by Kantorovich in 1942 [1, 2, 3]. In a simple way it can be interpreted as transporting a pile of sand to a hole, with the aim to minimize the transport costs.
Assume that μ and ν are two probability measures defined on two measure spaces X and Y. We model the transportation plan by probability measure π on a product space $X \times Y$.

$$\Pi(\mu,\nu) = \{\pi \in P(X \times Y) \mid \pi[A \times Y] = \mu[A], \quad \pi[X \times B] = \mu[B]\} \tag{1}$$

is the set of all probability measures with measurable subsets A of X and B of Y. With the measurable cost function $c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ we introduce the Kantorovich problem in the following way:

$$\min \; I[\pi] = \int_{X\times Y} c(x,y) d\pi(x,y) \quad \text{for } \pi \in \Pi(\mu,\nu). \tag{2}$$

If $T_c(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} I[\pi]$ exist, we name $T_c(\mu,\nu)$ the optimal transportation cost. The difference between Monge‘s optimal transport problem and Kantorovich‘s formulation is the fact that Monge‘s formulation does not allow the splitting of mass. The difficult nonlinear Monge problem was thereby replaced by a linear optimization problem over an abstract convex set, for which existence of a minimizer can be proved [2, 3].
In the year 1942 Kantorovich introduced the dual problem for (2). Let c be lower semicontinuous, so we can formulate the dual problem in the following way:

$$\sup_{\varphi,\psi} J(\varphi,\psi) = \sup_{\varphi,\psi} \left(\int_X \varphi(x) d\mu + \int_Y \psi(y) d\nu \right) \tag{3}$$

with measurable functions $(\varphi,\psi) \in L^1(d\mu) \times L^1(d\nu)$ under the assumptions $\varphi(x) + \psi(y) \leq c(x,y)$. We conclude that

$$\inf_\pi I[\pi] = \sup_{(\varphi,\psi)} J(\varphi,\psi). \tag{4}$$

For the proof we refer to [2].
If the cost function is a metric $c(x,y) = d(x,y)$ the Kantorovich problem admits a well-known dual formulation, introduced by Kantorovich and Rubinstein. With d a lower semi-continuous metric on X we can formulate

$$T_d(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times X} d(x,y) d\pi(x,y).$$

Let Lip(X) denote the space of all Lipschitz functions on X, and

$$\|\varphi\|_{Lip} \equiv sup_{x\neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)}.$$

Then

$$T_d(\mu,\nu) = \sup\left\{\int_X \varphi d(\mu-\nu); \varphi \in L^1(d|\mu-\nu|), \|\varphi\|_{Lip} \le 1\right\}.$$

References

[1] S. T. Rachev, L. Rüschendorf, *Mass Transportation Problems, Vol. 1: Theory*, Springer (1998).
[2] C. Villani, *Topics in Mass Transportation* , Graduate Studies in Mathematics, American Mathematical Society (2003).
[3] C. Villani, *Optimal Transport - Old and new* , Grundlehren, Springer, Berlin (1998).

Optimal transports and $c-$convex functions

Krzysztof Łatuszyński

The setting is as follows. Let $(\mathcal{X},\mu)$ and $(\mathcal{Y},\nu)$ be Polish probability spaces. Let functions $a : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ and $b : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be upper semicontinuous s.t. $a \in L^1(\mu)$, $b \in L^1(\nu)$. Let the cost function $c : \mathcal{X} \times \mathcal{Y} \to (-\infty,\infty]$ be lower semicontinuous and s.t. $c(x,y) \ge a(x) + b(y)$. By $\Pi(\mu,\nu)$ denote the family joint probability measures on $\mathcal{X} \times \mathcal{Y}$ with marginals μ and ν respectively. An element of Π will be denoted as π. From the first talk we know there exists $\pi \in \Pi(\mu,\nu)$ that minimizes $\mathbb{E}c(X,Y)$. Now we will discuss properties of an optimal plan π and in particular its support. The talk is based on [3], [2] and [1]. We start with definitions.

Definition 0.1.

- *A set $\Gamma \subset \mathcal{X} \times \mathcal{Y}$ is c−cyclically monotone if for any $N \in \mathbb{N}$ and for any $(x_1,y_1),\ldots,(x_N,y_N) \in \Gamma$ holds*

$$\sum_{i=1}^N c(x_i,y_i) \le \sum_{i=1}^N c(x_i,y_{i+1}) \qquad \text{with} \quad y_{N+1} \stackrel{def}{=} y_1.$$

- *We say that $\psi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is c−convex if it is $\psi \not\equiv +\infty$ and there exists $\xi : \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ such that for every $x \in \mathcal{X}$,*

$$\psi(x) = \sup_{y\in\mathcal{Y}}(\xi(y) - c(x,y)).$$

- *The c−transform of ψ is defined as*

$$\psi^c(y) = \inf_{x\in\mathcal{X}}(\psi(x) + c(x,y)).$$

- *The c−subdifferential of ψ is a c−cyclically monotone set defined by*

$$\partial_c\psi := \{(x,y) \in \mathcal{X} \times \mathcal{Y} : \psi^c(y) - \psi(x) = c(x,y)\},$$

and consequently the c−subdifferential of ψ at point x is

$$\partial_c\psi(x) := \left\{y \in \mathcal{Y}; (x,y) \in \partial_c\psi\right\} = \left\{y \in \mathcal{Y}; \forall_{z\in\mathcal{X}}, \psi(x) - \psi(z) \le c(z,y) - c(x,y)\right\}.$$

The $c-$transform can be seen as a general version of the Legendre transform. In particular if $\mathcal{X} = \mathcal{Y} = R^n$, then $c-$convexity is equivalent to convexity of $\psi(x)+|x|^2/2$. The following result provides an alternative, and often more tractable characterization of $c-$convexity.

Proposition 0.2. *Function ψ is $c-$convex if and only if $\psi^{cc} = \psi$, where*

$$\psi^{cc}(x) := (\psi^c)^c(x) = \sup_{y\in\mathcal{Y}} \inf_{\tilde{x}\in\mathcal{X}} (\psi(\tilde{x}) + c(\tilde{x}, y) - c(x, y)).$$

Sketch of a proof. First observe that for any $\phi : \mathcal{Y} \to \mathbb{R} \cup \{\infty\}$ one can write ϕ^{ccc} as sup inf sup and conclude that $\phi^{ccc} = \phi^c$. (c.f. [3] Proposition 5.8). Since ψ is $c-$convex, $\psi = \xi^c$ for some ξ. Thus $\psi^{cc} = \xi^{ccc} = \xi^c = \psi$. For the converse, it is enough to notice, that if $\psi^{cc} = \psi$, then ψ is $c-$convex as the $c-$transform of ψ^c. □

Next we show the Rüschendorf's theorem, namely that every $c-$cyclically monotone set can be included in a $c-$subdifferential of a $c-$convex function.

Theorem 0.3 (Rüschendorf's Theorem)**.** *A set $\Gamma \subset \mathcal{X} \times \mathcal{Y}$ is $c-$cyclically monotone $\Leftrightarrow$ there exists a $c-$convex function ψ on $\mathcal{X}$ such that $\Gamma \subset \partial_c\psi$.*

Proof. $\Leftarrow$. Take $(x_i, y_i) \in \Gamma \subset \partial_c\psi,\ 1 \le i \le n$. Then the definition of $\partial_c\psi(x_i)$ implies

$$\sum_{i=1}^{n} \Big(c(x_{i+1}, y_i) - c(x_i, y_i)\Big) \ge \sum_{i=1}^{n} (\psi(x_i) - \psi(x_{i+1})) = 0,$$

so Γ is $c-$cyclically monotone.
$\Rightarrow$. First note that $c-$convexity of ψ, is equivalent to $\psi(x) = \sup_{i\in I}(a_i - c(x, y_i))$, where I is an index set. Indeed, for $i \in I$ one can take $\xi(y_i) := a_i$ and $\xi(y) = -\infty$ otherwise. Conversely, if ψ is $c-$convex, then one can take $I := \mathcal{Y},\ y_y = y$ and $a_y = \xi(y)$. Now assuming that Γ is $c-$cyclically monotone and $(x_0, y_0) \in \Gamma$, define $\psi : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ as

$$\psi(x) := \sup_{\substack{(x_i,y_i)\in\Gamma \\ 1\le i\le n\in\mathbb{N}}} \Big(-c(x, y_n)+c(x_n, y_n)-c(x_n, y_{n-1})+c(x_{n-1}, y_{n-1}) \mp\cdots - c(x_1, y_0)+c(x_0, y_0)\Big).$$

Then ψ is $c-$convex and $\psi(x_0) \le 0$ as Γ is $c-$cyclically monotone. In fact, taking $n = 1,\ x_1 = x_0,\ y_1 = y_0$ gives $\psi(x_0) = 0$. Next we take $(x', y') \in \Gamma$ and show that $y' \in \partial_c\psi(x')$ and as a consequence $\Gamma \subset \partial_c\psi$. To this end take $\lambda < \psi(x')$, then there exist $(x_i, y_i) \in \Gamma,\ 1 \le i \le m \in \mathbb{N}$ such that

$$\lambda < -c(x', y_m)+c(x_m, y_m)-c(x_m, y_{m-1})+c(x_{m-1}, y_{m-1})\mp\cdots-c(x_1, y_0)+c(x_0, y_0).$$

Define $x_{m+1} := x'$ and $y_{m+1} := y'$. Then for $x \in \mathcal{X}$,

$$\begin{aligned}\psi(x) &\ge -c(x, y_{m+1}) + c(x_{m+1}, y_{m+1}) - c(x_{m+1}, y_m) + c(x_m, y_m) \mp \cdots - c(x_1, y_0) + c(x_0, y_0)\\ &\ge -c(x, y_{m+1}) + c(x_{m+1}, y_{m+1}) + \lambda.\end{aligned}$$

Thus $\psi(x) - \lambda \ge -c(x, y_{m+1}) + c(x_{m+1}, y_{m+1})$ and by letting $\lambda \to \psi(x')$ we obtain $\psi(x) - \psi(x') \ge -c(x, y') + c(x', y')$ as required. Moreover $\psi(x') < \infty$ since $\psi(x_0) = 0$. □

It is intuitively clear and also not very hard to prove formally that optimal plans have $c-$cyclically monotone supports under some regularity conditions for the cost function. So in view of the Rüschendorf's Theorem, optimal plans are supported in $c-$subdifferentials of $c-$convex functions. It is also intuitive, but much less obvious, that $c-$cyclically monotone plans are optimal. This property is stated precisely in the next Theorem (Theorem 5.10 of [3]). Here the function $\psi(x)$ can be interpreted as the price for buying goods at x and $\phi(y)$ as the price for selling the goods at y.

Theorem 0.4. *Let $(\mathcal{X}, \mu)$, $(\mathcal{Y}, \nu)$, c, a, b be as before. Then*

(1) *It c is real-valued and the optimal cost $C(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int c d\pi$ is finite, then there is a measurable $c-$cyclically monotone set $\Gamma \subset \mathcal{X} \times \mathcal{Y}$ (closed if a, b, c are continuous) s.t. for any $\pi \in \Pi(\mu, \nu)$ the following statements are equivalent.*
   - (a) *π is optimal;*
   - (b) *The support of π is $c-$cyclically monotone;*
   - (c) *There is a $c-$convex ψ s.t., $\pi-$a.s. $\psi^c(y) - \psi(x) = c(x, y)$;*
   - (d) *There exists $\psi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ and $\phi : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$, s.t. $\phi(y) - \psi(x) \leq c(x, y)$ for all (x, y), with equality $\pi-$a.s.*
   - (e) *π is concentrated on Γ.*

(2) *If c is real-valued, $C(\mu, \nu) < \infty$, and $c(x, y) \leq c_{\mathcal{X}}(x) + c_{\mathcal{Y}}(y)$, with $(c_{\mathcal{X}}, c_{\mathcal{X}}) \in L^1(\mu) \times L^1(\nu)$, then both the primal and the dual Kantorovich problems have solutions, so*

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y)$$
$$= \max_{(\psi,\phi) \in L^1(\mu) \times L^1(\nu); \phi - \psi \leq c} \left(\int_{\mathcal{Y}} \phi(y) d\nu(y) - \int_{\mathcal{X}} \psi(x) d\mu(x) \right)$$
$$= \max_{\psi \in L^1(\mu)} \left(\int_{\mathcal{Y}} \psi^c(y) d\nu(y) - \int_{\mathcal{X}} \psi(x) d\mu(x) \right),$$

and in the latter expressions one might as well impose that ψ be $c-$convex and $\phi = \psi^c$. If a, b, c are continuous, then there exists a closed $c-$cyclically monotone set $\Gamma \subset \mathcal{X} \times \mathcal{Y}$, s.t. for any $\pi \in \Pi(\mu, \nu)$ and for any $\psi \in L^1(\mu)$, we have
   - (a) *π is optimal in the Kantorovich problem iff $\pi[\Gamma] = 1$,*
   - (b) *ψ is optimal in the dual Kantorovich problem iff $\Gamma \subset \partial_c \psi$.*

References

[1] L. Rüschendorf, *On c−optimal random variables*, Stat. and Probab. Letters **27** (1996), 267–270.

[2] C. Villani, *Topics in Optimal Transportation*, vol. 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI (2003).

[3] C. Villani, *Optimal Transport, Old and New*, vol. 338 of *Grundlehren der mathematischen Wissenschaften* Springer, (2009).

Brenier's solution to the optimal transport problem in the Euclidean case, Polar factorization of vector-valued maps

Christian Selinger

For given probability measures μ and ν on $\mathbb{R}^d$ with finite second moment and the cost function $c : \mathbb{R}^d \times \mathbb{R}^d \ni (x,y) \mapsto \frac{||x-y||^2}{2} \in \mathbb{R}$ we know via Kantorovich's duality (see for instance [2]) that there exists a unique solution to the associated primal and dual optimal transport problem

$$\inf_\pi \int_{\mathbb{R}^d \times \mathbb{R}^d} c\, d\pi = \sup_{(\phi,\psi)} \left\{ \int_{\mathbb{R}^d} \phi\, d\mu + \int_{\mathbb{R}^d} \psi\, d\nu \right\},$$

where the infimum is taken over all product probability measures with given marginals μ and ν and the supremum is taken over all μ- resp. ν-integrable functions satisfying $\left\{\frac{||x||^2}{2} - \phi(x)\right\} + \left\{\frac{||y||^2}{2} - \psi(y)\right\} \leq c(x,y)$. Provided the above conditions Brenier showed in [1] the following

Theorem 1 If μ does not give mass to small sets then there exists a unique optimal transport plan π minimizing the transport cost, actually this plan turns out to be a map:

$$d\pi(y,z) = \delta(z - \nabla\phi(y))\mu(dy),$$

where $\nabla\phi$ is the μ-almost everywhere unique gradient of a convex function with $\nabla\phi\#\mu = \nu$. Reciprocally if ν does not give mass to small sets neither the optimal transport problem from ν to μ is solved by the $d\pi(y,z) = \delta(y - \nabla\phi^*(z))\nu(dz)$.

Proof. To prove this, two elementary lemmata come into play: Firstly, let us denote the Legendre transform of an integrable function f by $f^*(y) := \sup_x\{x.y - f(x)\}$. It is an easy calculation to show that $f(x) + f^*(y) = x.y$ if and only if $x \in \partial f^*(y) := \{z : f^*(y') - f^*(y) \geq z.(y'-y) \forall y' \in \mathbb{R}^d\}$ and $y \in \partial f(x)$. Secondly, Rademacher's theorem guarantees almost everywhere differentiability of Lipschitz continuous functions.
Now we can give a sketch of the proof: First we show that the solution to the dual problem is given by the pair (ϕ, ϕ^*): By definition $\psi^* \leq \psi$ and ψ^* is Lipschitz and respective equations hold for $(\psi^*)^*$. The existence of a solution to the dual Monge-Kantorovich problem let us conclude that $\phi = \psi^*$ μ-a.e. and that $\psi^*(y) + (\psi^*)^*(z) = y.z$. Furthermore $\psi = (\psi^*)^*$ a.e., the right-hand side being continuous and Lipschitz and the left hand side being convex by definition. By Rademacher $\nabla\psi$ is well-defined up to Lebesgue negligible sets and $\partial\psi(z) = \{\nabla\psi(z)\}$ a.e. From the first lemma mentioned at the beginning of the proof we obtain finally $dp(y,z) = \delta(z - \nabla\phi(y))\mu(dy)$. The reciprocal statement follows by the same reasoning in terms of the optimal pair (ψ^*, ψ). □

Proposition 1 (Stability) Given a lower semi-continuous cost function on $\mathbb{R}^d$ and probability measures μ and $\{\nu_k; k\mathbb{N}\}$ such that ν_k converges weakly to ν. Asumme furthermore that for each k there exists an optimal transport map T_k

from μ to ν_k and that there exists as well an optimal transport map T from μ to ν. Then T_k converges to T in probability, i.e. for any $\epsilon > 0$:

$$\mu\left(\{x \in \mathbb{R}^d : |T_k(x) - T(x)|^2 \geq \epsilon\}\right) \longrightarrow 0$$

Proof. With T_k and T we can build the unique optimal transport plans $p_k := (\mathrm{id}, T_k)\#\mu$ and $p := (\mathrm{id}, T)\#\mu$. Let $\epsilon > 0, \delta > 0$. By Lusin's theorem there exists a closed $K \subset \mathbb{R}^d$ such that $\mu(\mathbb{R}^d \setminus K) \leq \delta$ and $T|_K$ is continuous. It follows that

$$A_\epsilon := \left\{(x, y) \in K \times \mathbb{R}^d : |T(x) - y|^2 \geq \epsilon\right\} \subset K \times \mathbb{R}^d$$

is closed and

$$0 = p(A_\epsilon) \geq \limsup_k p_k(A_\epsilon) \geq \limsup_k \mu(\{x \in \mathbb{R}^d : |T(x) - T_k(x)|^2 \geq \epsilon\}) - \delta.$$

Letting δ tend to zero achieves the proof. □

As an application of Brenier's characterization of optimal transport plans we state

Theorem 2 (Polar factorization of vector-valued maps) Given two probability measures μ and ν satisfying the assumptions of Theorem 1, a Polish probability space $(X, \xi) \simeq ([0,1], |.|)$, the Hilbert space $H = L^2((X, \xi); \mathbb{R}^d)$ and $S = \{s : X \to \mathbb{R}^d; s\#\xi = \mu\}$ then, up to a negligible set $N = \{u \in H : u\#\xi(E) \neq 0$ for all E with $|E| = 0\}$, there exists the unique factorization

$$H \ni u = \nabla\psi \circ s,$$

where ψ is a convex function.

Proof. By Theorem 1 the optimal transport from μ to ν is given by the almost everywhere unique gradient of a convex function ψ. For the reciprocal transport the map $\nabla\psi^*$ is optimal and defining $s := \nabla\psi^* \circ u$ yields the factorization. □

Remark (Recovering the Helmholtz decomposition of vector fields) Let z be smooth vector field on $\mathbb{R}^d$ we define a smooth perturbation of the identity map for sufficiently small ϵ: $u(x) := x + \epsilon z(x)$. By Theorem 2 we know that $u = \nabla\psi \circ s$ and for small ϵ we can choose $\psi(x) = |x|^2/2 + \epsilon p(x) + o(\epsilon^2)$ resp. $s(x) = x + \epsilon d(x) + o(\epsilon^2)$ for a smooth function p and a vector field d. Now we have $u(x) = \nabla\psi(s(x)) = x + \epsilon(d(x) + \nabla p(x)) + o(\epsilon)$ wich entails $z = d + \nabla p$. Since s is measure preserving it holds for any bounded measurable f that $\int f(s(x))dx = \int f(x + \epsilon d(x) + o(\epsilon))dx = \int f(x)dx$ consequently the divergence of d is zero.

References

[1] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions.*, Comm. Pure Appl. Math. **44** (1991), 375-417.

[2] C. Villani, *Optimal transport, old and new*, Springer (2008).

The Monge-Ampère equation

Christian Seis

The Monge-Ampère equation is a nonlinear partial differential equation, which appears in different mathematical contexts. In optimal transportation, its special form

$$g(\nabla\varphi(x)) \det D^2\varphi(x) = f(x) \tag{1}$$

arises naturally from Brenier's solution $\nabla\varphi$, φ convex, to the Monge problem in the case of absolutely continuous measures $d\mu = f\,dx$, $d\nu = g\,dy$: At least formally (assuming φ strictly convex, smooth), (1) is a reformulation of $\nabla\varphi\#\mu = \nu$.

However, in general φ is only convex, and thus continuous and locally Lipschitz, but not necessarily twice differentiable. Therefore, $\det D^2\varphi$ is not well-defined. It turns out, that φ solves the Monge-Ampère equation (1) pointwise almost everywhere [3, Theorem 4.8 (iii)], if the Hessian $D^2\varphi$ is interpreted in the sense of Alexandrov's second derivative of convex functions, this is the absolutely continuous part of the distributional Hessian.

The proof makes use of the Hessian measure $\det_H D^2[E] = |\partial\varphi(E)|$. A crucial observation is the identification of the absolutely contiuous part of the Hessian measure with $\det D^2\varphi(x)\,dx$ by using the Radon-Nikodym theorem. It is convenient to restrict the analysis to a nice set M of points, where $D^2\varphi(x)$ exists and is invertible. Thanks to the above identification, the Hessian measure is concentrated on M. Applying once more the Radon-Nikodym theorem, a change of variable formula follows for $\nabla\varphi$.

Together with (1), McCann's change of variable formula [3, Theorem 4.8 (iv)]

$$\int U(g(y))\,dy = \int U\left(\frac{f(x)}{\det D^2\varphi(x)}\right) \det D^2\varphi(x)\,dx,$$

for nonnegative functions U with $U(0) = 0$ is derived. See also [2, Theorem 4.4].

As an application an elemantary proof of the Sobolev inequality [1, Theorem 2]

$$\|f\|_{L^{p^*}} \leq C_{n,p}\|\nabla f\|_{L^p},$$

with optimal constant $C_{n,p}$ is presented. Starting point is (1) for appropriate probability densities. The determinant can be estimated against the Laplacian using the arithmetic-geometric inequality.

References

[1] D. Cordero-Erausquin, B. Nazaret, C. Villani, *A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities*, Advances in Math **182** (2004), 307–332.

[2] R. McCann, *A convexity principle for interacting gases*, Advances in Mathematics **128** (1997), 153–179.

[3] C. Villani, *Topics in optimal transportation*, Graduate studies in Mathematics, American Mathematical Society (2003).

Caffarelli's regularity theory

Bernd Kirchheim, Thomas Viehmann

In the previous talk we have seen that the convex potentials φ of Brenier's solutions satisfy a partial differential equation, the Monge-Ampère equation, but only in a very weak sense. Also, Brenier's solutions yield a transport plan T as the gradient of the (Lipschitz) transportation potential. This talk discusses the question of C^1-regularity of the potential or, equivalently, continuity of T. These are the first steps of Caffarelli's regularity theory.

A simple example transporting mass from a ball to two separated half-balls shows that the transport map cannot in general expected to be continuous. However, this example can be refined to one in which the target measure has connected support. The example suggests that convexity of the support is the crucial property. We thus assume that source and target measure have convex supports and densities w.r.t. the Lebesgue measure that are bounded from above and below on their respective support.

In order to use the regularity theory for nonlinear elliptic partial differential equations we motivate that under these assumptions Alexandrov solutions enjoy comparison principles in the class of convex functions. In the language of PDEs they are viscosity solutions to the elliptic equation

$$0 < \lambda_1 \leq \det D^2\varphi \leq \lambda_2. \tag{1}$$

The first step is to construct suitable barriers (i.e comparison functions) in a very special geometry of regions comparable with balls. A crucial observation is now, that the existence of F.John's ellipsoid allows us to transform these estimates to general convex domains, obtaining a result of the following kind

Theorem *There is a $\mu_0 = \mu_0(\lambda_1, \lambda_2, n) > 0$ such that the following holds. Let φ be a convex function on a convex (bounded) body $\Omega \subset \mathbb{R}^n$ such that $\varphi_{|\partial\Omega} \equiv 0$ and φ a viscosity solution of (1). If $x, y \in \partial\Omega$ and $z = \mu x + (1-\mu)y \in \Omega$ with $\varphi(z) < \frac{1}{2}\min_\Omega \varphi$ then $\mu_0 \leq \mu \leq 1 - \mu_0$.*

Since the class of convex solutions of (1) is invariant under tilting, one obtains from this result that such functions must have a unique supporting plane, i.e. be differentiable and hence in C^1.

Similarly, using the Theorem one can show that for such a convex solution of (1) the (convex) set where φ attains its minimum can not have an extreme point inside Ω. From this we could similarly derive that the canonical globally lipschitz extension of φ to all of $\mathbb{R}^n$ would have to vanish on an entire line. But then φ has to constant along any parallel line and can not solve (1). Therefore, all solutions of (1) have to be strictly convex.

References

[1] L. A. Caffarelli, *A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity*, Annals of Mathematics **131** (1990), 129–134.

[2] L. A. Caffarelli, *The regularity of mappings with a convex potential*, Journal of the AMS **5** (1992), 99–104.

Optimal transport on Riemannian manifolds

Martin Huesmann

The talk and therefore this report are basically summaries of the very well written paper [1].

Throughout this report we consider a compact, connected, C^3-smooth (meaning that the metric tensor is twice continuously differentiable) Riemannian manifold (M, g) without boundary together with the cost function $c(x,y) = \frac{1}{2}d^2(x,y)$, one half times geodesic distance squared. Let μ and ν be two probability measures on M. $S(\mu,\nu)$ is the set of all Borel maps which push μ forward to ν. For $s \in S(\mu,\nu)$ we write $C(s) = \int_M c(x, s(x))d\mu(x)$. For the same probability measures, we set $Lip_c = \{u, v : M \to \mathbb{R} \quad \text{continuous}; u(x) + v(y) \leq c(x,y)\}$ and $J(u,v) = \int_M u d\mu + \int_M v d\nu$. Then the Kantorovich duality reads

$$\sup_{(u,v)\in Lip_c} J(u,v) \leq \inf_{s\in S(\mu,\nu)} C(s).$$

Given a pair $(u,v) \in Lip_c$ one can improve this pair by infimal convolution, i.e. $J(u,v) \leq J(u,u^c)$ and also $J(u,v) \leq J(v^c,v)$, where the infimal convolution is defined as $w^c(y) := \inf_{x\in M} c(x,y) - w(x)$. We say a function u on M is c-concave iff $u^{cc} = (u^c)^c = u$. We have the following lemma

Lemma 0.5. *Let (M,d) be a metric space with finite diameter. Then any c-concave ($c = \frac{d^2}{2}$) function u is either identically infinite, $u \equiv \pm\infty$, or Lipschitz continuous throughout.*

By Rademacher‘s theorem this directly implies that c-concave functions having values in the reals are differentiable almost everywhere and their gradient is Borel measurable on the set where it is defined. However, this statement is not strong enough to characterise the direction and distance of optimal transportation. For this, we need the notion of superdifferentiability.
We say a function $u : M \to \mathbb{R}$ is superdifferentiable in x with supergradient p if for $v \in T_xM$

$$u(\exp_x v) \leq u(x) + < p, v >_x + o(|v|_x).$$

Then one can prove using a Taylor expansion of the exponential map

Lemma 0.6. *Let (M,g) be as above and let $\sigma : [0,1] \to M$ be a minimizing geodesic from y to x parametrized with constant speed. Then $u(\cdot) = \frac{d^2}{2}(\cdot, y)$ is superdifferentiable with supergradient $\dot{\sigma}(1)$ at x.*

This enables us to prove the following uniqueness result.

Theorem 0.7 (uniqueness). *Let (M, g) be as above, $\mu << vol$ (the normalised Riemannian volume measure), $u : M \to \mathbb{R}$ c-concave. Then, $t(x) = \exp_x(-\nabla u(x))$ is the minimizer for $\inf_{s \in S(\mu, t_{\#}\mu)} C(s)$. Moreover, any other map $s \in S(\mu, t_{\#}\mu)$ must coincide with t μ a.e..*

The key ingredient to the proof of this theorem is the following tangency lemma. Together with the Kantorovich duality it directly gives the result. Furthermore, it also shows that in the Kantorovich duality we actually have an equality not just an inequality.

Lemma 0.8 (tangency). *(M, g) as above, $u : M \to \mathbb{R}$ c-concave. Then, $c(x, y) - u(x) - u^c(y) \geq 0$ for all $x, y \in M$. If u is differentiable in x, then equality holds iff $y = \exp_x(-\nabla u(x))$.*

The first part is obvious from the definition of infimal convolutions. The other part is not. Here one has to use the superdifferentiability of the cost function and the compactness of M together with the continuity of u. As u is c-concave, it is differentiable almost everywhere. Thus, we have almost everywhere equality for the pairs $(x, t(x))$ proving the first part of the theorem. The uniqueness statement also follows from this lemma as any other optimal map also has to satisfy the Kantorovich duality and therefore the equality in the tangency lemma which characterizes the map s almost everywhere.
For a given potential function we have found an optimal map between a given measure μ and a certain pushforward of this measure. That for any Borel measure such a potential function can be found, is the content of the next theorem.

Theorem 0.9. *Let (M, g) be as above, $\mu << vol$ and ν be an arbitrary Borel measure. Then, there is a c-concave function $u : M \to \mathbb{R}$ such that $t(x) = \exp_x(-\nabla u(x)) \in S(\mu, \nu)$ is optimal and unique a.e..*

By the first theorem uniqueness is clear as soon as one has found a suitable potential function. As this function has to maximize $J(u, u^c)$ (wlog we can take $v = u^c$ as $J(u, v) \leq J(u, u^c)$), the maximizer is the natural candidate. That this maximizer exists is a consequence of the compactness of the manifold. Then, one checks that the pushforward of μ by the exponential of the negative gradient of the maximizer coincides with ν.
If one additionally assumes that $\nu << vol$ as well, then one can use the symmetry of the problem to derive

Corollary 0.10. *Let $(M, g), u, t$ be as above, $\mu, \nu << vol$. Then $t^*(y) = \exp_y(-\nabla u^c(y)) \in S(\nu, \mu)$ is optimal and satisfies $t^*(t(x)) = x$ μ a.e. and $t(t^*(y)) = y$ ν a.e..*

This symmetry and these porperties are exactly those needed to generalise Brenier's polar factorization presented in talk number 3 to the setting of compact connected Riemannian manifolds.

The techniques used in this talk are not restricted to the case of the quadratic cost function. The key ingredients which are used are:

- compactness of M
- existence of minimizing geodesics between any two points
- superdifferentiability of the cost function

The first two points can be replaced by the same properties for the support of the two measures one considers. The last point is crucial because it characterizes the distance and direction of the optimal transport. However, there are many cost functions which are superdifferentiable, e.g.

$$c(x,y) = \int_0^{d(x,y)} f(\tau)d\tau$$

where f is continuously increasing. Of course, this yields another transport map. By mimicking the tangency lemma one easily finds the correct map, e.g. the map for the cost above would be -if everything else is left the same- $t(x) = \exp_x[-\frac{f^{-1}(|\nabla u|_x)}{|\nabla u|_x}\nabla u(x)]$.

References

[1] R.J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. 11 no.3 (2001), 589-608.

The solution to the original Monge problem

Clément Hongler

The original Monge problem, formulated in 1781 is the following: given two probability measures μ_1 and μ_2 supported on a convex compact set $X \subset \mathbb{R}^n$ with μ_0 absolutely continuous with respect to the Lebesgue measure, does there exist a map $\varphi : X \to X$ with $\varphi_{\#}(\mu_1) = \mu_2$ that minimizes the linear transport cost $\int_X |\varphi(x) - x|\, d\mu_1(x)$?

In one dimension such a map exists and one can give an explicit expression for it (although this map is not unique in general).

This existence problem was solved affirmatively in 1972 by Sudakov (see [2]); a gap in proof was found later and addressed by Ambrosio in 1992. We sketch here the explicit construction presented by Sudakov. By general techniques, there exists an optimal transport plan, or transport coupling, γ for μ_1 and μ_2, i.e. a probability measure on $X \times X$ with $\pi_1(\gamma) = \mu_1$ and $\pi_2(\gamma) = \mu_2$ (where $\pi_1, \pi_2 : X \times X \to X$ denote the projections on the first and second coordinates respectively) such that $\int_{X\times X} |x - y|\, d\gamma(x, y)$ is minimal in the set of all couplings of μ_1 and μ_2. We will construct a map with a transport cost reaching this minimum.

We can assume without loss of generality that for each $x \in X$ there exists $y \neq x$ such that $(x, y) \in \mathrm{Supp}(\gamma)$: the optimal transport map on the set of points where this condition is not satisfied is simply the identity map, and we are left with the problem of transporting the rest of the measure. Consider a Kantorovich potential u, i.e. a $1-$Lipschitz function u such that for any optimal transport plan $\tilde{\gamma}$ one has

$$u(y) - u(x) \quad = \quad \|y - x\|_{\mathbb{R}^n} \ \forall (x, y) \in \mathrm{Supp}(\tilde{\gamma}).$$

.

We define a transport ray as a maximal open segment $]x, y[\subset X$ such that $u(y) - u(x)$ (this notion is well-defined since u is 1-Lipschitz). By our initial assmuption on γ we have that μ_1-almost every $x \in X$ is contained in a unique transport ray, whose direction is given by the gradient of the potential u (which is well-defined on the transport ray).

The idea is to decompose X into transport rays, along which the measure is going to be transported and to reduce on each of these lines to the one-dimensional case. In order to do so, we need to desintegrate the optimal transport plan γ into a set of conditional measures $(\gamma_C)_{C\in\mathcal{C}}$ obtained in the following way. To each $(x, y) \in \operatorname{Supp}(\gamma)$ one associates the closure of the transport ray containing the segment $[x, y]$; this defines a map ψ. Then by general measure-theoretic results, one can factorize the measure γ as the product $\gamma_C \otimes (\psi_\#(\gamma))(C)$ where each γ_C is a probability measure on X supported on C. It is not very difficult to check that the measures γ_C are actually optimal transport plans between the measures $\pi_{1\#}(\gamma_C)$ and $\pi_{2\#}(\gamma_C)$ and to see that the cost of the optimal transport plan γ factorizes as the integral against $\psi_\#(\gamma)$ of the costs of the optimal transport plans γ_C.

Using results due to Ambrosio (see Theorem 7.1 in [1]), one can show that for $\psi_\#(\gamma)$-almost every C the desintegrated measure γ_C is actually absolutely continuous with respect to the one-dimensional Hausdorff measure restricted C (this follows from the fact that the segments C are gradient lines of a Lipschitz function). We then find for each C in the support of $\psi_\#(\gamma)$ an optimal transport map φ_C between $\pi_{1\#}(\gamma_C)$ and $\pi_{2\#}(\gamma_C)$ by monotone reparametrization (which exists in one dimension under the assumption that $\pi_{1\#}(\gamma_C)$ has no atom, this latter fact following from the absolute continuity of γ_C with respect to the one-dimensional Hausdorff measure restricted to C). Gluing all the transport maps φ_C we eventually obtain an optimal transport map.

References

[1] L. Ambrosio, *Lecture notes on optimal transport problems*, Topology **32** Notes from a CIME summer school in Madeira, 2-9 July 2000. To appear in proceedings of the CIME Course held in Marina Franca.

[2] V. N. Sudakov, *Geometric problems of the theory of infinite-dimensional probability distributions*, Trudy Mat. Inst. Steklov 141 (1976).

The Wasserstein space as a metric space, Benamou and Brenier's interpretation of the Wasserstein distance, and Arnold's geometry of the diffeomorphism group

Robert Philipowski

1. The Wasserstein space as a metric space

If the cost function in the Monge-Kantorovich problem is the p-th power of the distance function of a metric space X, one can use the optimal transportation cost between two probability measures to construct a distance on $\mathcal{P}(X)$. To be precise, let X be a complete and separable metric space, and $p \in [1, \infty)$. Then the *Wasserstein distance* of order p between two probability measures μ and ν on X is defined as

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d(x,y)^p \pi(dx, dy) \right)^{1/p},$$

where as usual the infimum is over all couplings π of μ and ν.

W_p has all properties of a distance, except that it can be infinite. One therefore defines the *Wasserstein space* of order p over X as the set $\mathcal{P}_p(X)$ of those probability measures μ on X for which $\int_X d(x_0, x)^p \mu(dx) < \infty$ for some (and then all) $x_0 \in X$. Then $(\mathcal{P}_p(X), W_p)$ is indeed a metric space (Theorem 7.3 in [7]).

Convergence with respect to W_p is equivalent to weak convergence together with convergence of p-th moments:

Theorem 1.1 (Theorem 7.12 in [7], Theorem 6.9 in [8])**.** *Let $(\mu_k)_{k \in \mathbb{N}}$ be a sequence in $\mathcal{P}_p(X)$, and $\mu \in \mathcal{P}_p(X)$. Then the following are equivalent:*

(1) *$W_p(\mu_k, \mu) \to 0$ as $k \to \infty$.*
(2) *$\mu_k \to \mu$ weakly, and $\int_X d(x_0, x)^p \mu_k(dx) \to \int_X d(x_0, x)^p \mu(dx)$ for some (and then all) $x_0 \in X$.*

The Wasserstein space inherits nice topological properties from the underlying space:

Theorem 1.2 (Theorem 6.18 in [8])**.** *If X is complete and separable, then so is $(\mathcal{P}_p(X), W_p)$.*

2. The Benamou-Brenier formula

Benamou and Brenier [4] observed that on $\mathbb{R}^n$ the square of the quadratic Wasserstein distance can be interpreted as the least action needed to transport one mass distribution to another one:

Theorem 2.1 (Proposition 1.1 in [4], Theorem 8.1 in [7])**.** *Let ρ_0 and ρ_1 be compactly supported probability densities on $\mathbb{R}^n$, and let $C(\rho_0, \rho_1)$ be the set of all pairs (ρ, v) consisting of a continuous curve $\rho = (\rho(t, \cdot))_{0 \le t \le 1}$ in the space* *of probability densities and a family* $v = (v(t,\cdot))_{0\le t\le 1}$ *of vector fields satisfying* $\int_0^1 \int_{\mathbb{R}^n} |v(t,x)|^2 \rho(t,x)dx < \infty$ *such that*

$$\frac{\partial \rho}{\partial t} + div(\rho v) = 0$$

and

$$\rho(0,\cdot) = \rho_0,\ \rho(1,\cdot) = \rho_1.$$

Then

$$W_2(\rho_0,\rho_1)^2 = \inf_{(\rho,v)\in C(\rho_0,\rho_1)} \int_0^1 \int_{\mathbb{R}^n} |v(t,x)|^2 \rho(t,x)dxdt.$$

A similar result holds on Riemannian manifolds M, see Remark 8.3 in [7] or Proposition 4.3 in [6].

Otto [5] (see also Chapter 8 in [7] and Chapter 15 in [8]) used the Benamou-Brenier formula to equip $\mathcal{P}_2(M)$ with the structure of a Riemannian manifold.

3. The geometry of the group of volume-preserving diffeomorphisms

Arnold [1] (see also [3] and Chapter 3.2 in [7]) observed that the Euler equation for an incompressible inviscid fluid (without external forces) in a domain Ω can be interpreted as geodesic motion on the group of volume-preserving diffeomorphisms of Ω. From a physical point of view this is very natural. Indeed, since the fluid is inviscid and incompressible, we are dealing with a mechanical system without friction subject to a constraint, so that in the absence of external forces the least action principle implies that the dynamics of the fluid is geodesic motion on an appropriate Riemannian manifold M (see e.g. Chapter 4 of [2]). Since the fluid is incompressible, the natural choice for M is the infinite-dimensional "manifold" $S\mathrm{Diff}(\Omega)$ of volume-preserving diffeomorphisms of Ω. From differential geometry it is well-known that the stability of geodesic flow on a Riemannian manifold is crucially influenced by its curvature: if the curvature is positive, geodesics tend to converge; if it is negative they tend to diverge. It is therefore natural to compute the curvature of $S\mathrm{Diff}(\Omega)$. In the case of the two-dimensional torus, $\Omega = T^2$, Arnold [1] found out that in "many" directions the curvature is negative (see Lemma 11 in [1] or Theorem 3.4 in [3] for the precise statement). This fact can be seen as a mathematical explanation for the unreliability of long-term weather forecasts, see Chapter IV.4.B in [3].

Arnold's result was later generalized to other manifolds, see Chapter IV.4.A in [3].

References

[1] V. I. Arnold, *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits*, Ann. Inst. Fourier, Grenoble **16** (1966), 319–361.

[2] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, Berlin, Heidelberg, 1978.

[3] V. I. Arnold and B. A. Khesin, *Topological Methods in Hydrodynamics*, Springer-Verlag, New York, Berlin, Heidelberg, 1998.

[4] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), 375–393.
[5] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), 101–174.
[6] F. Otto and M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal. **37** (2006), 1227–1255.
[7] C. Villani, *Topics in Optimal Transportation*, American Mathematical Society, Providence, RI, 2003.
[8] C. Villani, *Optimal Transport, Old and New*, Springer-Verlag, Berlin, Heidelberg, 2009.

Formal Riemannian structure for space of probability measures, its sectional curvature

Shiping Liu

In the previous talk, we saw that optimal transportation could be used to introduce a metric on the space of probability measures. In this talk we will show that this metric can be considered as induced from a formal Riemannian structure on that space. We also introduce an interesting geometrical property of the space itself.

1. Formal Riemannian Structure

We explain the formal Riemannian structure on

$$\mathcal{M} = \{\text{positive functions } \rho \text{ on } R^N \text{ with } \int \rho = 1\}$$

introduced by Otto [3]. Thinking of the tangent space as

$$T_\rho\mathcal{M} = \{\text{functions } s \text{ on } R^N \text{ with } \int s = 0\},$$

we impose the metric

$$g_\rho(s_1, s_2) = \int \rho \nabla p_1 \cdot \nabla p_2,$$

where $s_i \in T_\rho\mathcal{M}$, p_i satisfies the elliptic equation $s_i = -div(\rho\nabla p_i)$.

The idea to explore the geometrical properties of $\mathcal{M}$ is to see it as a "submanifold" of

$$\mathcal{M}^* = \{\text{diffeomorphisms } \Phi \text{ of } R^N\},$$

to which we impose the flat metric. Explicitly, we prove that for fixed $\rho_0 \in \mathcal{M}$ the map

$$\prod : \mathcal{M}^* \to \mathcal{M}, \ \ \Phi \mapsto \rho = \Phi_* \rho_0,$$

is an isometric submersion. Based on this observation, the distance function d induced by g should be the Wasserstein distance, that is,

$$d^2(\rho_0, \rho) = \inf_{\Phi_*\rho_0 = \rho} \int \mid x - \Phi(x) \mid \rho_0 dx, \ \ \forall \rho \in \mathcal{M}.$$

For the rigorous result about induced distance see [4], which is based on another heuristic argument in [5].

2. Sectional Curvature

O'Neill's formula describes a simple relation between the sectional curvatures of two manifolds linked by an isometric submersion. Therefore we can see formally that $\mathcal{M}$ should have nonnegative sectional curvature. This can be made rigorous in metric setting (see section 7.2 and 7.3 in [1]).

Theorem : Wasserstein space $\mathcal{P}_2(R^N)$ has nonnegative sectional curvature.

The definition of "nonnegative sectional curvature" in metric setting is to suppose some degree of concavity for distance functions.

The geodesics in $\mathcal{P}_2(R^N)$ connecting μ_1, μ_2 is just the displacement interpolation of an optimal transportation between μ_1 and μ_2.

With the above properties in hand, we can reduce the proof of the theorem eventually to the proof of the following lemma:

Gluing Lemma : Let $X_i,\ i = 1, 2$ be Polish spaces and $\mu_i \in P(X_i)$. Assume $\mu_{12} \in P(X_1 \times X_2)$ with marginals μ_1, μ_2, and $\mu_{23} \in P(X_2 \times X_3)$ with marginals μ_2, μ_3, then there exists a $\mu \in P(X_1 \times X_2 \times X_3)$ with marginals μ_{12} on $X_1 \times X_2$ and μ_{23} on $X_2 \times X_3$.

This is an easy corollary of the disintegration theorem of measures (see section 5.3 in [1]).

Lott [2] extends the formal observation on sectional curvature to space of probability measures on Riemannian manifolds. Sturm [6] proves more general rigorous results about the sectional curvature.

References

[1] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lecture in Mathematics ETH Zürich, Birkhäuser Verlag, Basel (2005).

[2] J. Lott, *Some geometric calculations on Wasserstein space*, Comm. Math. Phys. **277** (2008), 423–437.

[3] F. Otto, *The geometry of dissipative evolution equation: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), 101–174.

[4] F. Otto, M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal. **37** (2006), 1227–1255.

[5] F. Otto, C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal **173** no.2 (2000), 361–400.

[6] K.T. Sturm, *On the geometry of metric measure spaces I*, Acta Math. **196** no.1 (2006), 65–131.

McCann's displacement convexity

Tomonari Sei

The entropy functional on a Riemannian manifold is deeply related to the Riemannian structure of the underlying manifold. Here we briefly describe this relation in terms of McCann's displacement convexity. We begin with the Euclidean case to clarify the idea.

Let $\mathcal{P}_{\mathrm{ac}}(\mathbf{R}^n)$ be the set of all absolutely continuous probability measures on $\mathbf{R}^n$. Let ρ and ρ' be two measures in $\mathcal{P}_{\mathrm{ac}}(\mathbf{R}^n)$. Then from Brenier's result [1] there exists a unique monotone gradient map $\nabla\psi : \mathbf{R}^n \to \mathbf{R}^n$ such that $\rho' = (\nabla\psi)_\sharp\rho$, where for any measurable map T, $T_\sharp\rho$ denotes the push-forward measure: $(T_\sharp\rho)(A) = \rho(T^{-1}(A))$ for any measurable sets A. Obviously, we have $\rho = \mathrm{Id}_\sharp\rho$, where Id denotes the identity map. Now we consider a one-parameter set of maps $F_t = (1-t)\mathrm{Id} + t\nabla\psi$ for $t \in [0,1]$ and their push-forward measures

$$\rho_t = (F_t)_\sharp\rho = \{(1-t)\mathrm{Id} + t\nabla\psi\}_\sharp\rho.$$

One can show that the measure ρ_t is absolutely continuous by using Lipschitz continuity of $(F_t)^{-1}$ for $t \in [0,1)$. We call the set $(\rho_t)_{t\in[0,1]}$ *the displacement interpolation* of ρ and ρ', following McCann [3]. If ρ and ρ' has finite second moment, the displacement interpolation is considered as the 'geodesic' connecting ρ and ρ' with respect to the Wasserstein metric $W(\rho,\rho') = \min\{[\int |x-x'|^2\Gamma(dx,dx')]^{1/2} \mid \Gamma(\cdot,\mathbf{R}^n) = \rho, \Gamma(\mathbf{R}^n,\cdot) = \rho'\}$.

A function E from $\mathcal{P}_{\mathrm{ac}}(\mathbf{R}^n)$ to $\mathbf{R}$ is called *displacement convex* if for any $t \in [0,1]$

$$E(\rho_t) \le (1-t)E(\rho_0) + tE(\rho_1) \tag{1}$$

holds (see [3]). For example, the entropy functional $E(\rho) = \int \rho\log\rho\, dx$ is displacement convex for any dimension n. In fact, from the change-of-variables formula due to [3], we have

$$E[\rho_t] = \int \rho(x)\log\left(\frac{\rho(x)}{\det((1-t)I + t\nabla^2\psi(x))}\right)dx,$$

where I and $\nabla^2\psi$ denote the identity matrix and the Alexandrov Hessian, respectively. Since $\log\det((1-t)X + tY)$ is concave for any positive-definite matrices X and Y, we obtain the convexity of E. More generally, a functional $E[\rho] = \int A(\rho(x))dx$ is displacement convex if $A(0) = 0$ and the function $\lambda^n A(\lambda^{-n})$ of λ is convex and non-increasing. The proof is based on the geometric-arithmetic mean inequality: $\det^{1/n}((1-t)X + tY) \ge (1-t)\det^{1/n} X + t\det^{1/n} Y$.

We go on to the Riemannian case. Let (M,g) be an n-dimensional complete connected Riemannian manifold with the metric tensor g. The set of probability measures absolutely continuous to the volume measure $d\mathrm{Vol}$ is denoted by $\mathcal{P}_{\mathrm{ac}}(M)$. For any two points x,y in M, the geodesic distance between them is $d(x,y)$. The exponential map determined by a point $x \in M$ and a tangent vector $v \in T_xM$ is $\exp_x(v)$. In order to define the displacement convexity on M, we first describe McCann's theorem [4] on existence and uniqueness of optimal transport. The cost function is $c(x,y) = d(x,y)^2/2$. We call a function $\phi : M \to \mathbf{R}$ c-concave if there exists a function ξ on M such that $\phi(x) = \inf_{y\in M}(c(x,y) - \xi(y))$. Then McCann showed that for any $\rho,\rho' \in \mathcal{P}_{\mathrm{ac}}(M)$ there exists a c-concave function ϕ up to an arbitrary constant such that the map $x \mapsto \exp_x(-\nabla\phi(x))$ pushes ρ forward to ρ'. Now the displacement interpolation and displacement convexity for Riemannian manifolds are defined in the same manner as the Euclidean case. Namely, the displacement interpolation between ρ and ρ' is defined by $\rho_t = (F_t)_\sharp\mu$ with $F_t(x) = \exp_x(-t\nabla\phi(x))$. Remark that $\rho_0 = \rho$ and $\rho_1 = \rho'$. The displacement convexity of a functional E on $\mathcal{P}_{\mathrm{ac}}(M)$ is defined by (1).

Otto and Villani [5] observed that, by using formal calculus, the entropy functional $E[\rho] = \int_M \rho \log \rho \, d\mathrm{Vol}$ has its second-derivative

$$\frac{d^2 E[\rho_t]}{dt^2} = \int_M \left\{ \mathrm{tr}[(\nabla^2 \phi)(\nabla^2 \phi)] + \mathrm{Ric}(\nabla \phi, \nabla \phi) \right\} \rho_t \, d\mathrm{Vol}.$$

Hence E is displacement convex if M has non-negative Ricci curvature. Cordero-Erausquin et al. [2] refined this result as follows. Assume that M has non-negative Ricci curvature. Then, if a function $A : \mathbf{R} \to \mathbf{R}$ satisfies the conditions $A(0) = 0$ and $\lambda^n A(\lambda^{-n})$ convex non-increasing, then $E[\rho] = \int_M A(\rho) d\mathrm{Vol}$ is displacement convex. The proof is based on the Riemannian version of the geometric-arithmetic mean inequality

$$J_t^{1/n}(x) \geq (1-t)[v_{1-t}(F_1(x), x)]^{1/n} + t[v_t(x, F_1(x))]^{1/n} J_1^{1/n}(x),$$

where $J_t(x)$ is the Jacobian determinant of the map F_t, and $v_t(x,y)$ is the volume distortion coefficient (see [2] for the definition). The volume distortion coefficient is greater than or equal to 1 if M has the non-negative Ricci curvature.

Lastly, von Renesse and Sturm [6] showed that the Ricci curvature of M is non-negative (resp. bounded below from K) if and only if the entropy functional is displacement convex (resp. displacement K-convex) on $\mathcal{P}^2(M)$, where $\mathcal{P}^2(M)$ is the set of probability measures with finite second moment. Therefore the lower bound of the Ricci curvature is characterized by displacement convexity of the entropy functional.

References

[1] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), 375–417.

[2] D. Cordero-Erausquin, R. J. McCann and M. Schumuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Invent. Math. **146** (2001), 219–257.

[3] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math. textbf128 (1997), 153–179.

[4] R. J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. **11** (2001), 589–608.

[5] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361–400.

[6] M. K. von Renesse and K. T. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Comm. Pure. Appl. Math. **58** (2005), 923–940.

Ricci bounds for metric measure spaces

Ulrich Bunke

On the class of Riemannian manifolds we have the concept $\mathtt{curv} \geq K$ of sectional curvature bounded below by K. Riemannian manifolds can be embedded into the larger class of metric spaces. The concept of a lower curvature bound in the sense of Alexandrov extends the lower sectional curvature bound. The lower curvature bound is stable in the sense that if a sequence of metric spaces (M_i, d_i)

with $\mathtt{curv}(M_i, d_i) \geq K$ converges in the Gromov-Hausdorff sense to (M, d), then $\mathtt{curv}(M, d) \geq K$.

For the class of Riemannian manifolds we have a similar concept of lower Ricci curvature bound $\mathtt{Ricci} \geq K$. It turned out that an extension to metric spaces should take the volume measure into account. In this talk we report on the results of [1] and [2] on the extension of the concept of lower Ricci curvature bounds to the class of metric measure spaces and its stabilty.

We consider separable, complete metric spaces and locally finite Borel measures. A metric measure space is a triple (M, d, μ), where (M, d) is a metric space and μ is a measure. A Riemannian manifold (M, g) gives rise to the metric measure space (M, d_g, μ_g), where d_g is the Riemannian distance and μ_g is the Riemannian volume measure.

To a metric space (M, d) we can associate the metric space $P_2(M, W_2)$ of all probability measures with the Wasserstein distance W_2 which is given by

$$W_2(\mu, \nu)^2 := \inf_{\pi} \int_{M \times M} d^2 \, d\pi \ ,$$

where the infimum is taken over all measures π on $M \times M$ with $\mathtt{pr}_{1*}\pi = \mu$ and $\mathtt{pr}_{2*}\pi = \nu$.

By $P_2(M, d, \mu) \subseteq P_2(M, d)$ we denote the subset of absolute continuous measures. We define the entropy

$$\mathtt{Ent}(\dots|\mu) : P_2(M, d) \to [-\infty, \infty]$$

by

$$\mathtt{Ent}(\nu|\mu) := \begin{cases} \lim_{\epsilon \downarrow 0} \int_{\rho > \epsilon} \rho \log \rho \, d\mu & \nu \in P(M, d, \mu) \ , \ \ \nu = \rho\mu \\ \infty & \text{else} \end{cases}$$

We let $P^*(M, d, \mu) := \{\mathtt{Ent} < \infty\} \subseteq P_2(M, d, \mu)$. For a geodesic $c : [0, 1] \to P^*(M, d, \mu)$ we consider the condition (C_K):

$$\mathtt{Ent}(c(t)|\mu) \leq (1-t)\mathtt{Ent}(c(0)|\mu) + t\mathtt{Ent}(c(1)) - \frac{Kt(1-t)}{2} W_2(c(0), c(1)) \ , \forall t \in [0, 1] \ .$$

Definition 0.1. *A metric measure space (M, d, μ) has has Ricci curvature bounded below by K (we write $\mathtt{Ricci} \geq K$) if every pair $\nu_0, \nu_1 \in P^*(M, d, \mu)$ is connected by a geodesic which satisfies (C_K).*

This is indeed an extension of the Riemannian concept in view of the following theorem [1, Thm. 7.3], [2, Thm 4.9].

Theorem 0.2. *Let (M, g) be a complete Riemannian manifold. Then it has Ricci curvature bounded below by K if and only if the associated metric measure space (M, d_g, μ_g) has Ricci curvature bounded below by K.*

An ϵ-Gromov-Hausdorff approximation (ϵ-GHA) between two metric spaces (M, d) and (M', d') is a map $\phi : M \to M'$ such that $M' \subseteq B_\epsilon(\phi(M))$ and

$$|d(x, y) - d'(\phi(x), \phi(y))| \leq \epsilon \ , \quad \forall x, y \in M \ .$$

A sequence (M_i, d_i) converges in the Gromov-Hausdorff (GH) sense to (M, d) if there exists a sequence (ϵ_i) with $\epsilon_i \downarrow 0$ and ϵ_i-GHA's $\phi_i : M_i \to M$. Furthermore, a sequence of metric measure spaces (M_i, d_i, μ_i) converges to (M, d, μ) in the measured GH-sense, if there exists a sequence (ϵ_i) with $\epsilon_i \downarrow 0$ and ϵ_i-GHA's $\phi_i : M_i \to M$ such that $\phi_{i,*}\mu_i \stackrel{w}{\to} \mu$.

For example, the association $(M, d) \mapsto P_2(M, d)$ is GH-continuous. More precisely [1, Prop. 4.1]

Proposition 0.1. *If $\phi : (M, d) \to (M', d')$ is an ϵ-GHA, then $\phi_* : P_2(M, d) \to P_2(M', d')$ is an $\tilde{\epsilon}$-GHA, where $\tilde{\epsilon} = \tilde{\epsilon}(\epsilon, \mathtt{diam(M')})$ tends to zero as $\epsilon \downarrow 0$.*

The main theorem of this talk is [1, Thm. 4.15], [2, Thm 4.20] (the theorems in both references are more general)

Theorem 0.3. *Let (M_i, d_i, μ_i) be a sequence of compact metric measure spaces which converges in the measured GH-sense to (M, d, μ). Then $\mathtt{Ricci}(M_i, d_i, \mu_i) \geq K$ for all i implies $\mathtt{Ricci}(M, d, \mu) \geq K$.*

Let us sketch the main steps of the proof. We consider $\mu_0, \mu_1 \in P_2^*(M, d, \mu)$. Then we must find a geodesic from μ_0 to μ_1 which satisfies (C_K). By an approximation argument we can assume that $\mu_i = \rho_i \mu$ with $\rho_i \in C(M)$. We consider the sequence $(\phi_n : M_n \to M)$ of ϵ_n-GHA's. We define for $n >> 0$, $i = 0, 1$, the measure

$$\mu_{n,i} := \frac{\phi_n^* \rho_i}{\int_M \rho_i d\phi_{n*}(\mu_n)}$$

on M_n We then choose geodesics c_n in $P_2(M_n, d_n)$ from $\mu_{n,0}$ to $\mu_{n,1}$ which satisfy (C_K). After taking a subsequence $(\phi_{n,*}c_n)$ converges uniformly to a curve c in $P_2(M, d, m)$. We check that c is a geodesic from μ_0 to μ_1. We then show that c satisfies C_K. The main ingredients of this step (and also of the approximation argument above) are Proposition 0.1, the contraction property [1, Thm. B33]

$$\mathtt{Ent}(f_*\mu | f_*\mu) \leq \mathtt{Ent}(\nu|\mu) \ ,$$

and the lower semicontinuity

$$\mathtt{Ent}(\nu|\mu) \leq \liminf \mathtt{Ent}(\nu_n|\mu_n) \ , \ \text{if } (\nu_n, \mu_n) \stackrel{w}{\to} (\nu, \mu)$$

of the entropy.

References

[1] J. Lott, C. Villani, *Ricci curvature for metric-measure spaces via optimal transport,* Arxiv:math.DG/0412127.

[2] K. T. Sturm *On the geometry of metric measure spaces I*, Acta Math. **196** (2006), 65–131.

The curvature-dimension condition $\mathsf{CD}(K, N)$

Shin-ichi Ohta

We consider the *curvature-dimension condition* $\mathsf{CD}(K, N)$ introduced by Sturm, Lott and Villani. A metric measure space (X, d, m) is said to satisfy $\mathsf{CD}(0, N)$ for $N \in (1, \infty)$ if the *Rényi entropy* $S_N(\rho m) = N - N \int_X \rho^{1-1/N}\, dm$ is convex in the Wasserstein space $(\mathcal{P}_2(X), d_2^W)$. Passing to the limit, $\mathsf{CD}(0, \infty)$ is defined as the convexity of the *relative entropy* $S_\infty(\rho m) = \int_X \rho \log \rho\, dm$. General $\mathsf{CD}(K, N)$ for $K \neq 0$ is defined through a similar, but more complicated convexity condition involving sin (if $K > 0$) or sinh (if $K < 0$).

It is not difficult to see that the Euclidean space $(\mathbb{R}^n, \mathsf{L})$ with the Lebesgue measure L satisfies $\mathsf{CD}(0, n)$. Key ingredients are Brenier's convex function f that induces optimal transport $\Phi_t = (1-t)\mathrm{Id}_{\mathbb{R}^n} + t\nabla f$ between $\rho_0\mathsf{L}$ and $\rho_1\mathsf{L}$, the Monge-Ampère equation $|D\Phi_t| = \rho_0/(\rho_t \circ \Phi_t)$ (with $\rho_t\mathsf{L} = (\Phi_t)_\sharp(\rho_0\mathsf{L})$), and the inequality of arithmetic and geometric means

$$|D\Phi_t|^{1/n} \geq (1-t) + t|D\Phi_1|^{1/n}. \tag{1}$$

Here we denote by $|D\Phi|$ the Jacobian of a linear operator Φ. Similar, but more careful calculation shows that a weighted Euclidean space $(\mathbb{R}^n, e^{-\psi}\mathsf{L})$ satisfies

$$\|D\Phi_t\|^{1/N} \geq (1-t) + t\|D\Phi_1\|^{1/N} \tag{2}$$

and $\mathsf{CD}(0, N)$ if (and only if) ψ satisfies

$$\operatorname{Hess}\psi(v, v) - \frac{\langle \nabla\psi, v\rangle^2}{N-n} \geq 0,$$

where $\|D\Phi\|(x) := e^{\psi(x)-\psi(\Phi(x))}|D\Phi|(x)$ is the Jacobian taking the weight into account. In the infinite dimensional case $N = \infty$, there appear log concave measures (e.g., Gaussian spaces $(\mathbb{R}^n, e^{-|x|^2/2}\mathsf{L}(dx))$).

In order to obtain concavity similar to (1) or (2) in a weighted Riemannian manifold $(M, g, e^{-\psi}m_g)$ (m_g is the volume measure), besides $\operatorname{Hess}\psi(v, v) - \langle \nabla\psi, v\rangle^2/(N-n)$, we naturally need the Ricci curvature. This is because the optimal transport is performed along geodesics, say $\gamma : [0, 1] \longrightarrow M$. Then its variational vector fields are Jacobi fields J solving the Jacobi equation $D_{\dot\gamma}^2 J + R(J, \dot\gamma)\dot\gamma = 0$. Now the concavity of the density $\det(\langle J_i, J_j\rangle)$ (choosing orthonormal $\{J_i(0)\}_{i=1}^n$ with $J_1(0) = \dot\gamma(0)/|\dot\gamma(0)|$) is controlled by $\mathrm{Trace}(\langle R(\cdot, \dot\gamma)\dot\gamma, \cdot\rangle) = \mathrm{Ric}(\dot\gamma)$. It might be also helpful to recall the Bishop-Gromov volume comparison which asserts that the n-th root of the volume of concentric balls $m_g(B(x, r))^{1/n}$ is concave in $r \geq 0$ if $\mathrm{Ric} \geq 0$.

It indeed has been established by von Renesse, Sturm, Lott and Villani ([vRS], [St1], [St2], [LV1], [LV2]) that $\mathsf{CD}(K, N)$ is equivalent to the lower weighted Ricci curvature bound

$$\mathrm{Ric}(v) + \operatorname{Hess}\psi(v, v) - \frac{\langle \nabla\psi, v\rangle^2}{N-n} \geq K.$$

Thus the curvature-dimension condition gives a successful generalization of the lower Ricci curvature bound (along with the upper dimension bound) to general metric measure spaces without differentiable structure. $\mathsf{CD}(K,N)$ is stable under measured Gromov-Hausdorff convergence, and has a number of applications such as Brunn-Minkowski inequality (which was new even in the Riemannian setting), logarithmic Sobolev inequality and Lichnerowicz inequality.

Note that this kind of synthetic characterization had been known for lower sectional curvature bounds. Such spaces are called Alexandrov spaces, and deeply investigated in the last two decades. However, Ricci curvature bounds are in a sense a more natural condition, not only for its analytic implications, but also because such spaces form a precompact family with respect to the Gromov-Hausdorff convergence. Spaces appearing in the limit of the convergence had been studied in Cheeger and Colding's celebrated work [CC].

The equivalence between the curvature-dimension condition and lower Ricci curvature bounds is extended to general Finsler manifolds by introducing an appropriate notion of weighted Ricci curvature ([Oh]). For instance, $(\mathbb{R}^n, \|\cdot\|, e^{-\psi}\mathsf{L})$ with convex weight ψ satisfies $\mathsf{CD}(0,\infty)$ for any Minkowski norm $\|\cdot\|$. On the one hand, this result leads to a slightly disappointing fact (for Riemannian geometers) that the curvature-dimension condition can not characterize Riemannian spaces. On the other, the curvature-dimension condition turns out extremely useful also in Finsler geometry.

References

[CC] J. Cheeger and T. H. Colding, *On the structure of spaces with Ricci curvature bounded below. I, II, III*, J. Differential Geom. **46** (1997), 406–480; ibid. **54** (2000), 13–35; ibid. **54** (2000), 37–74.

[LV1] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. **169** (2009), 903–991.

[LV2] J. Lott and C. Villani, *Weak curvature conditions and functional inequalities*, J. Funct. Anal. **245** (2007), 311–333.

[Oh] S. Ohta, *Finsler interpolation inequalities*, to appear in Calc. Var. Partial Differential Equations.

[vRS] M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), 1–18.

[St1] K.-T. Sturm, *On the geometry of metric measure spaces*, Acta Math. **196** (2006), 65–131.

[St2] K.-T. Sturm, *On the geometry of metric measure spaces.* II, Acta Math. **196** (2006), 133–177.

Diffusions are gradient flows of entropy w. r. t. Wasserstein metric

Nicola Gigli

Mainly thank to the works [1] and [2], it has become clear the link between certain diffusion equations and the geometry of optimal transportation. The link being that such equations may be seen as gradient flow of appropriate energies w.r.t. the Wasserstein distance. The aim of this talk is to investigate this topic by: presenting the so called 'Otto calculus', and giving an overview on the techinques which allow a rigorous study of gradient flows in the Wasserstein space.

By Otto calculus, we mean the formal description of the Wasserstein space as Riemannian manifold, and the subsequent characterization of gradient flows. A way to see the Riemannian structure of $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is to verify that an absolutely continuous curve $t \mapsto \rho_t \in \mathcal{P}_2(\mathbb{R}^d)$ is a distributional solution of the continuity equation

$$\frac{d}{dt}\rho_t + \nabla \cdot (\nabla\varphi_t \rho_t) = 0, \tag{1}$$

where the vector fields $\nabla\varphi_t$ are uniquely identified by the above equation. It turns out that the metric length of the curve $\mathcal{L}(\rho_t)$ can be recovered by the formula

$$\mathcal{L}(\rho_t) = \int_0^1 \sqrt{\int |\nabla\varphi_t|^2 d\rho_t} dt.$$

Therefore it is natural to think that the tangent space at a measure ρ is the 'space of gradients endowed with the scalar product w.r.t. ρ itself'. In this setting, to perturb a measure ρ along the direction $\nabla\varphi$ consists in defining

$$\rho_t := \big(Id + t\nabla\varphi\big)_{\#}\rho, \tag{2}$$

which satisfies

$$\frac{d}{dt}\rho_t|_{t=0} + \nabla \cdot (\nabla\varphi\rho) = 0.$$

With this interpretation, it is possible to compute the gradient of a functional: we run explicit calculation for the case of the energy $E : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ defined by

$$E(\rho) := \frac{1}{m-1}\int \rho^m.$$

It is just a matter of algebraic manipulations to see that with ρ_t defined as in (2) it holds

$$\frac{d}{dt}E(\rho_t)|_{t=0} = -\int \rho^m \Delta\varphi = \int \left\langle \nabla\left(\frac{m}{m-1}\rho^{m-1}\right), \nabla\varphi \right\rangle \rho,$$

which indentifies the gradient $\nabla^W E$ of the energy E as

$$\nabla^W E = \nabla\Big(\frac{m}{m-1}\rho^{m-1}\Big)$$

Therefore, taking into account (1) the gradient flow equation of E is

$$\frac{d}{dt}\rho_t = \Delta(\rho_t^m),$$

which is the porous medium equation.

In the second part of the talk, we present the key ingredients needed to apply the theory of minimizing movements to functionals like the one described. The main idea here is that in order to replicate the well known results concerning gradient flows of convex functionals in Hilbert spaces (like, e.g., exponential convergence to equilibrium), what is needed is some sort of convexity property of both the functional and the squared distance. A complication which arises in the Wasserstein setting is that, being this space positively curved, the squared distance calculated along geodesics exhibits concavity properties, rather than convexity ones. In order to tackle this problem, in [3] it was proposed to work under the following Assumption. *Consider a metric space* (X, d) *and a lower semicontinuous functional* $E : X \to \mathbb{R} \cup \{+\infty\}$ *such that the following is true: for any* $x, y_0, y_1 \in X$ *there exists a curve* $\gamma : [0,1] \to X$ *(not necessarily a geodesic) such that* $\gamma(0) = y_0$*,* $\gamma(1) = y_1$ *and:*

i) $t \mapsto E\big(\gamma(t)\big)$ *is convex,*

ii) $t \mapsto \frac{1}{2}d^2\big(x, \gamma(t)\big)$ *is* 1*-convex.*

If this Assumption is satisfied, it is possible to show that:

(A) For any $x \in X$ and any $\tau > 0$ there exists one and only one minimizer x_τ of

$$y \mapsto E(y) + \frac{d^2(x,y)}{2\tau}.$$

(B) For any $x \in \overline{D(E)}$ and any $y \in D(E)$ it holds the discrete variational inequality

$$E(x_\tau) + \frac{1}{2\tau}d^2(x, x_\tau) + \frac{1}{2\tau}d^2(x_\tau, y) - \frac{1}{2\tau}d^2(x, y) \leq E(y),$$

where x_τ is the minimizer given by **(A)**.

The validity of **(A)** and **(B)** are the main bricks on which the theory of minimizing movements (in this setting) is built. Indeed **(A)** tells that there always exists a unique *discrete solution* of the minimizing movement scheme for any choice of the parameter τ, while **(B)** is the property which ensures the convergence of the scheme, by also giving information on the rate of convergence.

For our purpouses, the interest of this Assumption relies on its applicability to the study of functionals on $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, indeed:

- for any given $\rho, \nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ there is a natural choice of a curve $t \mapsto \nu_t$ such that the function $t \mapsto \frac{1}{2}W_2^2(\rho, \nu_t)$ is 1-convex: it is sufficient to define

$$\nu_t := \Big((1-t)T_\rho^{\nu_0} + tT_\rho^{\nu_1}\Big)_{\#}\rho,$$

  where $T_\rho^{\nu_i}$ are the optimal transport maps from ρ to ν_i, $i = 0, 1$,
- all the 'classical' energy functionals arising in this setting (i.e. the potential energy, the internal energy and the interaction energy functionals) are convex not only along geodesics, but also along this more general interpolating curves.

References

[1] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., **29** (1998), pp. 1–17.

[2] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, **26** (2001), pp. 101–174.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in spaces of probability measures*, Birkäuser, 2005.
[4] C. Villani, *Optimal transport, old and new*, Springer Verlag, 2008.

Contraction properties of Wasserstein metric under diffusions

Kazumasa Kuwada

The purpose of this talk is to derive a contraction property of solutions to a (non-linear) diffusion equation from the convexity property of a corresponding entropy functional with the aid of the theory of optimal transports. In the last talk, we have observed that a solution $\rho(t)$ to a diffusion equation is regarded as a gradient flow of an entropy functional $E(\rho)$ on L^2-Wasserstein space $\mathcal{P}_2(X)$. Let $\rho_0(t)$ and $\rho_1(t)$ are gradient flows of E and d_2^W the L^2-Wasserstein distance. As a general principle, our goal is expressed as follows:

If E is K-displacement convex on $\mathcal{P}_2(X)$ for some $K \in \mathbb{R}$,
then $d_2^W(\rho_0(t), \rho_1(t)) \le \mathrm{e}^{-Kt} d_2^W(\rho_0(0), \rho_1(0))$ holds.

Note that the same problem for gradient flows $\xi_0(t)$, $\xi_1(t)$ of a function f on a finite dimensional Riemannian manifold can be solved by a basic differential calculus. It follows from a combination of these two estimates:

(1) A calculation of $\dfrac{d}{dt} d(\xi_0(t), \xi_1(t))$ by the first variational formula.
(2) The Taylor expansion of $f \circ \gamma(s) - f \circ \gamma(1-s)$ up to second order, where $\{\gamma(s)\}_{s \in [0,1]}$ is a minimal geodesic joining $\xi_0(t)$ and $\xi_1(t)$.

When X is a complete Riemannian manifold, we can expect that the same strategy works on $\mathcal{P}_2(X)$ since $\mathcal{P}_2(X)$ possesses a formal Riemannian structure.

In this talk, we concentrate on introducing a result by Otto [1] where the formal argument is applied first rigorously. He shows a quantitative time asymptotic for solutions $\rho(t)$ to the porous medium equation

$$\frac{\partial}{\partial t} \rho(t) = \Delta(\rho(t)^m)$$

on $\mathbb{R}^N$ satisfying $m \ge 1 - N^{-1}$ and $m > N(N+2)^{-1}$. It is known that $\rho(t)$ approaches to a self-similar solution $\rho_*(t)$ as $t \to \infty$ in an appropriate sense. Under a suitable re-scaling using an exponent $\alpha = \{N(m-1)+2\}^{-1}$ of self-similarity, we obtain a new equation. Then the solution $\hat{\rho}(t)$ corresponding to $\rho(t)$ converges to a stationary solution $\hat{\rho}_*$ corresponding to $\rho_*(t)$. In this framework, we show the following estimate on the rate of convergence:

$$d_2^W(\hat{\rho}(t), \hat{\rho}_*) \le \mathrm{e}^{-\alpha t} d_2^W(\hat{\rho}(0), \hat{\rho}_*).$$

Indeed, $\hat{\rho}(t)$ is regarded as a gradient flow of a functional F and F is α-displacement convex. The proof is based on two key estimates corresponding to (1) and (2). After an approximation of the solution ensuring regularity, these two estimates follow by using the theory of optimal transport. For (1), a dynamical expression of gradient flows enables us to realize a variation of Wasserstein distances. For (2), a characterization of geodesics on $\mathcal{P}_2(\mathbb{R}^N)$ together with the change of variable formula plays a significant role. Note that $\rho(t)$ is also regarded as a gradient flow of an entropy functional E. But E is 0-displacement convex and has no information on the rate of convergence.

The contraction property can be obtained from the convexity of an entropy functional even when the underlying space X is a complete Riemannian manifold. We refer to [2] and Chapter 23 of [3] for developments in this direction.

References

[1] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no.1-2, 101–174.

[2] F. Otto, M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal. **37** (2006), 1227–1255

[3] C. Villani, *Optimal transport - Old and new*, Grundlehren, Springer, Berlin (2008)

Sobolev inequality, Talagrand inequality and applications

Jan Maas

Many inequalities with geometric content, such as Sobolev and Gagliardo-Nirenberg inequalities, can be elegantly proved using optimal transport methods [2], often with sharp constants. In this talk we focus on logarithmic Sobolev and Talagrand inequalities.

Logarithmic Sobolev inequalities. Let γ be a Borel probability measure on $\mathbb{R}^n$ having a smooth density $\frac{d\gamma}{dx} = e^{-V(x)}$. The measure γ is said to satisfy a logarithmic Sobolev inequality $LSI(\lambda)$ with parameter $\lambda > 0$ if for any probability measure μ on $\mathbb{R}^n$ having a smooth density ρ with respect to γ, one has

$$H_\gamma(\mu) \leq \frac{1}{\lambda} I_\gamma(\mu).$$

In this inequality,

$$H_\gamma(\mu) := \int_{\mathbb{R}^n} \rho(x) \log \rho(x) \, d\gamma(x)$$

denotes the relative entropy of μ with respect to γ, and

$$I_\gamma(\mu) := \int_{\mathbb{R}^n} |\nabla \log \rho(x)|^2 \, d\mu(x)$$

is the Fisher information.

Since the pioneering work of Gross in the seventies [3], logarithmic Sobolev inequalities play a prominent role in the analysis of Fokker-Planck equations associated with stochastic drift-diffusion equations

$$dX_t = -\nabla V(X_t) \, dt + \sqrt{2} dB_t, \tag{1}$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion on $\mathbb{R}^n$. In particular, $LSI(\lambda)$ implies exponential decay of the entropy along the transition semigroup $(P_t)_{t\geq 0}$ associated with (1). Moreover, if γ satisfies $LSI(\lambda)$, then γ satisfies the Poincaré inequality

$$\int_{\mathbb{R}^n} f^2 \, d\gamma \leq \frac{1}{\lambda} \int_{\mathbb{R}^n} |\nabla f|^2 \, d\gamma$$

for all smooth functions f with $\int_{\mathbb{R}^n} f \, d\gamma = 0$. This inequality expresses a spectral gap for the generator of the semigroup $(P_t)_{t\geq 0}$. Furthermore, $LSI(\lambda)$ is equivalent to the hypercontractivity estimate

$$\|P_t f\|_{L^q(\gamma)} \leq \|f\|_{L^p(\gamma)}$$

for $1 < p < q < \infty$ and $e^{2\lambda t} \geq \frac{q-1}{p-1}$.

Logarithmic Sobolev inequalities are stable under tensorisation: if μ and ν satisfy $LSI(\lambda)$, then the product measure $\mu \otimes \nu$ satisfies $LSI(\lambda)$ as well. This property makes logarithmic Sobolev inequalities very useful in high or infinite dimensional situations.

The most famous criterion for $LSI(\lambda)$ is due to Bakry and Emery [1]. These authors proved (in a more general Riemannian setting) that μ satisfies $LSI(\lambda)$ if

$$D^2 V(x) \geq \lambda$$

in the sense of positive quadratic forms. The proof of Bakry and Emery can be naturally interpreted using Otto's Riemannian structure on the Wasserstein space $P_2(\mathbb{R}^n)$ and the gradient flow formulation of the Fokker-Planck equation [5]. In fact, the proof is based on a comparison of the first and second derivatives of the entropy along the transition semigroup, and these quantities can be computed very efficiently using the gradient flow formulation.

Talagrand inequalities. We say that γ satisfies Talagrand's inequality $T(\lambda)$ if

$$W_2(\mu, \gamma) \leq \sqrt{\frac{H_\gamma(\mu)}{2\lambda}}$$

for every Borel probability measure μ on $\mathbb{R}^n$. In this inequality, W_2 denotes the Wasserstein metric associated with the quadratic cost. This inequality has first been established for the Gaussian measure [6].

Otto and Villani [5] discovered that $LSI(\lambda)$ implies Talagrand's inequality $T(\lambda)$. Their proof is based on the gradient flow formulation of the Fokker-Planck equation. Talagrand's inequality is important in probability theory and geometry, especially in high dimensions, as it implies the following concentration inequality, which expresses the fact that the measure of a set increases rapidly if the set is slightly expanded: there exists $C > 0$ such that for any Borel set $A \subseteq \mathbb{R}^n$ satisfying $\gamma(A) \geq \frac{1}{2}$, one has

$$\gamma(A^r) \geq 1 - e^{-Cr^2},$$

where $A^r := \{x \in \mathbb{R}^n : \mathrm{dist}(x, A) \leq r\}$.

Applications to spin systems. Logarithmic Sobolev inequalities are difficult to establish for spin systems with a non-convex Hamiltonian. We present an efficient criterion due to Otto and Reznikoff [4], which can be applied to spin systems with strong and weak interactions. A crucial ingredient in the proof is a covariance estimate related to Talagrand's inequality.

References

[1] D. Bakry and M. Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math. **1123** (1985), 177–206.

[2] D. Cordero-Erausquin, B. Nazaret, and C. Villani, *A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities*, Adv. Math. **182**, no. 2, (2004), 307–332.

[3] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083.

[4] F. Otto and M. G. Reznikoff, *A new criterion for the logarithmic Sobolev inequality and two applications*, J. Funct. Anal. **242** (2007), no. 1, 121–157.

[5] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), no. 2, 361–400.

[6] M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. **6**, no. 3, (1996), 587–600.

Optimal Transport and Ricci Flow

Oliver C. Schnürer

Let $g(\tau)$ solve reverse Ricci flow $\frac{\partial}{\partial\tau}g(\tau) = 2\mathrm{Ric}\,(g(\tau))$ on a closed manifold M. For a curve $\gamma : [\tau_1, \tau_2] \to M$, $0 < \tau_1 < \tau_2$, the $\mathcal{L}$-length is

$$\mathcal{L}(\gamma) = \int\limits_{\tau_1}^{\tau_2} \sqrt{\tau}\left(R(\gamma(\tau),\tau) + |\gamma'(\tau)|^2_{g(\tau)}\right) d\tau.$$

The $\mathcal{L}$-Wasserstein distance of two probability measures ν_1 and ν_2 is

$$V(\nu_1,\tau_1;\nu_2,\tau_2) := \inf_{\pi\in\Gamma(\nu_1,\nu_2)} \int\limits_{M\times M} \inf_{\gamma=\gamma(x,y)} \mathcal{L}(\gamma)\, d\pi(x,y),$$

where π has marginals ν_i and $\gamma : [\tau_1, \tau_2] \to M$ connects x and y. If ν_i, $i = 1, 2$, are backwards diffusions, $\nu_i = u_i(\tau)\, d\mu(\tau)$, and hence $\frac{\partial u_i}{\partial \tau} = \Delta u_i - R u_i$, P. Topping [1] has shown that

$$s \mapsto 2\left(\sqrt{\tau_2} - \sqrt{\tau_1}\right) e^{s/2} V\left(\nu_1\left(\tau_1 e^s\right), \tau_1 e^s; \nu_2\left(\tau_2 e^s\right), \tau_2 e^s\right) - 2n\left(\sqrt{\tau_2} - \sqrt{\tau_1}\right)^2 e^s$$

is weakly decreasing. This implies in particular that the Wasserstein distance $W_2(\nu_1(\tau), \nu_2(\tau), \tau)$ and Perelman's $\mathcal{W}$-functional are also weakly decreasing in τ.

References

[1] Peter Topping, *L-optimal transport for Ricci flow*, 2007, J. reine angew. Math., to appear.

Reporters: Bernhard Hader and Georg Menz

Participants

Dr. Jerome Bertrand
Institut de Mathematiques
Universite Paul Sabatier
118, Route de Narbonne
F-31062 Toulouse Cedex 4

Prof. Dr. Ulrich Bunke
NWF-I Mathematik
Universität Regensburg
93040 Regensburg

Dr. Codina Cotar
Zentrum Mathematik
TU München
Boltzmannstr. 3
85748 Garching bei München

Prof. Dr. Christopher Deninger
Mathematisches Institut
Universität Münster
Einsteinstr. 62
48149 Münster

Dr. Matthias Erbar
D160
48, bd Jourdan
F-75014 Paris

Marzena Franek
Mathematisches Institut
Universität Münster
Einsteinstr. 62
48149 Münster

Dr. Nicola Gigli
Mathematiques et Informatique
Universite Bordeaux I
351, cours de la Liberation
F-33405 Talence Cedex

Prof. Dr. Helge Glöckner
Institut für Mathematik
Universität Paderborn
Warburger Str. 100
33098 Paderborn

Dorian Anthony Goldman
Ferdinandstrasse 42
53127 Bonn

Bernhard Hader
Mathematisches Institut
Universität Bonn
Endenicher Allee 60
53115 Bonn

Michael Hinz
Mathematisches Institut
Friedrich-Schiller-Universität Jena
07737 Jena

Clement Hongler
Departement de Mathematiques
Universite de Geneve
Case Postale 64
2-4 rue du Lievre
CH-1211 Geneve 4

Xueping Huang
Fakultät für Mathematik
Universität Bielefeld
Universitätsstr. 25
33615 Bielefeld

Martin Huesmann
Universität Bonn
Institut für Angewandte Mathematik
Poppelsdorfer Allee 82/1
53115 Bonn

Dr. Bernd Kirchheim
Mathematisches Institut
Heinrich-Heine-Universität
Gebäude 25.22
Universitätsstraße 1
40225 Düsseldorf

Kazumasa Kuwada
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstr. 6
53115 Bonn

Dr. Krzysztof Latuszynski
Department of Statistics
University of Warwick
GB-Coventry CV4 7AL

Joseph Lehec
Equipe d'Analyse et Math. Appliq.
Universite de Marne-la-Vallee
Cite Descartes, 5 BD Descartes
Champs-sur-Marne
F-77454 Marne-La-Vallee Cedex

Dr. Adrian Lim
Laboratoire de Mathematique
Universite du Luxembourg
162 A, avenue de la Faiencerie
L-1511 Luxembourg

Shiping Liu
Max-Planck-Institut für Mathematik
in den Naturwissenschaften
Inselstr. 22 - 26
04103 Leipzig

Dr. Dejun Luo
Laboratoire de Mathematique
Universite du Luxembourg
162 A, avenue de la Faiencerie
L-1511 Luxembourg

Jan Maas
Delft Institute of Applied Mathematics
Faculty EEMCS
Delft University of Technology
P.O.Box 5031
NL-2600 GA Delft

Georg Menz
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstr. 10
53115 Bonn

Shin-ichi Ohta
Department of Mathematics
Kyoto University
Kitashirakawa, Sakyo-ku
Kyoto 606-8502
JAPAN

Prof. Dr. Felix Otto
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstr. 10
53115 Bonn

Dr. Robert Philipowski
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstr. 6
53115 Bonn

Dr. Oliver C. Schnürer
FB Mathematik & Informatik
Freie Universität Berlin
Arnimallee 3
14195 Berlin

Tomonari Sei
School of Information Science and
Technology
University of Tokyo
Hongo 7-3-1, Bunkyo-ku
Tokyo 113-8656
JAPAN

Christian Seis
Physikalisches Institut
Universität Bonn
Nußallee 12
53115 Bonn

Christian Selinger
Laboratoire de Mathematique
Universite du Luxembourg
162 A, avenue de la Faiencerie
L-1511 Luxembourg

Dr. Miles Simon
Mathematisches Institut
Universität Freiburg
Eckerstr. 1
79104 Freiburg

Dr. Emanuele Spadaro
Institut für Mathematik
Universität Zürich
Winterthurerstr. 190
CH-8057 Zürich

Prof. Dr. Karl-Theodor Sturm
Universität Bonn
Institut für Angewandte Mathematik
Poppelsdorfer Allee 82/1
53115 Bonn

Thomas Viehmann
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstr. 6
53115 Bonn

Boris Walter
Institut für Mathematik
Universität Paderborn
Warburger Str. 100
33098 Paderborn