OPTIMAL TRANSPORT AND CURVATURE

ALESSIO FIGALLI AND CÉDRIC VILLANI

Introduction

These notes record the six lectures for the CIME Summer Course held by the second author in Cetraro during the week of June 23-28, 2008, with minor modifications. Their goal is to describe some recent developements in the theory of optimal transport, and their applications to differential geometry. We will focus on two main themes:

(a) Stability of lower bounds on Ricci curvature under measured Gromov–Hausdorff convergence.
(b) Smoothness of optimal transport in curved geometry.

The main reference for all the material covered by these notes (and much more) is the recent book of the second author [45].

These notes are organized as follows:

• In Section 1 we recall some classical facts of metric and differential geometry; then in Section 2 we study the optimal transport problem on Riemannian manifolds. These sections introduce the basic objects and the notation.

• In Section 3 we reformulate lower bounds on Ricci curvature in terms of the "displacement convexity" of certain functionals, and deduce the stability. Then in Section 4 we address the question of the smoothness of the optimal transport on Riemannian manifold. These two sections, focusing on Problems (a) and (b) respectively, constitute the heart of these notes, and can be read independently of each other.

• Section 5 is devoted to a recap and the discussion of a few open problems; finally Section 6 gives a selection of the most relevant references.

1. Bits of metric geometry

The apparent redundancy in the title of this section is intended to stress the fact that we shall be concerned with geometry only from the metric point of view (rather than from the topological, or differential point of view), be it either in some possibly nonsmooth metric space, or in a smooth Riemannian manifold.

1.1. **Length.** Let (X, d) be a complete separable metric space. Given a Lipschitz curve $\gamma : [0, T] \to X$, we define its **length** by

$$L(\gamma) := \sup \left\{ \sum_{i=0}^{N} d(\gamma(t_i), \gamma(t_{i+1})) \,\Big|\, 0 = t_0 \le t_1 \le \ldots \le t_{N+1} = T \right\}.$$

It is easily checked that the length of a curve is invariant by reparameterization.

In an abstract metric space the velocity $\dot\gamma(t)$ of a Lipschitz curve does not make sense; still it is possible to give a meaning to the "modulus of the velocity", or metric derivative of γ, or **speed**:

$$|\dot\gamma(t)| := \limsup_{h\to 0} \frac{d(\gamma(t+h),\gamma(t))}{|h|}$$

For almost all t, the above limsup is a true limit [1, Theorem 4.1.6], and the following formula holds:

$$L(\gamma) = \int_0^T |\dot\gamma(t)|\,dt. \tag{1.1}$$

1.2. **Length spaces.** In the previous subsection we have seen how to write the length of curves in terms of the metric d. But once the length is defined, we can introduce a new distance on X:

$$d'(x,y) := \inf\left\{L(\gamma)\,\Big|\,\gamma\in \mathrm{Lip}([0,1],X),\ \gamma(0)=x,\ \gamma(1)=y\right\}.$$

By triangle inequality, $d'\geq d$. If $d'=d$ we say that (X,d) is a **length space**. It is worth recording that (X,d') defined as above is automatically a length space.

Example 1.1. Take $X=\mathbb{S}^1\subset\mathbb{R}^2$, and $d(x,y)=|x-y|$ the standard Euclidean distance in $\mathbb{R}^2$; then $d'(x,y)=2\arcsin\frac{|x-y|}{2}$, so (X,d) is not a length space. More generally, if X is a closed subset of $\mathbb{R}^n$ then (X,d) is a length space if and only if X is convex.

1.3. **Geodesics.** A curve $\gamma:[0,1]\to X$ which minimizes the length, among all curves with $\gamma(0)=x$ and $\gamma(1)=y$, is called a **geodesic**, or more properly a **minimizing geodesic**.

The property of being a minimizing geodesic is stable by restriction: if $\gamma:[0,1]\to X$ is a geodesic, then for all $a<b\in[0,1]$, $\gamma|_{[a,b]}:[a,b]\to X$ is a geodesic from $\gamma(a)$ to $\gamma(b)$.

A length space such that any two points are joined by a minimizing geodesic is called a **geodesic space**.

Example 1.2. By the (generalized) Hopf–Rinow theorem, any locally compact complete length space is a geodesic space [4].

It is a general fact that a Lipschitz curve γ can be reparameterized so that $|\dot\gamma(t)|$ is constant [1, Theorem 4.2.1]. Thus, any geodesic $\gamma:[0,1]\to X$ can be reparameterized so that $|\dot\gamma(t)|=L(\gamma)$ for almost all $t\in[0,1]$. In this case, γ is called a **constant-speed minimizing geodesic**. Such curves are minimizers of the action functional

$$A(\gamma) := \frac{1}{2}\int_0^1 |\dot\gamma(t)|^2\,dt.$$

More precisely, we have:

Proposition 1.3. *Let (X,d) be a length space. Then*

$$d(x,y) = \inf_{\gamma(0)=x,\,\gamma(1)=y}\sqrt{\int_0^1 |\dot\gamma(t)|^2\,dt} \qquad \forall\, x,y\in X.$$

Moreover, if (X,d) is a geodesic space, then minimizers of the above functional are precisely constant-speed minimizing geodesics.

Sketch of the proof. By (1.1) we know that

$$d(x,y) = \inf_{\gamma(0)=x,\,\gamma(1)=y} \int_0^1 |\dot\gamma(t)|\,dt \qquad \forall\, x,y \in X.$$

By Jensen's inequality,

$$\int_0^1 |\dot\gamma(t)|\,dt \le \sqrt{\int_0^1 |\dot\gamma(t)|^2\,dt},$$

with equality if and only if $|\dot\gamma(t)|$ is constant for almost all t. The conclusion follows easily. $\square$

1.4. **Riemannian manifolds.** Given an n-dimensional C^∞ differentiable manifolds M, for each $x \in M$ we denote by T_xM the tangent space to M at x, and by $TM := \cup_{x\in M}(\{x\} \times T_xM)$ the whole tangent bundle of M. On each tangent space T_xM, we assume that is given a symmetric positive definite quadratic form $g_x : T_xM \times T_xM \to \mathbb{R}$ which depends smoothly on x; $g = (g_x)_{x\in M}$ is called a **Riemannian metric**, and (M,g) is a **Riemannian manifold**.

A Riemannian metric defines a scalar product and a norm on each tangent space: for each $v,w \in T_xM$

$$\langle v,w\rangle_x := g_x(v,w), \qquad |v|_x := \sqrt{g_x(v,v)}.$$

Let U be an open subset of $\mathbb{R}^n$ and $\Phi : U \to \Phi(U) = V \subset M$ a chart. Given $x = \Phi(x^1,\ldots,x^n) \in V$, the vectors $\frac{\partial}{\partial x^i} := \frac{\partial \Phi}{\partial x_i}(x^1,\ldots,x^n)$, $i = 1,\ldots,n$, constitute a basis of T_xM: any $v \in T_xM$ can be written as $v = \sum_{i=1}^n v^i \frac{\partial}{\partial x^i}$. We can use this chart to write our metric g in coordinates inside V:

$$g_x(v,v) = \sum_{i,j=1}^n g_x\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) v^i v^j = \sum_{i,j=1}^n g_{ij}(x) v^i v^j,$$

where by definition $g_{ij}(x) := g_x\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)$. We also denote by g^{ij} the coordinates of the inverse of g: $g^{ij} = (g_{ij})^{-1}$; more precisely, $\sum_j g^{ij} g_{jk} = \delta^i_k$, where δ^i_k denotes Kronecker's delta:

$$\delta^i_k = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{if } i \neq k. \end{cases}$$

In the sequel we will use these coordinates to perform many computations. Einstein's convention of summation over repeated indices will be used systematically: $a_k b^k = \sum_k a_k b^k$, $g_{ij} v^i v^j = \sum_{i,j} g_{ij} v^i v^j$, etc.

1.5. **Riemannian distance and volume.** The notion of "norm of a tangent vector" leads to the definition of a distance on a Riemannian manifold (M,g), called **Riemannian distance**:

$$\begin{aligned} d(x,y) &= \inf_{\gamma(0)=x,\,\gamma(1)=y} \int_0^1 \sqrt{g_{\gamma(t)}\big(\dot\gamma(t),\dot\gamma(t)\big)}\,dt \\ &= \inf_{\gamma(0)=x,\,\gamma(1)=y} \sqrt{\int_0^1 g_{\gamma(t)}\big(\dot\gamma(t),\dot\gamma(t)\big)\,dt} \qquad \forall\, x,y \in X. \end{aligned}$$

This definition makes (M,d) a length space.

A Riemannian manifold (M, g) is also equipped with a natural reference measure, the **Riemannian volume**:

$$\mathrm{vol}(dx) = n\text{-dimensional Hausdorff measure on } (M, d) = \sqrt{\det(g_{ij})}\, dx^1 \dots dx^n.$$

This definition of the volume allows to write a change of variables formula, exactly as in $\mathbb{R}^n$ (see for instance [45, Chapter 1]).

1.6. **Differential and gradients.** Given a smooth map $\varphi : M \to \mathbb{R}$, its differential $d\varphi : TM \to \mathbb{R}$ is defined as

$$d\varphi(x) \cdot v := \frac{d}{dt}\Big|_{t=0} \varphi(\gamma(t)),$$

where $\gamma : (-\varepsilon, \varepsilon) \to M$ is any smooth curve such that $\gamma(0) = x$, $\dot\gamma(0) = v$ (this definition is independent of the choice of γ). Thanks to the Riemannian metric, we can define the **gradient** $\nabla\varphi(x)$ at any point $x \in M$ by the formula

$$\big\langle \nabla\varphi(x), v \big\rangle_x := d\varphi(x) \cdot v.$$

Observe carefully that $\nabla\varphi(x)$ is a tangent vector (i.e. an element of T_xM), while $d\varphi(x)$ is a cotangent vector (i.e. an element of $T_x^*M := (T_xM)^*$, the dual space of T_xM). Using coordinates induced by a chart, we get

$$g_{ij}(\nabla\varphi)^i v^j = (d\varphi)_i v^i \quad \Longrightarrow \quad (\nabla\varphi)^i = g^{ij}(d\varphi)_j.$$

1.7. **Geodesics in Riemannian geometry.** On a Riemannian manifold, constant-speed minimizing geodesics satisfy a second order differential equation:

$$\ddot\gamma^k + \Gamma^k_{ij}\dot\gamma^i\dot\gamma^j = 0, \tag{1.2}$$

where Γ^k_{ij} are the Christoffel symbols defined by

$$\Gamma^k_{ij} = \frac{1}{2} g^{k\ell}\left(\frac{\partial g_{j\ell}}{\partial x^i} + \frac{\partial g_{i\ell}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^\ell}\right).$$

Exercise 1.4. Prove the above formula.
Hint. Consider the action functional $A(\gamma) = \frac{1}{2}\int_0^1 g_{\gamma(t)}\big(\dot\gamma(t), \dot\gamma(t)\big)\, dt$ for a geodesic γ and, working in charts, make variations of the form $A(\gamma + \varepsilon h)$, with h vanishing at the end points. Then use $\frac{d}{d\varepsilon}|_{\varepsilon=0} A(\gamma + \varepsilon h) = 0$ and the arbitrariness of h.

1.8. **Exponential map and cut locus.** From now on, by a geodesic we mean a solution of the geodesic equation (1.2), and we will explicitly mention whether it is or not minimizing. We also assume M to be complete, so that geodesics are defined for all times.

The **exponential map** $\exp : TM \to M$ is defined by

$$\exp_x(v) := \gamma_{x,v}(1),$$

where $\gamma_{x,v} : [0, +\infty) \to M$ is the unique solution of (1.2) starting at $\gamma_{x,v}(0) = x$ with velocity $\dot\gamma_{x,v}(0) = v$.

We observe that the curve $(\exp_x(tv))_{t\geq 0}$ is a geodesic defined for all times, but in general is not minimizing for large times (on the other hand, it is possible to prove that $\exp_x(tv)$ is always minimizing between x and $\exp_x(\varepsilon v)$ for $\varepsilon > 0$ sufficiently small). We define the **cut time** $t_c(x, v)$ as

$$t_c(x, v) := \inf\Big\{t > 0 \,|\, s \mapsto \exp_x(sv) \text{ is not minimizing between } x \text{ and } \exp_x(tv)\Big\}.$$

Example 1.5. On the sphere $\mathbb{S}^n$, the geodesics starting from the north pole $N = (0, \dots, 0, 1)$ with unit speed describe great circles passing through the south pole $S = (0, \dots, 0, -1)$. These geodesics are minimizing exactly until they reach S after a time π. Thus $t_c(N, v) = \pi$ for any $v \in T_N M$ with unit norm. By homogeneity of the sphere and time-rescaling, we get $t_c(x, v) = \frac{\pi}{|v|_x}$ for any $x \in \mathbb{S}^n$, $v \in T_x M \setminus \{0\}$.

Given two points $x, y \in M$, if there exists a unique minimizing geodesic $(\exp_x(tv))_{0 \le t \le 1}$ going from x to y in time 1, we will write (with a slight abuse of notation) $v = (\exp_x)^{-1}(y)$.

Given $x \in M$, we define the **cut locus** of x as

$$\mathrm{cut}(x) := \Big\{\exp_x\big(t_c(x, \xi)\xi\big) \,|\, \xi \in T_x M,\ |\xi|_x = 1\Big\}$$

We further define

$$\mathrm{cut}(M) := \{(x, y) \in M \times M \;|\; y \in \mathrm{cut}(x)\}.$$

Example 1.6. On the sphere $\mathbb{S}^n$, the cut locus of a point consists only of its antipodal point, i.e. $\mathrm{cut}(x) = \{-x\}$.

It is possible to prove that, if $y \notin \mathrm{cut}(x)$, then x and y are joined by a unique minimizing geodesic. The converse is close to be true: $y \notin \mathrm{cut}(x)$ if and only if there are neighborhoods U of x and V of y such that any two points $x' \in U$, $y' \in V$ are joined by a unique minimizing geodesic. In particular $y \notin \mathrm{cut}(x)$ if and only if $x \notin \mathrm{cut}(y)$.

1.9. **First variation formula and (super)differentiability of squared distance.** Exactly as in the computation for the geodesic equations (Exercise 1.4), one can compute the first variation of the action functional at a geodesic: let $\gamma : [0, 1] \to M$ be a constant-speed minimizing geodesic from x to y, and let $x' \simeq x + \delta x$ to $y' \simeq y + \delta y$ be perturbations of x and y respectively. (When $\delta x \in T_x M$ with $|\delta x| \ll 1$, $x + \delta x$ is an abuse of notation for, say, $\exp_x(\delta x)$, or for $h(1)$, where $h(s)$ is any smooth path with $h(0) = x$ and $\dot{h}(0) = \delta x$.) The **formulation of first variation** states that

$$A(\gamma') = A(\gamma) + \big(\langle \dot\gamma(1), \delta y\rangle_y - \langle \dot\gamma(0), \delta x\rangle_x\big) + O(|\delta y|^2) + O(|\delta x|^2).$$

Here γ' can be any curve from x' to y', geodesic or not, the important point is that γ' be a C^1 perturbation of γ. Below is a more rigorous statement:

Proposition 1.7. *Let $\gamma : [0, 1] \to M$ be a constant-speed minimizing geodesic from x to y. Consider a C^1 family of curves $\gamma^\varepsilon : [0, 1] \to M$, $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$ with $\gamma^0 = \gamma$, and let X be the vector field along γ defined by $X(t) = (d/d\varepsilon)|_{\varepsilon=0}\gamma^\varepsilon(t)$. Then*

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} A(\gamma^\varepsilon) = \langle \dot\gamma(1), X(1)\rangle_y - \langle \dot\gamma(0), X(0)\rangle_x.$$

The proof of this fact is analogous to the proof of the formula of the geodesic equation, with the only exception that now the variation we consider do not vanish at the boundary points, and so when doing integration by parts one has to take care of boundary terms.

We can now compute the (super)differential of the squared distance; the result is most conveniently expressed in terms of the gradient. Fix $x_0 \in M$, and consider the function $F(x) := \frac{1}{2} d(x_0, x)^2$.

Proposition 1.8. *If x_0 and x are joined by a unique minimizing geodesic, then F is differentiable at x and $\nabla F(x) = -(\exp_x)^{-1}(x_0)$.*

Sketch of the proof. Let $\gamma : [0,1] \to M$ be the unique constant-speed minimizing geodesic from x_0 to x, and let $x_\varepsilon \simeq x + \varepsilon w$ be a perturbation of x. Let γ_ε be a minimizing geodesic connecting x_0 to x_ε; so $\gamma_\varepsilon(t) = \exp_{x_0}(tv_\varepsilon)$ for some $v_\varepsilon \in T_{x_0}M$. Up to extraction of a subsequence, γ_ε converges to some minimizing geodesic, which is necessarily γ, and v_ε converges to $(\exp_{x_0})^{-1}(x)$. So $\gamma_\varepsilon(t)$ is a C^1 perturbation of γ, and the first variation formula yields

$$F(x_\varepsilon) = A(\gamma_\varepsilon) = A(\gamma) + \varepsilon \langle \dot\gamma(1), w \rangle_x + O(\varepsilon^2) = F(x) - \varepsilon \big\langle (\exp_x)^{-1}(x_0), w \big\rangle_x + O(\varepsilon^2).$$

□

In case x_0 and x are joined by several minimizing geodesics, the above argument fails. On the other hand, one still has **superdifferentiability**: there exists $p \in T_xM$ such that

$$F(x') \le F(x) + \langle p, \delta x \rangle_x + O(|\delta x|^2).$$

Proposition 1.9. *For any $x \in M$, F is superdifferentiable at x.*

Sketch of the proof. Let $\gamma : [0,1] \to M$ be a constant-speed minimizing geodesic from x_0 to x. Then, for any perturbation $x' \simeq x + \delta x$ of x, we can perturb γ into a smooth path γ' (not necessarily minimizing!) connecting x_0 to x'. The first variation formula yields

$$F(x') \le A(\gamma') = A(\gamma) + \langle \dot\gamma(1), \delta x \rangle_x + O(|\delta x|^2),$$

so $\dot\gamma(1)$ is a supergradient for F at x. □

By the above proposition we deduce that, although the (squared) distance is not smooth, its only singularities are upper crests. By the above proof we also see that F is differentiable at x if and only if x_0 and x are joined by a unique minimizing geodesic. (Indeed, a superdifferentiable function F is differentiable at x if and only if F has only one supergradient at x.)

1.10. **Hessian and second order calculus.** Let $\varphi : M \to \mathbb{R}$ be a smooth function. The **Hessian** $\nabla^2\varphi(x) : T_xM \to T_xM$ is defined by

$$\langle \nabla^2 \varphi(x) \cdot v, v \rangle_x := \frac{d^2}{dt^2}\bigg|_{t=0} \varphi(\gamma(t)),$$

where $\gamma(t) = \exp_x(tv)$. Observe that, when we defined the differential of a function, we could use any curve starting from x with velocity v. In this case, as the second derivative of $\varphi(\gamma(t))$ involves $\ddot\gamma(0)$, we are not allowed to choose an arbitrary curve in the definition of the Hessian.

1.11. **Variations of geodesics and Jacobi fields.** Let us consider a family $(\gamma_\theta)_{-\varepsilon \le \theta \le \varepsilon}$ of constant-speed geodesics $\gamma_\theta : [0,1] \to M$. Then, for each $t \in [0,1]$, we can consider the vector field

$$J(t) := \frac{\partial}{\partial \theta}\bigg|_{\theta=0} \gamma_\theta(t) \in T_{\gamma(t)}M.$$

The vector field J is called a **Jacobi field** along $\gamma = \gamma_0$. By differentiating the geodesic equations with respect to θ, we get a second order differential equation for J:

$$\frac{\partial}{\partial\theta}\left(\ddot{\gamma}_\theta^k + \Gamma_{ij}^k(\gamma_\theta)\dot{\gamma}_\theta^i\dot{\gamma}_\theta^j\right) = 0$$

gives

$$\ddot{J}^k + \frac{\partial\Gamma_{ij}^k}{\partial x^\ell}J^\ell\dot{\gamma}^i\dot{\gamma}^j + 2\Gamma_{ij}^k\dot{J}^i\dot{\gamma}^j = 0.$$

This complicated equation takes a nicer form if we choose time-dependent coordinates determined by a moving orthonormal basis $\{e_1(t), \dots, e_n(t)\}$ of $T_{\gamma(t)}M$, such that

$$\dot{e}_\ell^k(t) + \Gamma_{ij}^k(\gamma(t))e_\ell^i(t)\dot{\gamma}^j(t) = 0$$

(in this case, we say that the basis is **parallel transported** along γ). With this choice of the basis, defining $J^i(t) := \langle J(t), e_i(t)\rangle_{\gamma(t)}$ we get

$$\ddot{J}_i(t) + R_i^j(t)J_j(t) = 0.$$

For our purposes it suffices to know that R_i^j is a symmetric matrix; in fact one can show that $R_i^j(t) = \langle \mathrm{Riem}(\dot{\gamma}, e_i)\cdot\dot{\gamma}, e_j\rangle$, where Riem denotes the **Riemann tensor** of (M, g).

We now write the **Jacobi equation** in matrix form: let $\boldsymbol{J}(t) = (J_1(t), \dots, J_n(t))$ be a matrix of Jacobi fields, and define $J_{ij}(t) := \langle J_i(t), e_j(t)\rangle_{\gamma(t)}$, with $\{e_1(t), \dots, e_n(t)\}$ parallel transported as before. Then

$$\ddot{\boldsymbol{J}}(t) + R(t)\boldsymbol{J}(t) = 0,$$

where $R(t)$ is a symmetrix matrix involving derivatives of the metric $g_{ij}(\gamma(t))$ up to the second order, and such that (up to identification) $R(t)\dot{\gamma}(t) = 0$.

1.12. **Sectional and Ricci curvatures.** The matrix R appearing in the Jacobi fields equation allows to define the **sectional curvature** in a point x along a plane $P \subset T_xM$: let $\{e_1, e_2\}$ be an orthonormal basis of P, and consider γ the geodesic starting from x with velocity e_1. We now complete $\{e_1, e_2\}$ into an orthonormal basis of T_xM, and we construct $\{e_1(t), \dots, e_n(t)\}$ as above. Then the sectional curvature at x along P is given by

$$\mathrm{Sect}_x(P) := R_{22}(0)$$

Remark 1.10. The sectional curvature has the following geometric interpretation: given $v, w \in T_xM$ unit vectors, with (non-oriented) angle θ,

$$d\big(\exp_x(tv), \exp_x(tw)\big) = \sqrt{2(1-\cos\theta)}\,t\left(1 - \frac{\sigma\cos^2(\theta/2)}{6}t^2 + O(t^4)\right), \tag{1.3}$$

where σ is the sectional curvature at x along the plane generated by v and w. Thus the sectional curvatures infinitesimally measure the tendency of geodesics to converge ($\sigma > 0$) or diverge ($\sigma < 0$). We observe that formula (1.3) implies Gauss's Theorema Egregium, namely that the sectional curvature is invariant under local isometry.

The **Ricci curvature** at point $x \in M$ is a quadratic form on the tangent space defined as follows: fix $\xi \in T_xM$, and complete ξ to an orthonormal basis $\{\xi = e_1, e_2, \dots, e_n\}$. Denoting by $[e_i, e_j]$ the plane generated by e_i and e_j for $i \neq j$, we define

$$\mathrm{Ric}_x(\xi,\xi) := \sum_{j=2}^{n} \mathrm{Sect}_x([e_1, e_j]).$$

Another equivalent definition consists in considering the geodesic starting from x with velocity ξ, take $\{e_1(t), \dots, e_n(t)\}$ obtained by parallel transport, and define $\mathrm{Ric}_x(\xi,\xi) = \mathrm{tr}\big(R(0)\big)$.

1.13. **Interpretation of Ricci curvature bounds.** In this paragraph we give a geometric interpretation of the Ricci curvature. For more details we refer to [45, Chapter 14] and references therein.

Let ξ be a C^1 vector field (i.e. $\xi(x) \in T_xM$) defined in a neighborhood of $\{\gamma(t) \,|\, 0 \leq t \leq 1\}$, and consider the map

$$T_t(x) := \exp_x\big(t\xi(x)\big).$$

We want to compute the Jacobian of this map.

Think of $d_xT_t : T_xM \to T_{T_t(x)}M$ as an array $(J_1(t), \dots, J_n(t))$ of Jacobi fields. Expressed in an orthonoral basis $\{e_1(t), \dots, e_n(t)\}$ obtained by parallel transport, we get a matrix $\boldsymbol{J}(t)$ which solves the Jacobi equations $\ddot{\boldsymbol{J}} + R\boldsymbol{J} = 0$. Moreover, since $T_0(x) = x$ and $\dot{T}_0(x) = \xi(x)$, we have the system

$$\begin{cases} \ddot{\boldsymbol{J}} + R\boldsymbol{J} = 0, \\ \boldsymbol{J}(0) = I_n, \\ \dot{\boldsymbol{J}}(0) = \nabla\xi, \end{cases}$$

where $\nabla\xi$ is defined as

$$\langle \nabla\xi \cdot e_i, e_j \rangle := \frac{d}{ds}\bigg|_{s=0} \Big\langle \xi\big(\exp_x(se_i)\big), e_j\big(\exp_x(se_i)\big) \Big\rangle_{\exp_x(se_i)},$$

and $e_j\big(\exp_x(se_i)\big)$ is obtained by parallel transport along $s \mapsto \exp_x(se_i)$.

We now define $\mathscr{J}(t) := \mathrm{Jac}_x T_t = \det \boldsymbol{J}(t)$. Then

$$\frac{d}{dt} \log \mathscr{J}(t) = \mathrm{tr}\left(\dot{\boldsymbol{J}}(t)\boldsymbol{J}(t)^{-1}\right)$$

as long as $\det \boldsymbol{J}(t) > 0$. Let $U(t) := \dot{\boldsymbol{J}}(t)\boldsymbol{J}(t)^{-1}$. Using the Jacobi equation for $\boldsymbol{J}$, we get

$$\dot{U}(t) = -R(t) - U(t)^2.$$

Taking the trace, we deduce the important formula

$$\frac{d}{dt}\mathrm{tr}\big(U(t)\big) + \mathrm{tr}\left(U(t)^2\right) + \mathrm{Ric}_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t)) = 0.$$

Remark 1.11. The above formula is nothing else than the Lagrangian version of the celebrated **Bochner formula**, which is Eulerian in nature:

$$-\nabla \cdot \big((\xi \cdot \nabla)\xi\big) + \xi \cdot \nabla\big(\nabla \cdot \xi\big) + \mathrm{tr}\left((\nabla\xi)^2\right) + \mathrm{Ric}(\xi,\xi) = 0.$$

Assume now that $U(0)$ is symmetric (this is the case for instance if $\xi = \nabla\psi$ for some function ψ, as $U(0) = \nabla^2\psi$). In this case, since R is symmetric, U and U^* solves the same differential equation with the same initial condition, and by uniqueness $U(t)$ is symmetric too. We can therefore apply the inequality

$$\operatorname{tr}\big(U(t)^2\big) \geq \frac{1}{n}\big[\operatorname{tr}\big(U(t)\big)\big]^2$$

(a version of the Cauchy–Schwarz inequality). Combining all together we arrive at:

Proposition 1.12. *If $T_t(x) = \exp_x\big(t\nabla\psi(x)\big)$, then $\mathscr{J}(t) = \mathrm{Jac}_x T_t$ satisfies*

$$\frac{d^2}{dt^2}\log\mathscr{J}(t) + \frac{1}{n}\left(\frac{d}{dt}\log\mathscr{J}(t)\right)^2 + \mathrm{Ric}_{\gamma(t)}(\dot\gamma(t),\dot\gamma(t)) \leq 0.$$

From this proposition, we see that lower bounds on the Ricci curvature estimate the *averaged* tendency of geodesics to converge (in the sense of the Jacobian determinant).

The above proposition can be reversed, and one can prove for instance

$$\mathrm{Ric} \geq 0 \text{ throughout } M \qquad \Longleftrightarrow \qquad \frac{d^2}{dt^2}\log\mathrm{Jac}_x\big(\exp_x(t\nabla\psi(x))\big) \leq 0 \quad \forall\psi, \tag{1.4}$$

where ψ is arbitrary in the class of semiconvex functions defined in the neighborhood of x, such that $\mathrm{Jac}_x\big(\exp_x(t\nabla\psi)\big)$ remains positive on $[0,1]$. A more precise and more general discussion can be found in [45, Chapter 14].

1.14. **Why look for curvature bounds?** Sectional upper and lower bounds, and Ricci lower bounds, turn out to be very useful in many geometric applications. For instance, Ricci bounds appears in inequalities relating gradients and measures, such as:

- Sobolev inequalities;
- heat kernel estimates;
- compacteness of families of manifolds;
- spectral gap;
- diameter control.

For example, the Bonnet–Myers theorem states

$$\mathrm{Ric}_x \geq K g_x,\ K > 0 \quad \Longrightarrow \quad \mathrm{diam}(M) \leq \pi\sqrt{\frac{n-1}{K}},$$

while Sobolev's inequalities on n-dimensional compact manifolds say that, if $\mathrm{Ric}_x \geq K g_x$ for some $K \in \mathbb{R}$, then

$$\|f\|_{L^{(n-1)/n}(d\mathrm{vol})} \leq C(n, K, \mathrm{diam}(M))\,\big(\|f\|_{L^1(d\mathrm{vol})} + \|\nabla f\|_{L^1(d\mathrm{vol})}\big) \qquad \forall f.$$

1.15. **Stability issue and (measured) Gromov–Hausdorff convergence.** Sectional and Ricci curvatures are nonlinear combinations of derivatives of the metric g up to the second order. Therefore it is clear that if a sequence of Riemannian manifolds (M_k, g_k) converges (in charts) in C^2-topology to a Riemannian manifold (M, g), then both sectional and Ricci curvatures pass to the limit.

However much more is true: *lower bounds on these quantities pass to the limit under much weaker notions of convergence* (which is an indication of the stability/robustness of these bounds).

To make an analogy, consider the notion of convexity: a C^2 function $\phi : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if $\nabla^2\phi \geq 0$ everywhere. In particular if a sequence of convex functions ϕ_k converges to a ϕ is C^2-topology, then ϕ is convex. However it is well-known that convexity pass to the limit under much weaker notions of convergence (for instance, pointwise convergence).

In a geometric context, a powerful weak notion of convergence is the **Gromov–Hausdorff convergence**:

Definition 1.13. A sequence $(X_k, d_k)_{k\in\mathbb{N}}$ of compact length spaces is said to converge in the Gromov–Hausdorff topology to a metric space (X, d) if there are functions $f_k : X_k \to X$ and positive numbers $\varepsilon_k \to 0$, such that f_k is an ε_k-approximate isometry, i.e.

$$\begin{cases} |d(f_k(x), f_k(y)) - d_k(x,y)| \leq \varepsilon_k \quad \forall\, x, y \in X_k, \\ \\ \mathrm{dist}_d(f(X_k), X) \leq \varepsilon_k. \end{cases}$$

Here $\mathrm{dist}_d(A,B)$ denotes the distance between two sets $A, B \subset X$ measured with respect to d, namely $\mathrm{dist}_d(A,B) = \max(\sup_{x\in A}\inf_{y\in B} d(x,y),\ \sup_{y\in B}\inf_{x\in A} d(x,y))$.

One pleasant feature of the Gromov–Hausdorff convergence is that a limit of length spaces is a length space, and a limit of geodesic spaces is a geodesic space. The key property is summarized in the following:

Exercise 1.14. Let γ_k be a geodesic in X_k for each $k \in X_k$. Prove that, although $f_k \circ \gamma_k$ is a priori a discontinuous curve in X, up to extraction $f_k \circ \gamma_k$ converges to a geodesic in X.
Hint. Argue by contradiction.

In a Riemannian context, by abuse of notation one will say that a sequence of compact Riemannian manifolds (M_k, g_k) converges to a Riemannian manifold (M, g) if (M_k, d_k) converges to (M, d), where d_k (resp. d) is the geodesic distance on (M_k, g_k) (resp. (M, g)). It is remarkable that sectional curvature bounds do pass to the limit under Gromov–Hausdorff convergence [4]), however weak the latter notion:

Theorem 1.15. *Let (M_k, g_k) be a sequence of Riemannian manifolds converging in the Gromov–Hausdorff topology to a Riemannian manifold (M, g). If all sectional curvatures of (M_k, g_k) are all bounded from below by some fixed number $\kappa \in \mathbb{R}$, then also the sectional curvatures of (M, g) are bounded from below by κ.*

A slightly weaker notion of convergence takes care not only of the distances, but also of the measures; it is the **measured Gromov–Hausdorff convergence**, introduced in the present form by Fukaya (related notions were studied by Gromov):

Definition 1.16. A sequence $(X_k, d_k, \mu_k)_{k\in\mathbb{N}}$ of compact length spaces, equipped with reference Borel measures μ_k, is said to converge in the measured Gromov–Hausdorff topology to a measured metric space (X, d, μ) if there are functions $f_k : X_k \to X$, and positive numbers $\varepsilon_k \to 0$, such that f_k is an ε_k-approximate isometry, and $(f_k)_\#\mu_k$ converge in the weak topology to μ.

By abuse of notation, we shall say that a sequence of compact Riemannian manifolds (M_k, g_k) converges to another Riemannian manifold (M, g) in the measured Gromov–Hausdorff topology if $(M_k, d_k, \mathrm{vol}_k)$ converges to (M, d, vol), where d_k and vol_k (resp. d and vol) are the geodesic distance and volume measure associated to (M_k, g_k) (resp. (M, g)). After these preparations, we can state the stability result on Ricci lower bounds, which will be proved in Section 3 using elements of calculus of variations, optimal transport, and its relation to Ricci curvature. The following statement is a particular case of more general results proved independently in [28] and [38, 39]:

Theorem 1.17. *Let (M_k, g_k) be a sequence of compact Riemannian manifolds converging in the measured Gromov–Hausdorff topology to a compact Riemannian manifold (M, g). If the Ricci curvature of (M_k, g_k) is bounded below by $K\, g_k$, for some number $K \in \mathbb{R}$ independent of k, then also the Ricci curvature of (M, g) is bounded from below by $K\, g$.*

2. Solution of the Monge problem in Riemannian geometry

2.1. **The Monge problem with quadratic cost.** Let (M, g) be a Riemannian manifold, d its geodesic distance, and let $P(M)$ denote the space of probability measures on M. The **Monge problem with quadratic cost** on M is the following: given $\mu, \nu \in P(M)$, consider the minimization problem

$$\inf_{T_\# \mu = \nu} \int_M d(x, T(x))^2 \, d\mu(x).$$

Here the infimum is taken over all measurable maps $T : M \to M$ such that the **push-forward** $T_\# \mu$ of μ by T (i.e. the Borel probability measure defined by $T_\# \mu(A) := \mu(T^{-1}(A))$ for all Borel subsets A of M) coincides with ν.

This problem has a nice *engineering interpretation*: if we define $c(x, y) = \frac{1}{2} d(x, y)^2$ to be the cost to move a unit mass from x to y, then the above minimization problem simply consists in minimizing the total cost (=work) by choosing the destination $T(x)$ for each x.

It is also possible to give an equivalent *probabilistic interpretation* of the problem:

$$\inf \Big\{ \mathbb{E}[c(X, Y)] \,|\, \mathrm{law}(X) = \mu,\, \mathrm{law}(Y = \nu) \Big\},$$

so that we are minimizing a sort of correlation of two random variables, once their law is given.

Example 2.1. If $c(x, y) = -x \cdot y$ in $\mathbb{R}^n \times \mathbb{R}^n$, then we are just maximizing the correlation of the random variables X and Y, in the usual sense.

2.2. **Existence and uniqueness on compact manifolds.** In [32], McCann generalized Brenier's theorem [3] to compact Riemannian manifolds (see [12] or [45, Chapter 10] for the case of more general cost functions on arbitrary Riemannian manifolds). McCann proved:

Theorem 2.2. *Let (M, g) be a compact connected Riemannian manifold, let $\mu(dx) = f(x)\mathrm{vol}(dx)$ and $\nu(dy) = g(y)\mathrm{vol}(dy)$ be probability measures on M, and consider the cost $c(x, y) = \frac{1}{2} d(x, y)^2$. Then*

(1) There exists a unique solution T to the Monge problem.
(2) T is characterized by the structure $T(x) = \exp_x\big(\nabla\psi(x)\big)$ for some $\frac{d^2}{2}$-convex function $\psi : M \to \mathbb{R}$.
(3) For μ_0-almost all x, there exists a unique minimizing geodesic from x to $T(x)$, which is given by $t \mapsto \exp_x\big(t\nabla\psi(x)\big)$.
(4) $\mathrm{Jac}_x T = \frac{f(x)}{g(T(x))}$ μ-almost everywhere.

In the sequel of this section we shall explain this statement, and provide a sketch of the proof; much more details are in [45, Part I].

2.3. **c-convexity and c-subdifferential.** We recall that a function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semicontinuous convex if and only if

$$\varphi(x) = \sup_{y \in \mathbb{R}^n} \big[x \cdot y - \varphi^*(y)\big],$$

where

$$\varphi^*(x) := \sup_{x \in \mathbb{R}^n} \big[x \cdot y - \varphi(x)\big].$$

This fact is the basis for the definition of c-**convexity**, where $c : X \times Y \to \mathbb{R}$ is an arbitrary function:

Definition 2.3. A function $\psi : X \to \mathbb{R} \cup \{+\infty\}$ is c-convex if

$$\forall x \qquad \psi(x) = \sup_{y \in Y} \big[\psi^c(y) - c(x,y)\big],$$

where

$$\forall y \qquad \psi^c(y) := \inf_{x \in X} \big[\psi(x) + c(x,y)\big].$$

Moreover, for a c-convex function ψ, we define its c-**subdifferential** at x as

$$\partial^c \psi(x) := \big\{y \in Y \mid \psi(x) = \psi^c(y) - c(x,y)\big\}$$

With this general definition, when $c(x,y) = -x \cdot y$, the usual convexity coincides with the c-convexity and the usual subdifferential coindides with the c-subdifferential.

Remark 2.4. In the case of the Euclidean $\mathbb{R}^n$, a function ψ is $\frac{d^2}{2}$-convex if and only if $\psi(x) + \frac{|x|^2}{2}$ is convex.

The following facts are useful (see [45, Chapter 13]):

Proposition 2.5. *Let M be a compact Riemannian manifold. Then*

(*a*) *If $\psi : M \to \mathbb{R}$ is $\frac{d^2}{2}$-convex, then ψ is semiconvex (i.e. in any chart can be written as the sum of a convex and a smooth function).*

(*b*) *There exists a small number $\delta(M) > 0$ such that any function $\psi : M \to \mathbb{R}$ with $\|\psi\|_{C^2(M)} \leq \delta(M)$ is $\frac{d^2}{2}$-convex.*

(*c*) *If $\psi : M \to \mathbb{R}$ is a $\frac{d^2}{2}$-convex function of class C^2, then*

$$\nabla^2 \psi(x) + \frac{\nabla_x^2 d\big(x, \exp_x(\nabla \psi(x))\big)^2}{2} \geq 0,$$

where $\nabla_x^2 d(x,y)^2$ denotes the second derivative of the function $d^2(x,y)$ with respect to the x variable.

Remark 2.6. A natural question is whether condition (c) is also sufficient for $\frac{d^2}{2}$-convexity (at least for C^2 functions). As we will see in Subsection 4.5, this is the case under a suitable (forth order) condition on the cost function $\frac{d^2}{2}$.

2.4. **Sketch of the proof Theorem 2.2.** There are several ways to establish Theorem 2.2. One possibility is to go through the following five steps:

Step 1: Solve the Kantorovich problem
In [20, 21], Kantorovich introduced a notion of weak solution of the optimal transport problem: look for **transport plans** instead of transport maps. A transport map between two probability measures μ and ν is a measurable map T such that $T_{\#}\mu = \nu$; while a transport plan is a probability measure π on $M \times M$ whose marginals are μ and ν, i.e.

$$\int_{M\times M} h(x)\, d\pi(x,y) = \int_M h(x)\, d\mu(x), \qquad \int_{M\times M} h(y)\, d\pi(x,y) = \int_M h(y)\, d\nu(y),$$

for all $h : M \to \mathbb{R}$ bounded continuous. Denoting by $\Pi(\mu,\nu)$ the set of transport plans between μ and ν, the new minimization problem becomes

$$\inf_{\pi\in\Pi(\mu,\nu)} \left\{ \int_{M\times M} d(x,y)^2\, d\pi(x,y) \right\}. \tag{2.1}$$

A solution of this problem is called an **optimal transport plan**. The connection between the formulation of Kantorovich and that of Monge is the following: any transport map T induces the plan defined by $(\mathrm{Id}_X \times T)_{\#}\mu$ which is concentrated on the graph of T. Conversely, if a transport plan is concentrated on the graph of a measurable function T, then it is induced by this map.

By weak compactness of the set $\Pi(\mu,\nu)$ and continuity of the function $\pi \mapsto \int d(x,y)^2\, d\pi$, it is simple to prove the existence of an optimal transport plan $\bar\pi$; so to prove the existence of a solution to the Monge problem it suffices to show that $\bar\pi$ is automatically concentrated on the graph of a measurable map T, i.e.

$$y = T(x) \qquad \text{for } \bar\pi\text{-almost every } (x,y).$$

Once this fact is proved, the uniqueness of optimal maps will follow from the observation that, if T_1 and T_2 are optimal, then $\pi_1 := (\mathrm{Id}_X \times T_1)_{\#}\mu$ and $\pi_2 := (\mathrm{Id}_X \times T_2)_{\#}\mu$ are both optimal plans, so by linearity $\bar\pi = \frac{1}{2}(\pi_1 + \pi_2)$ is optimal. If it is concentrated on a graph, this implies $T_1 = T_2$ μ-almost everywhere

Step 2: The support of $\bar\pi$ is c-cyclically monotone
A set $S \subset M \times M$ is called **c-cyclically monotone** if, for all $N \in \mathbb{N}$, for all $\{(x_i, y_i)\}_{0\le i\le N} \subset S$, one has

$$\sum_{i=0}^N c(x_i, y_i) \le \sum_{i=0}^N c(x_i, y_{i+1}),$$

where by convention $y_{N+1} = y_0$.

The above definition heuristically means that, sending the point x_i to the point y_i for $i = 0, \dots, N$ is globally less expensive than sending the point x_i to the point y_{i+1}. It is therefore intuitive that, since $\bar\pi$ is optimal, its support is c-cyclically monotone (see [17] or [45, Chapter 5] for a proof).

Step 3: Any c-cyclically monotone set is contained in the c-subdifferential of a c-convex function
A proof of this fact (which is due to Rockafellar for $c(x,y) = -x \cdot y$, and Rüschendorf for the

general case) consists in constructing explicitly a c-convex function which does the work: given S c-cyclically monotone, we define

$$\psi(x) := \sup_{N\in\mathbb{N}} \sup_{\{(x_i,y_i)\}_{1\le i\le N}\subset S} \Big\{ \big[c(x_0,y_0)-c(x_1,y_0)\big]+\big[c(x_1,y_1)-c(x_2,y_1)\big]+\ldots+\big[c(x_N,y_N)-c(x,y_N)\big]\Big\},$$

where (x_0, y_0) is arbitrarily chosen in S. We leave as an exercise for the reader to check that with this definition ψ is c-convex, and that $S \subset \partial^c\psi(x) := \cup_{x\in M}(\{x\} \times \partial^c\psi(x))$.

Step 4: $\bar\pi$ is concentrated on a graph

Applying Steps 2 and 3, we know that the support of $\bar\pi$ is contained in the c-subdifferential of a c-convex function $\bar\psi$. Moreover, as said in Proposition 2.5, c-convex functions with $c = \frac{d^2}{2}$ are semiconvex. In particular $\bar\psi$ is Lipschitz, and so it is differentiable vol-almost everywhere. Since μ is absolutely continuous with respect to vol, we deduce that $\bar\psi$ is differentiable μ-almost everywhere. This further implies that, for $\bar\pi$-almost every (x, y), $\bar\psi$ is differentiable at x.

Now, let us fix a point $(\bar x, \bar y) \in \mathrm{supp}(\pi)$ such that $\bar\psi$ is differentiable at $\bar x$. To prove that $\bar\pi$ is concentrated on a graph, it suffices to prove that $\bar y$ is uniquely determined as a function of $\bar x$. To this aim, we observe that:

(a) Since the support of $\bar\pi$ is contained in the c-subdifferential of $\bar\psi$, we have $\bar y \in \partial^c\bar\psi(\bar x)$, and this implies that the function $x \mapsto \bar\psi(x) + c(x, \bar y)$ is subdifferentiable at $\bar x$ (and 0 belongs to the subdifferential).

(b) As shown in Subsection 1.9, $c(x, \bar y) = \frac{1}{2}d(x, \bar y)^2$ is superdifferentiable everywhere.

(c) $\bar\psi$ is differentiable at $\bar x$.

The combination of (a), (b) and (c) implies that $c(x, \bar y)$ is both upper and lower differentiable at $x = \bar x$, hence it is differentiable at $\bar x$. Since $\bar x$ was an arbitrary point where $\bar\psi$ is differentiable, this proves that

$$\nabla\bar\psi(x) + \nabla_x c(x, y) = 0 \qquad \text{for } \bar\pi\text{-almost every } (x, y).$$

By the first variation formula and the discussion in Subsection 1.9, this implies that there exists a unique geodesic joining x to y, and $\nabla\bar\psi(x) = (\exp_x)^{-1}(y)$. Thus we conclude that, for $\bar\pi$-almost every (x, y),

$$\begin{cases} y = \exp_x\big(\nabla\bar\psi(x)\big) \text{ (in particular } y \text{ is a function of } x), \\ \\ t \mapsto \exp_x\big(t\nabla\bar\psi(x)\big) \text{ is the unique minimizing geodesic between } x \text{ and } y. \end{cases}$$

Step 5: Change of variable formula

Here we give just a formal proof of the Jacobian equation, and we refer to [45, Chapter 11] for a rigorous proof.

Since π has marginals μ and ν, and is concentrated on the graph of T, for all bounded continuous functions $\zeta : M \to \mathbb{R}$ we have

$$\int_M \zeta(y)\, d\nu(y) = \int_{M\times M} \zeta(y)\, d\pi(x, y) = \int_M \zeta(T(x))\, d\mu(x),$$

that is $T_\#\mu = \nu$. Recalling that $\mu = f\mathrm{vol}$ and $\nu = g\mathrm{vol}$, we get, by change of variables,

$$\int_M \zeta(T(x))f(x)\, d\mathrm{vol}(x) = \int_M \zeta(y)g(y)\, d\mathrm{vol}(y) = \int_M \zeta(T(x))g(T(x))|\det(d_xT)|\, d\mathrm{vol}(x).$$

By the arbitrariness of ζ we conclude that $f(x) = g(T(x))|\det(d_xT)|$ μ-almost everywhere.

2.5. **Interpretation of the function $\bar{\psi}$.** The function $\bar{\psi}$ appearing in the formula for the optimal transport map has an interpretation as the solution of a dual problem:

$$\sup_{\psi}\left[\int_M \psi^c(y)\,d\nu(y) - \int_M \psi(x)\,d\mu(x)\right].$$

The above maximization problem has the following economical interpretation: $\psi(x)$ is the price at which a "shipper" buys material at x, while $\psi^c(y)$ is the price at which he sells back the material at y. Then, since

$$\psi^c(y) = \inf_x\big[\psi(x) + c(x,y)\big] = \sup\Big\{\varphi(y)\,|\,\varphi(y) \le \psi(x) + c(x,y)\Big\},$$

this means that $\psi^c(y)$ is the maximum selling price which is below the sum "buy price + transportation cost", that is the maximum price to be "competitive". In other words, the shipper is trying to maximize his profit.

To prove that $\bar{\psi}$ solves the above maximization problem, we observe that:

1) $\bar{\psi}^c(y) - \bar{\psi}(x) = c(x,y)$ on $\mathrm{supp}(\bar{\pi})$ for an optimal plan $\bar{\pi}$.
2) For any c-convex function ψ, $\psi^c(y) - \psi(x) \le c(x,y)$ on $M \times M$.

Combining these two facts, we get

$$\begin{aligned}\int_M \bar{\psi}^c(y)\,d\nu - \int_M \bar{\psi}(x)\,d\mu &= \int_{M\times M}\big[\bar{\psi}^c(y) - \bar{\psi}(x)\big]\,d\bar{\pi}(x,y) = \int_{M\times M} c(x,y)\,d\bar{\pi}(x,y)\\ &\ge \int_{M\times M}\big[\psi^c(y) - \psi(x)\big]\,d\bar{\pi}(x,y) = \int_M \psi^c\,d\nu - \int_M \psi\,d\mu,\end{aligned}$$

and so $\bar{\psi}$ is a maximizer.

3. Synthetic formulation of Ricci bounds

As we have seen in the last section, the optimal transport allows to construct maps $T(x) = \exp_x\big(\nabla\psi(x)\big)$, with $\psi : M \to \mathbb{R}$ globally defined. Moreover it involves a Jacobian formula for T. For this reason, it turns out to be a natural candidate for a global reformulation of Ricci curvature bounds. More precisely, Ricci curvature bounds can be reformulated in terms of *convexity inequalities* for certain nonlinear functionals of the density, along geodesics of optimal transport. This fact will provide the stability of such bounds, and other applications.

3.1. **The 2-Wasserstein space.** Let (M,g) be a compact Riemannian manifold, equipped with its geodesic distance d and its volume measure vol. We denote with $P(M)$ the set of probability measures on M. Let

$$W_2(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)}\left\{\int_{M\times M} d^2(x,y)\,d\pi(x,y)\right\}^{\frac{1}{2}}.$$

The quantity $W_2(\mu,\nu)$ is called the **Wasserstein distance of order 2** between μ and ν. It is well-known that it defines a finite metric on $P(M)$, and so one can speak about geodesic in the metric space $P_2(M) := (P(M), W_2)$. This space turns out to be a geodesic space (see e.g. [45, Chapter 7]). We denote with $P_2^{ac}(M)$ the subset of $P_2(M)$ that consists of the Borel probability measures on M that are absolutely continuous with respect to vol.

3.2. **Geodesics in $P_2(M)$.** Given $\mu_0, \mu_1 \in P_2(M)$, we want to construct a geodesic between them. In general the geodesic is not unique, as can be seen considering $\mu_0 = \delta_x$ and $\mu_1 = \delta_y$, where x and y can be joined by several minimizing geodesics. Indeed if $\gamma : [0,1] \to M$ is a geodesic form x to y, then $\mu_t := \delta_{\gamma(t)}$ is a Wasserstein geodesic from δ_x to δ_y. (Check it!)

On the other hand, if μ_0 (or equivalently μ_1) belongs to $P_2^{ac}(M)$, this problem has a simple answer (see [45, Chapter 7] for a general treatment):

Proposition 3.1. *Assume $\mu_0 \in P_2^{ac}(M)$, and let $T(x) = \exp_x\big(\nabla\psi(x)\big)$ be the optimal map between μ_0 and μ_1. Then the unique geodesic from μ_0 to μ_1 is given by $\mu_t := (T_t)_\#\mu_0$, with $T_t(x) := \exp_x\big(t\nabla\psi(x)\big)$.*

Proof. To prove that μ_t is a geodesic, we observe that

$$\begin{aligned} W_2(\mu_s,\mu_t)^2 &\le \int_M d\big(\exp_x\big(s\nabla\psi(x)\big), \exp_x\big(t\nabla\psi(x)\big)\big)\, d\mu_0(x) \\ &= (s-t)^2 \int_M |\nabla\psi(x)|_x^2\, d\mu_0(x) = (t-s)^2 W_2(\mu_0,\mu_1)^2. \end{aligned}$$

This implies that $W_2(\mu_s,\mu_t) \le |t-s| W_2(\mu_0,\mu_1)$ for all $s,t \in [0,1]$, so the length of the path $(\mu_t)_{0\le t\le 1}$ satisfies

$$L\big((\mu_t)_{0\le t\le 1}\big) \le W_2(\mu_0,\mu_1).$$

But since the converse inequality is always true, we get $L\big((\mu_t)_{0\le t\le 1}\big) = W_2(\mu_0,\mu_1)$, so that μ_t is a geodesic.

The fact that μ_t is the unique geodesic is a consequence Theorem 2.2(c), together with the general fact that any Wasserstein geodesic takes the form $\mu_t = (e_t)_\#\Pi$, where Π is a probability measure on the set Γ of minimizing geodesics, and $e_t : \Gamma \to M$ is the evaluation map at time t: $e_t(\gamma) := \gamma(t)$ (see [45, Theorem 7.21 and Corollary 7.23]). ☐

Remark 3.2. One can write down the geodesic equations for $P_2(M)$, which has to be understood in a suitable weak sense (see [45, Chapter 13]):

$$\begin{cases} \dfrac{\partial \mu_t}{\partial t} + \operatorname{div}(\mu_t \nabla\psi_t) = 0, \\[2ex] \dfrac{\partial \psi_t}{\partial t} + \dfrac{|\nabla\psi_t|^2}{2} = 0. \end{cases}$$

3.3. **Approximate geodesics in Wasserstein space.** A key property of the Wasserstein space is that it depends continuously on the basis space, when the topology is the Gromov–Hausdorff topology. The following statement is proven in [28, Proposition 4.1]: *If $f_k : M_k \to M$ are ε_k-approximate isometries, with $\varepsilon_k \to 0$, then $(f_k)_\# : P_2(M_k) \to P_2(M)$ are $\tilde\varepsilon_k$-approximate isometries, with $\tilde\varepsilon_k \to 0$.*

3.4. **Reformulation of** $\mathrm{Ric} \ge 0$. Let

$$H(\mu) := H_{\mathrm{vol}}(\mu) = \begin{cases} \displaystyle\int_M \rho\log(\rho)\, d\mathrm{vol} & \text{if } \mu = \rho\mathrm{vol}, \\ +\infty & \text{otherwise.} \end{cases}$$

This is the **Boltzmann H functional**, or negative of the **Boltzmann entropy**. As shown in [37] (as a development of the works in [36, 9]), the inequality $\mathrm{Ric} \geq 0$ can be reformulated in terms of the convexity of H along Wasserstein geodesics:

Theorem 3.3. *Let (M, g) be a compact Riemannian manifold. Then* $\mathrm{Ric} \geq 0$ *if and only if $t \mapsto H(\mu_t)$ is a convex function of $t \in [0, 1]$ for all Wasserstein geodesics $(\mu_t)_{0 \leq t \leq 1}$.*

More generally, $\mathrm{Ric} \geq K\,g$ if and only if, for all $\mu_0, \mu_1 \in P_2^{ac}(M)$,

$$H(\mu_t) \leq (1-t)H(\mu_0) + tH(\mu_1) - K\,\frac{t(1-t)}{2} W_2(\mu_0, \mu_1)^2 \qquad \forall t \in [0, 1], \tag{3.1}$$

where $(\mu_t)_{0 \leq t \leq 1}$ is the Wasserstein geodesic between μ_0 and μ_1.

Sketch of the proof. Let us consider just the case $K = 0$. By using the Jacobian equation (Theorem 2.2(d)), we get

$$\begin{aligned} H(\mu_t) &= \int_M \rho_t(x) \log\bigl(\rho_t(x)\bigr)\, d\mathrm{vol}(x) = \int_M \rho_t(T_t(x)) \log\bigl(\rho_t(T_t(x))\bigr) \mathrm{Jac}_x T_t\, d\mathrm{vol}(x) \\ &= \int_M \rho_0(x) \log\left(\frac{\rho_0(x)}{\mathrm{Jac}_x T_t}\right) d\mathrm{vol}(x) = H(\mu_0) - \int_M \log \mathrm{Jac}_x\bigl(\exp_x(t\nabla\psi)\bigr)\, d\mu_0. \end{aligned}$$

Then the direct implication follows from (1.4). The converse implication is obtained also from (1.4), using Proposition 2.5(b) to explore all tangent directions by minimizing geodesics in Wasserstein space. Details appear e.g. in [45, Chapter 17]. □

3.5. **Application: stability.** Let us give a sketch of the proof of Theorem 1.17; we refer to [28] and [45, Chapters 28 and 29] for details.

First of all, we reformulate the inequality $\mathrm{Ric}(M_k) \geq K\,g_k$ in terms of the convexity inequality (3.1); the goal is to prove that the inequality (3.1) holds on the limit manifold M.

Let $\mu_0 = \rho_0 \mathrm{vol}, \mu_1 = \rho_1 \mathrm{vol} \in P_2^{ac}(M)$, and define on M_k the probability measures

$$\mu_0^k := \frac{\rho_0 \circ f_k \mathrm{vol}_k}{\displaystyle\int_{M_k} \rho_0 \circ f_k\, d\mathrm{vol}_k}, \qquad \mu_1^k := \frac{\rho_1 \circ f_k \mathrm{vol}_k}{\displaystyle\int_{M_k} \rho_1 \circ f_k\, d\mathrm{vol}_k},$$

where $f_k : M_k \to M$ are the approximate isometries appearing in the definition of the measured Gromov–Hausdorff convergence. Let $(\mu_t^k)_{0 \leq t \leq 1}$ be the Wasserestein geodesic between μ_0^k and μ_1^k. Up to extraction of a subsequence, $(f_k)_\# \mu_t^k$ converges, uniformly in $t \in [0, 1]$, to a geodesic μ_t in $P_2(M)$ between μ_0 and μ_1 (recall Exercise 1.14). It remains to show that the inequality (3.1) passes to the limit. Let us consider separately the three terms in this inequality

(a) The term $H_k(\mu_0^k)$ passes to the limit

By an approximation argument, it suffices to consider the case $\rho_0 \in C(M)$. Then

$$Z_k := \int_{M_k} \rho_0 \circ f_k\, d\mathrm{vol}_k = \int_M \rho_0\, d(f_k)_\# \mathrm{vol}_k \longrightarrow \int_M \rho_0\, d\mathrm{vol} = 1,$$

so that (with obvious notation)

$$H_k(\mu_0^k) = \int_{M_k} \left(\frac{\rho_0 \circ f_k}{Z_k}\right) \log\left(\frac{\rho_0 \circ f_k}{Z_k}\right) d\mathrm{vol}_k \simeq \int_{M_k} \rho_0 \circ f_k \log(\rho_0 \circ f_k)\, d\mathrm{vol}_k$$
$$= \int_M \rho_0 \log(\rho_0)\, d(f_k)_\# \mathrm{vol}_k \longrightarrow \int_M \rho_0 \log(\rho_0)\, d\mathrm{vol} = H(\mu_0).$$

The case of $H_k(\mu_1^k)$ is analogous.

(b) The term $W_2(\mu_0^k, \mu_1^k)^2$ passes to the limit
This follows from the fact that $(f_k)_\#$ are $\tilde{\varepsilon}_k$-approximate isometries, and the Wasserstein distance on a compact manifold metrizes the weak convergence: so

$$W_2(\mu_0^k, \mu_1^k) \simeq W_2\big((f_k)_\# \mu_0^k, (f_k)_\# \mu_1^k\big) \longrightarrow W_2(\mu_0, \mu_1).$$

(c) The term $H_k(\mu_t^k)$ is lower semicontinuous under weak convergence
This comes from the following general property: If $U : \mathbb{R}^+ \to \mathbb{R}$ is convex and continuous, then

- $P_2(M) \times P_2(M) \ni (\mu, \nu) \longmapsto \int U\left(\frac{d\mu}{d\nu}\right) d\nu$ is a convex lower semicontinuous functional.
- $\displaystyle\int U\left(\frac{d(f_\#\mu)}{d(f_\#\nu)}\right) d(f_\#\nu) \leq U\left(\frac{d\mu}{d\nu}\right) d\nu$ for any function $f : M \to M$.

Combining these two facts, we get

$$H(\mu_t) \leq \liminf_{k\to\infty} H_{(f_k)_\#\mathrm{vol}_k}\big((f_k)_\# \mu_t^k\big) \leq \liminf_{k\to\infty} H_{\mathrm{vol}_k}\big(\mu_t^k\big),$$

which is the desired inequality.

4. The smoothness issue

Let (M, g) be a compact connected Riemannian manifold, let $\mu(dx) = f(x)\mathrm{vol}(dx)$ and $\nu(dy) = g(y)\mathrm{vol}(dy)$ be probability measures on M, and consider the cost $c(x, y) = \frac{1}{2} d(x, y)^2$. Assume f and g are C^∞ and strictly positive on M. Is the optimal map T smooth?

A positive answer to this problem has been given in the Euclidean space [43, 10, 5, 6, 7, 44] and in the case of the flat torus [8], but the general question of Riemannian manifolds remained open until the last years. Only recently, after two key papers of Ma, Trundinger and Wang [30] and Loeper [24], did specialists understand a way to attack this problem; see [45, Chapter 12] for a global picture and references.

There are several motivations for the investigation of the smoothness of the optimal map:

- It is a typical PDE/analysis question.
- It is a step towards a qualitative understanding of the optimal transport map.
- If it is a general phenomenon, then nonsmooth situations may be treated by regularization, instead of working directly on nonsmooth objects.

Moreover, as we will see, the study of this regularity issue allows to understand some geometric properties of the Riemannian manifold itself.

4.1. **The PDE.** Starting from the Jacobian equation

$$\big|\det(d_x T)\big| = \frac{f(x)}{g(T(x))},$$

and the relation $T(x) = \exp_x\big(\nabla\psi(x)\big)$, we can write a PDE for ψ. Indeed, since

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

differentiating with respect to x and using the Jacobian equation we get

$$\det\Big[\nabla^2\psi(x) + \nabla_x^2 c\big(x, \exp_x\big(\nabla\psi(x)\big)\big)\Big] = \frac{f(x)}{g(T(x))\big|\det(d_{\nabla\psi(x)}\exp_x)\big|}.$$

(Observe that, since ψ is c-convex, the matrix appearing in the left-hand side is positive definite by Proposition 2.5(c).)

We see that ψ solves a Monge–Ampère type equation with a perturbation $\nabla_x^2 c\big(x, \exp_x\big(\nabla\psi(x)\big)\big)$ which is of first order in ψ. Unfortunately, for Monge–Ampère type equations lower order terms do matter, and it turns out that it is exactly the term $\nabla_x^2 c\big(x, \exp_x\big(\nabla\psi(x)\big)\big)$ which can create obstructions to the smoothness.

4.2. **Obstruction I: Local geometry.** We now show how negative sectional curvature is an obstruction to regularity (indeed even to continuity) of optimal maps. We refer to [45, Theorem 12.4] for more details on the construction given below.

Let $M = \mathbb{H}^2$ be the hyperbolic plane (or a compact quotient thereof). Fix a point O as the origin, and fix a local system of coordinates in a neighborhood of O such that the maps $(x_1, x_2) \mapsto (\pm x_1, \pm x_2)$ are local isometries (it suffices for instance to consider the model of the Poincaré disk, with O equal to the origin in $\mathbb{R}^2$). Then define the points

$$A^\pm = (0, \pm\varepsilon), \quad B^\pm = (\pm\varepsilon, 0) \qquad \text{for some } \varepsilon > 0.$$

Take a measure μ symmetric with respect to 0 and concentrated near $\{A^+\}\cup\{A^-\}$ (say 3/4 of the total mass belongs to a small neighborhood of $\{A^+\}\cup\{A^-\}$), and a measure ν symmetric with respect to 0 and concentrated near $\{B^+\}\cup\{B^-\}$. Moreover assume that μ and ν are absolutely continuous, and have strictly positive densities everywhere. We denote by T the unique optimal transport map, and we assume by contradiction that T is continuous. By symmetry, we deduce that $T(O) = O$. Then, by counting the total mass, we deduce that there exists a point A' close to A^+ which is sent to a point B' near, say, B^+.

But, by negative curvature (if A' and B' are close enough to A and B respectively), Pythagore's Theorem becomes an inequality: $d(O, A')^2 + d(O, B')^2 < d(A', B')^2$, and this contradicts the c-cyclical monotonicity of the support of an optimal plan (see Step 2 of Theorem 2.2).

4.3. **Obstruction II: Topology of the c-subdifferential.** Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex function; its differential $\partial\varphi(x)$ is given by

$$\partial\varphi(x) = \big\{y \in \mathbb{R}^n \,|\, \varphi(x) + \varphi^*(y) = x \cdot y\big\} = \big\{y \in \mathbb{R}^n \,|\, \varphi(z) - z\cdot y \geq \varphi(x) - x \cdot y \quad \forall\, z \in \mathbb{R}^n\big\}.$$

Then $\partial\varphi(x)$ is a convex set, a fortiori connected.

If we now consider $\psi : M \to \mathbb{R}$ a c-convex function, $c = \frac{d^2}{2}$, then (see Subsection 2.3)

$$\partial^c\psi(x) = \big\{y \in M \,|\, \psi(x) = \psi^c(y) - c(x,y)\big\} = \big\{y \in M \,|\, \psi(z) + c(z,y) \geq \psi(x) + c(x,y) \quad \forall z \in M\big\}.$$

In this generality there is no reason for $\partial^c\psi(x)$ to be connected — and in general, this is not the case!

Following the construction given in [30, Section 7.3], Loeper showed that under adequate assumptions the connectedness of the c-subdifferential is a necessary condition for the smoothness of optimal transport [24] (see also [45, Theorem 12.7]):

Theorem 4.1. *Assume that there exist $\bar{x} \in M$ and $\psi : M \to \mathbb{R}$ c-convex such that $\partial^c\psi(\bar{x})$ is not (simply) connected. Then one can construct f and g, C^∞ strictly positive probability densities on M, such that the optimal map T from $f$vol to $g$vol is discontinuous.*

4.4. **Conditions for the connectedness of $\partial^c\psi$.** We now wish to find some simple enough conditions implying the connectedness of sets $\partial^c\psi$.

First attempt: Let us look at the simplest c-convex functions:

$$\psi(x) := -c(x, y_0) + a_0.$$

Assume that $\bar{x} \notin \mathrm{cut}(y)$, and let $\bar{y} \in \partial^c\psi(\bar{x})$. Then the function $\psi(x) + c(x, \bar{y})$ achieves its minimum at $x = \bar{x}$, so $\bar{x} \notin \mathrm{cut}(\bar{y})$ (see the argument in Step 4 of Theorem 2.2) and

$$-\nabla_x c(\bar{x}, y_0) + \nabla_x c(\bar{x}, \bar{y}) = 0.$$

Thus $(\exp_{\bar{x}})^{-1}(y_0) = (\exp_{\bar{x}})^{-1}(\bar{y})$ (see Subsection 1.9), which implies $\bar{y} = y_0$. In conclusion $\partial^c\psi(\bar{x}) = \{y_0\}$ is a singleton, automatically connected — so we do not learn anything!

Second attempt: The second simplest example of c-convex functions are

$$\psi(x) := \max\big\{-c(x, y_0) + a_0, -c(x, y_1) + a_1\big\}.$$

Take a point $\bar{x} \notin \mathrm{cut}(y)$ belonging to the set $\{-c(x, y_0)+a_0 = -c(x, y_1)+a_1\}$, and let $\bar{y} \in \partial^c\psi(\bar{x})$. Since $\psi(x) + c(x, \bar{y})$ attains its minimum at $x = \bar{x}$, we get

$$0 \in \nabla^-_{\bar{x}}\big(\psi + c(\cdot, \bar{y})\big),$$

or equivalently

$$-\nabla_x c(\bar{x}, \bar{y}) \in \nabla^-\psi(\bar{x})$$

(recall that by Proposition 2.5(a) ψ is a semiconvex function, so that its subgradient $\nabla^-\psi(\bar{x})$, which is defined in charts as the set $\{p \,|\, \psi(\bar{x} + \delta x) \geq \psi(\bar{x}) + \langle p, \delta x\rangle + o(|\delta x|)\}$, is convex and non-empty). From the above inclusion we deduce that $\bar{y} \in \exp_{\bar{x}}\big(\nabla^-\psi(\bar{x})\big)$ (see Subsection 1.9). Moreover, it is not difficult to see that

$$\nabla^-\psi(\bar{x}) = \{(1-t)v_0 + tv_1 \,|\, t \in [0,1]\}, \qquad v_i := \nabla_x c(\bar{x}, y_i) = (\exp_{\bar{x}})^{-1}(y_i), \quad i = 0, 1.$$

Therefore, denoting by $[v_0, v_1]$ the segment joining v_0 and v_1, we have proved the inclusion

$$\partial^c\psi(\bar{x}) \subset \exp_{\bar{x}}\big([v_0, v_1]\big).$$

The above formula suggests the following definition of **c-segment**:

Definition 4.2. Let $\bar{x} \in M$, $y_0, y_1 \notin \mathrm{cut}(\bar{x})$. Then we define the c-segment from y_0 to y_1 with base $\bar{x}$ as

$$[y_0, y_1]_{\bar{x}} := \big\{y_t = \exp_{\bar{x}}\big((1-t)(\exp_{\bar{x}})^{-1}(y_0) + t(\exp_{\bar{x}})^{-1}(y_1)\big) \,|\, t \in [0,1]\big\}.$$

In [24], Loeper proved (a sligtly weaker version of) the following result (see [45, Chapter 12] for the general result):

Theorem 4.3. *The following conditions are equivalent:*

(1) For any ψ c-convex, for all $\bar{x} \in M$, $\partial^c \psi(\bar{x})$ is connected.
(2) For any ψ c-convex, for all $\bar{x} \in M$, $(\exp_{\bar{x}})^{-1}\big(\partial^c\psi(\bar{x}) \setminus \mathrm{cut}(\bar{x})\big)$ is convex.
(3) For all $\bar{x} \in M$, for all $y_0, y_1 \notin \mathrm{cut}(\bar{x})$, if $[y_0, y_1]_{\bar{x}} = \{y_t\}_{t\in[0,1]}$ does not meet $\mathrm{cut}(\bar{x})$, then

$$d(x,y_t)^2 - d(\bar{x},y_t)^2 \geq \min\big[d(x,y_0)^2 - d(\bar{x},y_0)^2, d(x,y_1)^2 - d(\bar{x},y_1)^2\big] \tag{4.1}$$

for all $x \in M$, $t \in [0,1]$.
(4) For all $\bar{x} \in M$, for all $y \notin \mathrm{cut}(\bar{x})$, for all $\eta, \xi \in T_{\bar{x}}M$ with $\xi \perp \eta$,

$$\frac{d^2}{ds^2}\bigg|_{s=0} \frac{d^2}{dt^2}\bigg|_{t=0} d\big(\exp_{\bar{x}}(t\xi), \exp_{\bar{x}}(p+s\eta)\big)^2 \leq 0,$$

where $p = (\exp_{\bar{x}})^{-1}(y)$.

Moreover, if these conditions are not satisfied, C^1 c-convex functions are not dense in Lipschitz c-convex functions.

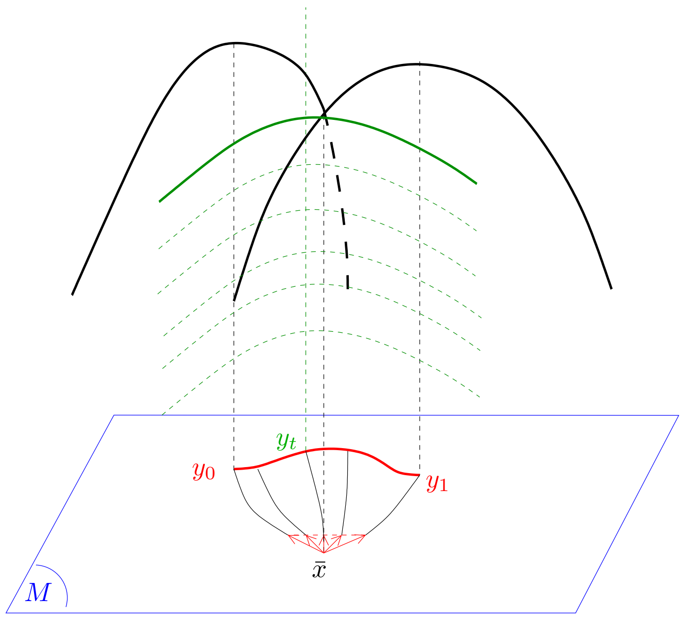

FIGURE 1. Property (3): the mountain grown from y_t emerges exactly at the pass between the mountains centered at y_0 and y_1.

Sketch of the proof. We give here only some elements of the proof.

(2) $\Rightarrow$ (1): since $(\exp_{\bar{x}})^{-1}\big(\partial^c\psi(\bar{x}) \setminus \mathrm{cut}(\bar{x})\big)$ is convex, it is connected, and so its image by $\exp_{\bar{x}}$ is connected too.

(1) ⇒ (2): for $\psi_{\bar{x},y_0,y_1} := \max\big\{-c(\cdot,y_0)+c(\bar{x},y_0), -c(\cdot,y_1)+c(\bar{x},y_1)\big\}$ we have $(\exp_{\bar{x}})^{-1}\big(\partial^c\psi(\bar{x})\big) \subset [(\exp_{\bar{x}})^{-1}(y_0), \exp_{\bar{x}})^{-1}(y_1)]$, which is a segment. Since in this case connectedness is equivalent to convexity, if (1) holds we obtain $\partial^c\psi_{\bar{x},y_0,y_1} = [y_0,y_1]_{\bar{x}}$.

In the general case, we fix $y_0, y_1 \in \partial^c\psi(\bar{x})$. Then it is simple to see that

$$\partial^c\psi(\bar{x}) \supset \partial^c\psi_{\bar{x},y_0,y_1}(\bar{x}) = [y_0,y_1]_{\bar{x}},$$

and the result follows.

(2) ⇔ (3): condition (4.1) is equivalent to $\partial^c\psi_{\bar{x},y_0,y_1} = [y_0,y_1]_{\bar{x}}$. Then the equivalence between (2) and (3) follows arguing as above.

(3) ⇒ (4): fix $\bar{x} \in M$, and let $y := \exp_{\bar{x}}(p)$. Take ξ, η orthogonal and with unit norm, and define

$$y_0 := \exp_{\bar{x}}(p - \varepsilon\eta), \quad y_1 := \exp_{\bar{x}}(p + \varepsilon\eta) \qquad \text{for some } \varepsilon > 0 \text{ small.}$$

Moreover, let

$$h_0(x) := c(\bar{x},y_0) - c(x,y_0), \qquad h_1(x) := c(\bar{x},y_1) - c(x,y_1), \qquad \psi := \max\big\{h_0,h_1\big\} = \psi_{\bar{x},y_0,y_1}.$$

We now define $\gamma(t)$ as a curve contained in the set $\{h_0 = h_1\}$ such that $\gamma(0) = \bar{x}$, $\dot{\gamma}(0) = \xi$. (See Figure 2.)

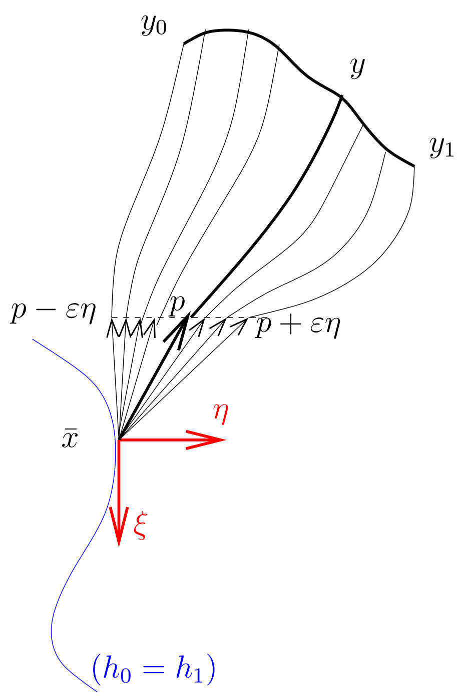

FIGURE 2. Proof of (3) ⇒ (4); y belongs to the c-segment with base $\bar{x}$ and endpoints y_0, y_1; ξ is tangent to the local hypersurface $(h_0 = h_1)$.

Since $y \in [y_0, y_1]_{\bar{x}}$, by (3) we get $y \in \partial^c\psi(\bar{x})$, so that

$$\frac{1}{2}\big[h_0(\bar{x})+h_1(\bar{x})\big]+c(\bar{x},y) = \psi(\bar{x})+c(\bar{x},y) \le \psi(\gamma(t))+c(\gamma(t),y) = \frac{1}{2}\big[h_0(\gamma(t))+h_1(\gamma(t))\big]+c(\gamma(t),y),$$

where we used that $h_0 = h_1$ along γ. Recalling the definition of h_0 and h_1, we deduce

$$\frac{1}{2}\big[c(\gamma(t),y_0)+c(\gamma(t),y_1)\big] - c(\gamma(t),y) \le \frac{1}{2}\big[c(\bar{x},y_0)+c(\bar{x},y_1)\big] - c(\bar{x},y),$$

so the function $t \mapsto \frac{1}{2}\big[c(\gamma(t),y_0)+c(\gamma(t),y_1)\big] - c(\gamma(t),y)$ achieves its maximum at $t=0$. This implies

$$\frac{d^2}{dt^2}\bigg|_{t=0}\Big[\frac{1}{2}\big(c(\gamma(t),y_0)+c(\gamma(t),y_1)\big) - c(\gamma(t),y)\Big] \le 0,$$

i.e.

$$\Big\langle \Big[\frac{1}{2}\big(\nabla_x^2 c(\bar{x},y_0)+\nabla_x^2 c(\bar{x},y_1)\big) - \nabla_x^2 c(\bar{x},y)\Big]\cdot\xi,\xi\Big\rangle \le 0$$

(here we used that $\nabla_x c(\bar{x},y) = \frac{1}{2}\big[\nabla_x c(\bar{x},y_0)+\nabla_x c(\bar{x},y_1)\big]$). Thus the function

$$\eta \mapsto \big\langle \nabla_x^2 c\big(\bar{x}, \exp_{\bar{x}}(p+\eta)\big)\cdot\xi,\xi\big\rangle$$

is concave, and proves (4). □

The above theorem leads to the definition of the **regularity** property:

Definition 4.4. The cost function $c = \frac{d^2}{2}$ is said to be regular if the properties listed in Theorem 4.3 are satisfied.

To understand why the above properties are related to smoothness, consider properties (3) in Theorem 4.3. It says that, if we take the function $\psi_{\bar{x},y_0,y_1} = \max\big\{-c(\cdot,y_0)+c(\bar{x},y_0), -c(\cdot,y_1)+c(\bar{x},y_1)\big\}$, then we are able to touch the graph of this function from below at $\bar{x}$ with the family of functions $\{-c(\cdot,y_t)+c(\bar{x},y_t)\}_{t\in[0,1]}$. This suggests that we could use this family to regularize the cusp of $\psi_{\bar{x},y_0,y_1}$ at the point $\bar{x}$, by slightly moving above the graphs of the functions $-c(\cdot,y_t)+c(\bar{x},y_t)$. (See Figure 1.) On the other hand, if (3) does not holds, it is not clear how to regularize the cusp preserving the condition of being c-convex.

By what we said above, we see that the regularity property seems mandatory to develop a theory of smoothness of optimal transport. Indeed, if it is not satisfied, we can construct C^∞ strictly positive densities f, g such that the optimal map is not continuous. The next natural question is: when is it satisfied?

4.5. **The Ma-Trudinger-Wang tensor.** As we have seen in Theorem 4.3, the regularity of $c = \frac{d^2}{2}$ is equivalent to

$$\frac{d^2}{ds^2}\bigg|_{s=0}\frac{d^2}{dt^2}\bigg|_{t=0} c\big(\exp_x(t\xi), \exp_x(p+s\tilde{\eta})\big) \le 0, \tag{4.2}$$

for all $p, \xi, \tilde{\eta} \in T_xM$, with ξ and $\tilde{\eta}$ orthogonal, $p = (\exp_x)^1(y)$ for some $y \notin \mathrm{cut}(x)$.

Introduce a local system of coordinates $(x^1,\dots,x^n)$ around x, and a system $(y^1,\dots,y^n)$ around y. We want to express the above condition ony in terms of c, using the relation $\nabla_x c(x,y) + (\exp_x)^{-1}(y) = 0$. By the definition of gradient (see Subsection 1.6), this relation is also equivalent to

$$-d_x c(x,y) = \langle(\exp_x)^{-1}(y),\cdot\rangle_x. \tag{4.3}$$

We now start to write everything in coordinates. We will write $c_j = \frac{\partial c}{\partial x^j}$, $c_{jk} = \frac{\partial^2 c}{\partial x^j \partial x^k}$, $c_{i,j} = \frac{\partial^2 c}{\partial x^i \partial y^j}$, and so on; moreover $(c^{i,j})$ will denote the coordinates of the inverse matrix of $(c_{i,j})$. Then (4.3) becomes

$$-c_i \xi^i = g_{jk} p^j \xi^k \qquad \forall \xi \in T_x M.$$

Differentiating at y in a direction $\eta \in T_y M$ we get

$$-c_{i,j} \xi^i \eta^j = g_{ij} [(d_p \exp_x)^{-1}(\eta)]^j \xi^i.$$

Thus we get a formula for $-d^2_{x,y} c : T_x M \times T_y M \to \mathbb{R}$:

$$-c_{i,j}(x,y) = g_{ik}(x) [(d_p \exp_x)^{-1}(\eta)]^k_j, \qquad y = \exp_x(p). \tag{4.4}$$

As shown in [23], $-d^2_{x,y} c$ defines a pseudo-metric on $T_x M \times T_y M$, which coincides with g along the diagonal $\{x = y\}$, and it is possible to interpret the regularity condition for the cost in terms of this pseudo-metric.

1. Rewriting the orthogonality condition

The operator $-d^2_{x,y} c$ can be used to transport tangent vectors at y in cotangent vectors at x, and viceversa. In particular, if we consider the covector $\tilde{\eta}_i := -c_{i,j} \eta^j$ ($\eta \in T_y M$), the orthogonality condition $g_x(\xi, \tilde{\eta}) = 0$ appearing in the definition of the regularity of the cost is equivalent to $\tilde{\eta}_i \xi^i = 0$, i.e.

$$0 = -c_{i,j} \xi^i \eta^j.$$

Thanks to (4.4), we have the formula $\eta = d_p \exp_x(\tilde{\eta})$. (Note: η is **not** the parallel transport of $\tilde{\eta}$ along the geodesic $\exp_x(tp)$!) In particular, if the sectional curvature of the manifold in non-negative everywhere, then the exponential map is 1-Lipschitz, and so $|\eta|_y \le |\tilde{\eta}|_x$.

2. Rewriting the Ma–Trudinger–Wang condition

Equation (4.2) can also be written as

$$\frac{\partial^2}{\partial p_\eta^2} \frac{\partial^2}{\partial x_\xi^2} c(x,y) \le 0. \tag{4.5}$$

The meaning of the left-hand side in (4.5) is the following: first freeze y and differentiate $c(x,y)$ twice with respect to x in the direction $\xi \in T_x M$. Then, considering the result as a function of y, parameterize y by $p = -\nabla_x c(x,y)$, and differentiate twice with respect to p in the direction $\eta \in T_y M$ (see the discussion in the next paragraph). By the relation $p_i = -c_i(x,y)$ we get $\frac{\partial p_i}{\partial y^j} = -c_{i,j}$, which gives $\frac{\partial y^k}{\partial p_\ell} = -c^{k,\ell}$. Then, using $-c_{i,j}$ and $-c^{i,j}$ to raise and lower indices ($\eta^k = -c^{k,l} \eta_i$, etc.), it is just a (tedious) exercise to show that the expression in (4.5) is equal to

$$\sum_{ijklrs} (c_{ij,kl} - c_{ij,r} c^{r,s} c_{s,kl})\, \xi^i \xi^j \eta^k \eta^l,$$

where we used the formula $d(M^{-1}) \cdot H = -M^{-1} H M^{-1}$.

We can now define the **Ma-Trudinger-Wang tensor** (in short **MTW tensor**):

$$\mathfrak{S}_{(x,y)}(\xi, \eta) := \frac{3}{2} \sum_{ijklrs} (c_{ij,r} c^{r,s} c_{s,kl} - c_{ij,kl})\, \xi^i \xi^j \eta^k \eta^l.$$

In terms of this tensor, the regularity condition for the cost functions becomes

$$\mathfrak{S}_{(x,y)}(\xi,\eta) \geq 0 \qquad \text{whenever } (x,y) \in (M\times M)\setminus \mathrm{cut}(M),\ -c_{i,j}\xi^i\eta^j = 0.$$

4.6. **Invariance of $\mathfrak{S}$.** By the computations of the last paragraph, we have seen that $\mathfrak{S}$ is constructed by the expression in (4.5). Since that expression involves second derivatives (which are not intrinsic and depend on the choice of the coordinates), it is not a priori clear whether $\mathfrak{S}$ depends or not on the choice of coordinates. On the other hand, we can hope it does not, because of the (intrinsic) geometric interpretation of the regularity.

To see that $\mathfrak{S}$ is indeed independent of any choice of coordinates (so that one does not even need to use geodesic coordinates, as in (4.2)), we observe that, if we do a change of coordinates and compute first the second derivatives in x, we get some additional terms of the form

$$\Gamma_{ij}^k(x)c_k(x,y) = -\Gamma_{ij}^k(x)p_k(x,y) = \Gamma_{ij}^k(x)g_{k\ell}(x)p^\ell(x,y).$$

But when we differentiate twice with respect to p, this additional term disappears.

4.7. **Relation to curvature.** Let $\xi,\eta \in T_xM$ two orthogonal unit vectors, and consider the functions

$$F(t,s) := \frac{d\big(\exp_x(t\xi),\exp_x(s\eta)\big)^2}{2}.$$

As $\frac{\partial}{\partial s}|_{s=0}F(t,s)$ and $\frac{\partial}{\partial t}|_{t=0}F(t,s)$ identically vanish, we have the Taylor expansion

$$F(t,s) \simeq At^2 + Bs^2 + Ct^4 + Dt^2s^2 + Es^4 + \dots$$

Since $F(t,0) = t^2$ and $F(0,s) = s^2$ we deduce $A = B = 1$ and $C = E = 0$. Hence by (1.3) we recover the identity

$$\mathfrak{S}_{(x,x)}(\xi,\eta) = -\frac{3}{2}\,\frac{\partial^2}{\partial s^2}\bigg|_{s=0}\frac{\partial^2}{\partial t^2}\bigg|_{t=0} F(t,s) = \mathrm{Sect}_x([\xi,\eta]),$$

first proved by Loeper [24]. This fact shows that the MTW tensor is a non-local version of the sectional curvature. In fact, as shown by Kim and McCann [23], $\mathfrak{S}$ is the sectional curvature of the manifold $M\times M$, endowed with the pseudo-metric $-d^2_{xy}c$. Combining the above identity with Theorems 4.1 and 4.3, we get the following important negative result:

Theorem 4.5. *Let (M,g) be a (compact) Riemannian manifolds, and assume that there exist $x\in M$ and a plane $P\subset T_xM$ such that $\mathrm{Sect}_x(P) < 0$. Then there exist C^∞ strictly positive probability densities f and g such that the optimal map is discontinuous.*

After this negative result, one could still hope to develop a regularity theory on any manifold with non-negative sectional curvature. But such is not the case: as shown by Kim [22], the regularity condition is strictly stronger than the condition of nonnegativity of sectional curvatures.

4.8. **The Ma-Trudinger-Wang condition.**

Definition 4.6. We say that (M,g) satisfies the MTW(K) condition if, for all $(x,y)\in(M\times M)\setminus\mathrm{cut}(M)$, for all $\xi\in T_xM$, $\eta\in T_yM$,

$$\mathfrak{S}_{(x,y)}(\xi,\eta) \geq K|\xi|_x^2|\tilde\eta|_x^2 \qquad \text{whenever } -c_{i,j}(x,y)\xi^i\eta^j = 0,$$

where $\tilde\eta^i = -g^{i,k}(x)c_{k,j}(x,y)\eta^j \in T_xM$.

Some example of manifolds satisfying the Ma-Trudinger-Wang condition are given in [24, 25, 23, 15]:

- $\mathbb{R}^n$ and $\mathbb{T}^n$ satisfy MTW(0).
- $\mathbb{S}^n$ and its quotients satisfy MTW(1).
- Products of spheres satisfy MTW(0).

We observe that the MTW condition is a nonstandard curvature condition, as it is fourth order and nonlocal. Therefore an important open problem is whether this condition is stable under pertubation. More precisely, we ask for the following

Question: assume that (M, g) satisfies the MTW(K) condition for $K > 0$, and let g_ε be a C^4-perturbation of g. Does (M, g_ε) satisfy the MTW(K') condition for some $K' > 0$?

The answer to this question is easily seen to be affirmative for manifolds with nonfocal cut-locus like the projective space $\mathbb{RP}^n$ (see [26]). Moreover, as proven by Rifford and the first author [15], the answer is affirmative also for the 2-dimensional sphere $\mathbb{S}^2$.

The next property, called **Convexity of Tangent Injectivity Loci**, or **(CTIL)** in short, is useful to prove regularity and stability results [26, 15]:

Definition 4.7. We say that (M, g) satisfies CTIL if, for all $x \in M$, the set

$$\mathrm{TIL}(x) := \{tv \in T_xM \,|\, 0 \leq t < t_c(x, v)\} \subset T_xM$$

is convex.

As shown by the second author [46], if CTIL is satisfied, then the MTW condition is stable under Gromov–Hausdorff convergence:

Theorem 4.8. *Let (M_k, g_k) be a sequence of Riemannian manifolds converging in the Gromov–Hausdorff topology to a Riemannian manifold (M, g). If (M_k, g_k) satisfy* MTW(0) *and* CTIL*, then also (M, g) satisfies* MTW(0).

The proof of this result uses that, under CTIL, MTW(0) is equivalent to the connectedness of the c-subdifferential of all c-convex functions ψ which solve the dual Kantorovich problem (see Subsection 2.5).

4.9. **Local to global.** Under CTIL, one can prove that the MTW(0) condition is equivalent to the regularity condition (4.1) (see [46]).

Here we want to show that an "improved" MTW condition allows to prove an "improved" regularity condition, which in turns implies (Hölder) continuity of the optimal map.

Definition 4.9. Let $K, C \geq 0$. We say that (M, g) satisfies the MTW(K, C) condition if for all $(x, y) \in (M \times M) \setminus \mathrm{cut}(M)$, for all $\xi \in T_xM$, $\eta \in T_yM$,

$$\mathfrak{S}_{(x,y)}(\xi, \eta) \geq K|\xi|_x^2|\tilde{\eta}|_x^2 - C\big|\langle \xi, \tilde{\eta}\rangle_x\big||\xi|_x|\tilde{\eta}|_x,$$

where $\tilde{\eta}^i = -g^{i,k}(x)c_{k,j}(x, y)\eta^j \in T_xM$.

We observe that the second term appearing in the right hand side vanishes if $-c_{i,j}(x, y)\xi^i\eta^j = 0$. Moreover, by the Cauchy–Schwarz inequality and $\mathfrak{S}_{(x,x)}(\xi, \xi) = 0$, we must have $C \geq K$. Next, if MTW(K, C) holds for some $K > 0$, then the sectional curvatures are bounded from below by K, and so by Bonnet–Myers's Theorem the manifold is compact. We finally remark that, in a subset of $M \times M$ where c is smooth, a compactness argument shows that $\mathrm{MTW}(K)$ implies $\mathrm{MTW}(K, C)$ for some $C > 0$ [26, Lemma 2.3]. So the refinement from $\mathrm{MTW}(K)$ to $\mathrm{MTW}(K, C)$ is interesting only when the cost function loses its smoothness, i.e. close to the cut locus.

Example 4.10. As proved in [15], the sphere $\mathbb{S}^n$ and its quotients satisfy $\mathrm{MTW}(K, K)$ for some $K > 0$, and C^4-perturbations of $\mathbb{S}^2$ satisfy $\mathrm{MTW}(K, C)$ for some $K, C > 0$.

We now show that $\mathrm{MTW}(K, C)$ with $K > 0$ implies an "improved" regularity inequality. For simplicity, here we give a simpler version of the lemma, where we assume that (M, g) satisfies CTIL (otherwise one would need to apply an approximation lemma proved by the authors in [16]). (See Figure 3.)

Lemma 4.11. *Let (M, g) satisfies CTIL and $\mathrm{MTW}(K, C)$ with $K > 0$. For any $\bar{x} \in M$, let $(p_t)_{0 \le t \le 1}$ be a C^2 curve drawn in $\mathrm{TIL}(\bar{x})$, and let $y_t = \exp_{\bar{x}}(p_t)$; let further $x \in M$. If*

$$|\ddot{p}_t|_{\bar{x}} \le \varepsilon_0 \, d(\bar{x}, x) |\dot{p}_t|^2_{\bar{x}}, \tag{4.6}$$

then there exists $\lambda = \lambda(K, C, \varepsilon_0) > 0$ such that, for any $t \in (0, 1)$,

$$\begin{aligned} d(x, y_t)^2 - d(\bar{x}, y_t)^2 \ge \min\Big(d(x, y_0)^2 - d(\bar{x}, y_0)^2, \, d(x, y_1)^2 - d(\bar{x}, y_1)^2 \Big) \\ + 2\lambda t(1-t) d(\bar{x}, x)^2 |p_1 - p_0|^2_{\bar{x}}. \end{aligned} \tag{4.7}$$

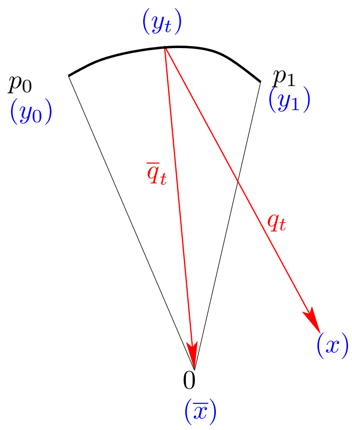

FIGURE 3. Lemma 4.11: p_0, p_1 are tangent vectors at $\bar{x}$; q_t, $\bar{q}_t$ are tangent at $y_t = \exp_{\bar{x}} p_t$.

This result, first proved by Loeper and the second author [26], and then sligthly modified by Rifford and the first author [15], is a refinement of the proof given by Kim and McCann [23] for the implication (4) $\Rightarrow$ (3) in Theorem 4.3.

Sketch of the proof. Define the function

$$h(t) := -c(x, y_t) + c(\bar{x}, y_t) + \delta t(1-t),$$

with $c = \frac{d^2}{2}$. We want to prove that, for $\delta = \lambda d(\bar{x}, x)^2 |p_1 - p_0|^2_{\bar{x}}$, if λ is small enough then $h(t) \leq \max\bigl(h(0), h(1)\bigr)$. The idea of the proof is by the maximum principle: if we show that $\ddot{h} > 0$ whenever $\dot{h} = 0$, this will imply the result.

Define $q_t := (\exp_{y_t})^{-1}(x)$, $\bar{q}_t := (\exp_{y_t})^{-1}(\bar{x})$, $\eta := q_t - \bar{q}_t$. Then, since

$$\dot{y}_t^i = c^{i,j}\dot{p}_j,$$

$$\ddot{y}_t^i = -c^{i,k}c_{k,\ell j}c^{\ell,r}c^{j,s}\dot{p}_r\dot{p}_s - c^{i,r}\ddot{p}_r,$$

(everything being evaluated at $(\bar{x}, y_t)$), after some computations one obtains

$$\dot{h}(t) = -c_{i,j}(\bar{x}, y_t)\eta^i\dot{y}_t^j + \delta(1-2t),$$

$$\ddot{h}(t) = -\bigl([c_{,ij}(x, y_t) - c_{,ij}(\bar{x}, y_t)] - \eta^k c_{k,ij}(\bar{x}, y_t)\bigr)\dot{y}_t^i\dot{y}_t^j + c_{i,j}\eta^i c^{j,r}\ddot{p}_r - 2\delta. \tag{4.8}$$

We now observe that the first term in the right hand side can be written as

$$\Phi(q_t) - \Phi(\bar{q}_t) - d_{\bar{q}_t}\Phi\cdot(q_t - \bar{q}_t),$$

with $\Phi(q) := c_{,ij}\bigl(\exp_{y_t}(q), y_t\bigr)\dot{y}_t^i\dot{y}_t^j$; therefore it is equal to

$$\int_0^1 \frac{d^2}{ds^2}\Phi(sq_t + (1-s)\bar{q}_t)\,ds = -\frac{2}{3}\int_0^1 \mathfrak{S}_{(y_t, x_s)}(\dot{y}_t, \eta)\,ds$$

with $x_s := \exp_{y_t}\bigl(sq_t + (1-s)\bar{q}_t\bigr)$, and we get

$$\begin{aligned}\ddot{h}(t) &= \frac{2}{3}\int_0^1 \mathfrak{S}_{(y_t, x_s)}(\dot{y}_t, \eta)\,ds + c_{i,j}\eta^i c^{j,r}\ddot{p}_r - 2\delta \\ &\geq \frac{2}{3}K|\tilde{\eta}|^2_{y_t}|\dot{y}_t|^2_{y_t} - \frac{2}{3}C\bigl|\langle\tilde{\eta}, \dot{y}_t\rangle_{y_t}\bigr||\dot{y}_t|_{y_t}|\tilde{\eta}|_{y_t} + c_{i,j}(\bar{x}, y_t)\eta^i c^{j,r}\ddot{p}_r - 2\delta,\end{aligned}$$

where $\tilde{\eta} := (d_{p_t}\exp_{y_t})^{-1}(\eta) = (d_{p_t}\exp_{y_t})^{-1}(q_t - \bar{q}_t)$. To understand now why the result is true, we remark that for $\delta = 0$ the condition $\dot{h} = 0$ means $\langle\tilde{\eta}, \dot{y}_t\rangle_{y_t} = -c_{i,j}(\bar{x}, y_t)\eta^i\dot{y}_t^j = 0$, which gives

$$\ddot{h}(t) \geq \frac{2}{3}K|\tilde{\eta}|^2_{y_t}|\dot{y}_t|^2_{y_t} + c_{i,j}\eta^i c^{j,r}\ddot{p}_r,$$

and thanks to the assumption of smallness on $|\ddot{p}|_{\bar{x}}$ one gets $\ddot{h} > 0$.

In the general case $\delta > 0$ small, the condition $\dot{h} = 0$ gives $|c_{i,j}(\bar{x}, y_t)\eta^i\dot{y}_t^j| \leq \delta$, and using that $|\tilde{\eta}|_{\bar{x}} \geq |\eta|_{y_t} \geq d(x, \bar{x})$ (since the sectional curvature of the manifold is non-negative) one obtains the desired result. $\square$

Remark 4.12. This local-to-global argument can also be used to give a differential characterization of c-convex functions (see Proposition 2.5). More precisely one has: assume that (M, g) satisfies MTW(0) and CTIL. Then $\psi \in C^2(M)$ is c-convex if and only if

$$\nabla^2\psi(x) + \frac{\nabla_x^2 d\bigl(x, \exp_x(\nabla\psi(x))\bigr)^2}{2} \geq 0 \qquad \forall\, x \in M.$$

4.10. **A smoothness result.**

Theorem 4.13. *Let (M, g) be a (compact) Riemannian manifold satifying* $\mathrm{MTW}(K, C)$ *with* $K > 0$*. Assume morever that all* $\mathrm{TIL}(x)$ *are uniformly convex, and let f and g be two probability densities on M such that $f \le A$ and $g \ge a$ for some $A, a > 0$. Then the optimal map between $\mu = f\,\mathrm{vol}$ and $\nu = g\,\mathrm{vol}$, with cost function $c = d^2/2$, is continuous.*

As shown by Loeper [24], this theorem can be refined into a Hölder regularity for the transport map. The argument of the proof, originally due to Loeper, has been simplified first by Kim and McCann [23], and then by Loeper and the second author [26]. The argument we present is borrowed from [26].

Proof. By Theorem 2.2, we know that the optimal map T can be written as $\exp_x(\nabla\psi(x))$, so it suffices to prove that ψ is C^1. Since ψ is semiconvex, we need to show that the subgradient $\nabla^-\psi(x)$ is a singleton for all $x \in M$. The proof is by contradiction.

Assume that there is $\bar{x} \in M$ and $p_0, p_1 \in \nabla^-\psi(\bar{x})$. Let $y_0 = \exp_{\bar{x}} p_0$, $y_1 = \exp_{\bar{x}} p_1$. Since the cost is regular, we have $y_i \in \partial^c\psi(\bar{x})$ for $i = 0, 1$, that is

$$\psi(\bar{x}) + c(\bar{x}, y_i) = \min_{x\in M}\Big[\psi(x) + c(x, y_i)\Big], \qquad i = 0, 1.$$

In particular

$$c(x, y_i) - c(\bar{x}, y_i) \ge \psi(\bar{x}) - \psi(x), \qquad i = 0, 1. \tag{4.9}$$

For $\varepsilon \in (0, 1)$, we define $D_\varepsilon \subset \overline{\mathrm{TIL}}(\bar{x})$ as follows: D_ε consists of the set of points $p \in T^*_{\bar{x}}M$ such that there exists a path $(p_t)_{0\le t\le 1} \subset \overline{\mathrm{TIL}}(\bar{x})$ from p_0 to p_1 such that, if we define $y_t = \pi_1 \circ \phi_1^H(\bar{x}, p_t)$, we have $\ddot{p}_t = 0$ for $t \notin [1/4, 3/4]$, $|\ddot{p}_t|_{y_t} \le \varepsilon\eta_0|\dot{y}_t|^2_{y_t}$ for $t \in [1/4, 3/4]$, and $p = p_t$ for some $t \in [1/4, 3/4]$ (this is like a "sausage", see Figure 4).

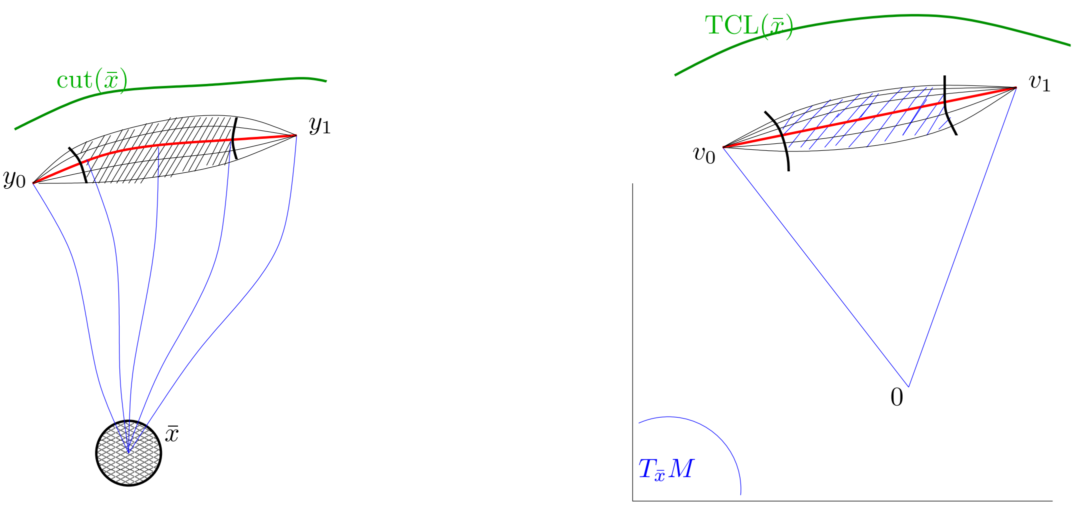

FIGURE 4. Proof of Theorem 4.13: the volume of the ball $B(\bar{x}, \varepsilon)$ is much smaller than the volume of the "sausage" with base $\bar{x}$, endpoints y_0, y_1 and width $O(\varepsilon)$; $\mathrm{TCL}(\bar{x})$ is the tangent cut locus at $\bar{x}$.

By uniform convexity of $\overline{\mathrm{TIL}}(\bar{x})$, if η_0 is sufficiently small then D_ε lies a positive distance σ away from the tangent cut locus $\mathrm{TCL}(\bar{x}) = \partial\big(\mathrm{TIL}(\bar{x})\big)$, with $\sigma \sim |p_0 - p_1|_{\bar{x}}^2$. Thus all paths $(p_t)_{0\le t\le 1}$ used in the definition of D_ε satisfy

$$|\dot{y}_t|_{y_t} \ge c|p_0 - p_1|_{\bar{x}} \qquad \forall\, t \in [1/4, 3/4].$$

Moreover condition (4.6) is satisfied if $\eta_0 \le \varepsilon_0$ and $d(\bar{x}, x) \ge \varepsilon$. By simple geometric consideration, we see that D_ε contains a parallelepiped E_ε centered at $(p_0 + p_1)/2$ with one side of length $\sim |p_0 - p_1|_{\bar{x}}$, and the other sides of length $\sim \varepsilon|p_0 - p_1|_{\bar{x}}^2$, such that all points y in this parallelepiped can be written as y_t for some $t \in [1/3, 2/3]$, with y_t as in the definition of D_ε. Therefore

$$\mathscr{L}^n(E_\varepsilon) \ge c(M, \eta_0, |p_0 - p_1|_{\bar{x}})\varepsilon^{n-1},$$

with $\mathscr{L}^n$ denoting the Lebesgue measure on $T_{\bar{x}}M$. Since E_ε lies a positive distance from $\partial\big(\mathrm{TIL}(\bar{x})\big)$, we obtain

$$\mathrm{vol}(Y_\varepsilon) \sim \mathscr{L}^n(E_\varepsilon) \ge c(M, \eta_0, |p_0 - p_1|_{\bar{x}})\varepsilon^{n-1}, \qquad Y_\varepsilon := \exp_{\bar{x}}(E_\varepsilon).$$

We wish to apply Theorem 4.13 to the paths $(p_t)_{0\le t\le 1}$ used in the definition of D_ε. Since p_0, p_1 belong to $\overline{\mathrm{TIL}}(\bar{x})$ but not necessarily to $\mathrm{TIL}(\bar{x})$, we first apply the theorem with $(\theta p_t)_{0\le t\le 1}$ with $\theta < 1$, and then we let $\theta \to 1$; in the end, for any $y \in Y_\varepsilon$ and $x \in M \setminus B_\varepsilon(\bar{x})$,

$$d(x,y)^2 - d(\bar{x},y)^2 \ge \min\Big(d(x,y_0)^2 - d(\bar{x},y_0)^2,\, d(x,y_1)^2 - d(\bar{x},y_1)^2\Big) + \lambda\varepsilon^2|p_0 - p_1|_{\bar{x}}^2,$$

for some $\lambda > 0$. Combining this inequality with (4.9), we conclude that

$$\text{for any } y \in Y_\varepsilon, \qquad y \notin \partial^c\psi(x) \qquad \forall\, x \in M \setminus B_\varepsilon(\bar{x}).$$

This implies that all the mass brought into Y_ε by the optimal map comes from $B_\varepsilon(\bar{x})$, and so

$$\mu(B_\varepsilon(\bar{x})) \ge \nu(Y_\varepsilon).$$

Since $\mu(B_\varepsilon(\bar{x})) \le A\mathrm{vol}(B_\varepsilon(\bar{x})) \sim \varepsilon^n$ and $\nu(Y_\varepsilon) \ge a\mathrm{vol}(Y_\varepsilon) \gtrsim \varepsilon^{n-1}$, we obtain a contradiction as $\varepsilon \to 0$. ☐

5. Recap and perspectives

In these notes, we have seen two different connections of optimal transport to curvature:

1. Ricci curvature: the optimal transport is a way to give a syntethic formulation of lower Ricci curvature bounds
2. Sectional curvature: a regularity theory for optimal transport on a manifold depends on the MTW condition, which reinforces non-negative sectional curvature.

We remark that, in both cases, the optimal transport goes well with lower bounds only.

A good thing is that in both cases there is a "soft" reformulation in terms of optimal transport:

$$\mathrm{Ric} \ge 0 \quad \Longleftrightarrow \quad t \mapsto H(\mu_t) \text{ is convex, } \forall\,(\mu_t)_{0\le t\le 1} \text{ geodesic in } P_2(M);$$

$$\mathfrak{S} \ge 0 \quad \Longleftrightarrow \quad \partial^c\psi(x) \text{ is connected, } \forall\,\psi \text{ solution of the dual Kantorovich problem.}$$

Observe that these reformulations have the advantage of being very stable, and at the same time can be used to generalize some differential concepts out of the Riemannian setting.

5.1. **The curvature-dimension condition.** As we already said, while the sectional curvature gives a pointwise control on distances, the Ricci curvature gives an averaged control, and is related to Jacobian estimates with respect to a reference measure (which in Section 3 was always the volume measure). For this reason, a natural general setting where one can study Ricci bounds is the one of measured metric spaces (X, d, ν) (see Subsection 1.15).

As an example, consider the measured metric spaces $(M^n, g, e^{-V}\mathrm{vol})$, with $V \in C^2(M)$. Modify the classical Ricci tensor into $\mathrm{Ric}_{N,\nu} := \mathrm{Ric} + \nabla^2 V - \frac{\nabla V \otimes \nabla V}{N-n}$ for $N \geq n$ (where by convention $\frac{0}{0} = 0$, and N plays the role of an "effective" dimension). Then the **curvature-dimension** condition $\mathrm{CD}(K, N)$, classically used in probability theory and geometry, consists in $\mathrm{Ric}_{N,\nu} \geq K$. Exactly as $\mathrm{Ric} \geq K$, also this more general condition can be reformulated in terms of optimal transport. Up to minor variations, the following definition was introduced independently by Sturm [38, 39] and by Lott and the second author [28, 29] (recall the definition of H from Subsection 3.4):

Definition 5.1. A compact measured metric space (X, d, ν) is said to satisfy $\mathrm{CD}(K, \infty)$ if, for all $\mu_0, \mu_1 \in P_2(X)$, there exists a Wasserstein geodesic $(\mu_t)_{0 \leq t \leq 1}$ between μ_0 and μ_1 such that

$$H_\nu(\mu_t) \leq (1-t) H_\nu(\mu_0) + t H_\nu(\mu_1) - K \frac{t(1-t)}{2} W_2(\mu_0, \mu_1)^2 \qquad \forall t \in [0, 1].$$

A similar definition for $\mathrm{CD}(0, N)$ is obtained by choosing $K = 0$ and replacing the nonlinearity $r \log r$ by $-r^{1-1/N}$. There is also a more complicated definition which works for the general $\mathrm{CD}(K, N)$ criterion [39, 45].

Example 5.2. $(\mathbb{R}^n, |\cdot|, e^{-|x|^2/2} dx)$ satisfies $\mathrm{CD}(1, \infty)$; $(\mathbb{R}^n, \|\cdot\|, dx)$ satisfies $\mathrm{CD}(0, N)$ for any norm $\|\cdot\|$.

Ohta [33] recently performed some exploration of Finsler geometry along these lines.

5.2. **Open problem: locality.** A natural question is whether the above definition of $\mathrm{Ric}_{\infty,\nu} \geq K$ is local or not. As shown in [38, 39] and [45, Chapter 30], this question has an affirmative answer if geodesics are non-branching, but it is open in the general case. The answer would however be affirmative if the following (elementary but tricky) conjecture from [45] were true:

Conjecture 5.3. Let $0 < \theta < 1$, $0 \leq \alpha \leq \pi$, and assume that $f : [0, 1] \to \mathbb{R}^+$ satisfies

$$f\big((1-\lambda)t + \lambda t'\big) \geq (1-\lambda) \left(\frac{\sin\big((1-\lambda)\alpha|t-t'|\big)}{(1-\lambda)\sin\big(\alpha|t-t'|\big)} \right)^\theta f(t) + \lambda \left(\frac{\sin\big(\lambda\alpha|t-t'|\big)}{\lambda \sin\big(\alpha|t-t'|\big)} \right)^\theta f(t')$$

for $|t - t'|$ small. Then the above inequality holds true for all $t, t' \in [0, 1]$.

5.3. **Ricci and diffusion equations.** The non-negativity of the Ricci curvature is important to get contraction properties of solutions of the heat equation on a manifold. Indeed the following result holds (see [35, 36, 37]):

Theorem 5.4. (M, g) *satisfies* $\mathrm{Ric} \geq 0$ *if and only if, for all* μ_t *and* $\tilde{\mu}_t$ *solutions of the heat equation,* $W_2(\mu_t, \tilde{\mu}_t)$ *is non-increasing in time.*

Another way to state this theorem is to say that $\mathrm{Ric} \geq 0$ if and only if $(e^{t\Delta})_{t \geq 0}$ is a contraction in $P_2(M)$.

Recalling that Ric ≥ 0 is equivalent to the convexity of the entropy functional $H(\mu_t)$ (see Subsection 3.4), the above equivalence if formally explained by the Jordan-Kinderlehrer-Otto theorem [18]: *the heat flow is the gradient flow of H in $P_2(M)$.* A generalization of the above result has been given McCann and Topping [31] (see also [27]):

Theorem 5.5. *A family of Riemannian metric $g(t)$ are super-solutions of the backward Ricci-flow $\frac{\partial g_t}{\partial t} \leq 2\mathrm{Ric}_{g_t}$ if and only if, for all μ_t and $\tilde{\mu}_t$ solutions of the heat equation in (M, g_t) (i.e. $\partial_t \mu_t = \Delta_{g_t} \mu_t$), $W_2(\mu_t, \tilde{\mu}_t)$ is non-increasing in time.*

Let us also notice that Lott [27] and Topping [40] have recently studied properties of the Ricci flow with the help of the optimal transport, and they can for instance recover Perelman's monotonicity formula.

5.4. **Discretization.** A natural question in probability theory and statistical mechanics is how to define a notion of curvature on discrete spaces. The optimal transport allows to answer to this question: the idea is to "discretize the synthetic formulation":

- replace length space by δ-length space, etc.
- allow for errors, either in the heat formulation (Markov kernels, etc.) as done by Ollivier [34] and Joulin [19], or in the optimal transport formulation as done by Bonciocat and Sturm [2].

Example 5.6. Consider the metric space $\{0, 1\}^N$ endowed with the Hamming metric (i.e. each edge is of length 1). Then Ric $\gtrsim \frac{1}{N}$ at scale $O(1)$ (see [34, Example 8]).

5.5. **Smoothness.** Regarding the smoothness of the optimal transport map, two main questions arise:

- Find further examples of manifolds satisfying the MTW conditions. (Recall $\mathbb{S}^n$ and its quotients, $\mathbb{RP}^n$ and its perturbations, perturbations of $\mathbb{S}^2$, products of spheres.)
- Find results yielding regularity in terms of the MTW condition, the shape of the cut locus, assumptions on μ and ν...

As an example, the following theorem was proven by Loeper and the second author [26]:

Theorem 5.7. *Let (M, g) be a compact Riemannian manifold such that there is no focalization at the cut locus (i.e., $d_{t_c(x,v)v} \exp_x$ is invertible for all x, v). Assume that (M, g) satisfies* MTW(K) *for some $K > 0$, and let f and g be two probability densities on M such that $f \leq A$ and $g \geq a$ for some $A, a > 0$. Then the optimal map T is $C^{0,\alpha}$, with $\alpha = \frac{1}{4n-1}$.*

Examples of manifold satifying the assumptions of the above theorem are the projective space $\mathbb{RP}^n$ and its perturbations, and a challenge is to understand what happens when one removes the "no-focalization" assumption.

The following conjecture was formulated in [26]:

Conjecture 5.8. MTW implies CTIL.

This conjecture has been proved by Loeper and the second author [26] under MTW(K) with $K > 0$, and the no-focalization assumption. Using a variant of the MTW condition, Rifford and the first author proved the following result [15]:

Theorem 5.9. *If (M, g) is a C^4-perturbation of $\mathbb{S}^2$, then* CTIL *holds.*

Through these considerations we see that the MTW condition, originally introduced as a way to explore the regularity of optimal transport, turns out to give a new kind of geometric information on the manifold (compare with the Bonnet–Myers theorem, or with classical theorems on the rectifiability or local description of the cut locus).

6. Selected references

This section provides a list of references which seem to us the most significant. To this list should be added the book [45], which is attempts at providing a synthetic overview of all the links between geometry and optimal transport.

6.1. **Links between optimal transport and Ricci curvature.** The reference [36] by Otto–Villani may be considered as the founding paper for this topic. A body of technical tools was developed by Cordero-Erausquin–McCann–Schmuckenschläger [9] to make progress on these issues, and solve open problems stated in [36]. At that time the emphasis was rather on applications from calculus of variations and functional inequalities.

Later the interest of these links for geometric applications was realized, and explicitly noted by von Renesse–Sturm [37]. The synthetic theory of Ricci curvature bounds in the general setting of metric-measure spaces was developed independently by Sturm [38, 39] and by Lott–Villani [28].

6.2. **Optimal transport and Ricci flow.** The links between these objects were suspected for some time, and hinted for by a preliminary result by McCann–Topping [31]. Finally this link was made precise in contributions of Lott [27] and Topping [40].

6.3. **Discrete Ricci curvature.** This theory is in construction with preliminary works by Joulin and Ollivier; one may consult in particular [34] by the latter.

6.4. **Smoothness of optimal transport, cut locus and MTW tensor.** After the landmark papers by Ma–Trudinger–Wang [30] and Loeper [24], the theory was partly simplified and rewritten by Kim–McCann [23]. Applications to the geometry of the cut locus were first investigated by Loeper–Villani [26], and further developed by Figalli–Rifford [15].

References

[1] L. Ambrosio and P. Tilli: Topics on analysis in metric spaces. *Oxford Lecture Series in Mathematics and its Applications, 25.* Oxford University Press, Oxford, 2004

[2] A.-I. Bonciocat and K.-T. Sturm: Mass transportation and rough curvature bounds for discrete spaces. *J. Funct. Anal. 256 (2009), no. 9,* 2944–2966 .

[3] Y. Brenier: Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math. 44 (1991), no. 4,* 375–417.

[4] D. Burago, Y. Burago and S. Ivanov: A course in metric geometry. *Graduate Studies in Mathematics, 33.* American Mathematical Society, Providence, RI, 2001.

[5] L. A. Caffarelli: The regularity of mappings with a convex potential. *J. Amer. Math. Soc. 5 (1992), no. 1,* 99–104.

[6] L. A. Caffarelli: Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math. 45 (1992), no. 9,* 1141–1151.

[7] L. A. Caffarelli: Boundary regularity of maps with convex potentials. II. *Ann. of Math. (2) 144 (1996), no. 3,* 453–496.

[8] D. Cordero-Erausquin: Sur le transport de mesures priodiques. *C. R. Acad. Sci. Paris Sèr. I Math. 329 (1999), no. 3,* 199–202.

[9] D. Cordero-Erasquin, R. J. McCann and M.Schmuckenschlager: A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math. 146 (2001), no. 2,* 219-257.

[10] P. Delanoë: Classical solvability in dimension two of the second boundary-value problem associated with the Monge–Ampère operator. *Ann. Inst. H. Poincar Anal. Non Linaire 8 (1991), no. 5,* 443–457.

[11] P. Delanoë and Y. Ge: Regularity of optimal transportation maps on compact, locally nearly spherical, manifolds. *J. Reine Angew. Math.,* to appear.

[12] A. Fathi and A. Figalli: Optimal transportation on non-compact manifolds. *Israel J. Math., 175 (2010), no. 1,* 1–59.

[13] A. Figalli, Y. H. Kim and R. J. McCann: Continuity and injectivity of optimal maps for non-negatively cross-curved costs. Preprint, 2009.

[14] A. Figalli and G. Loeper: C^1 regularity of solutions of the Monge–Ampère equation for optimal transport in dimension two. *Calc. Var. Partial Differential Equations, 35 (2009), no. 4,* 537–550.

[15] A. Figalli and L. Rifford: Continuity of optimal transport maps on small deformations of $\mathbb{S}^2$. *Comm. Pure Appl. Math., 62 (2009), no. 12,* 1670–1706.

[16] A. Figalli and C. Villani: An approximation lemma about the cut locus, with applications in optimal transport theory. *Methods Appl. Anal., 15 (2008), no. 2,* 149–154.

[17] W. Gangbo and R.J. McCann: The geometry of optimal transportation. *Acta Math. 177 (1996), no. 2,* 113–161.

[18] R. Jordan, D. Kinderlehrer and F. Otto: The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal. 29 (1998), no. 1,* 1–17.

[19] A. Joulin: A new Poisson-type deviation inequality for Markov jump process with positive Wasserstein curvature. *Bernoulli 15 (2009), no. 2,* 532–549.

[20] L. V. Kantorovich: On mass transportation. Reprinted from *C. R. (Doklady) Acad. Sci. URSS (N.S.) 37 (1942), no. 7-8.*

[21] L. V. Kantorovich: On a problem of Monge. Reprinted from *C. R. (Doklady) Acad. Sci. URSS (N.S.) 3 (1948), no. 2.*

[22] Y. H. Kim: Counterexamples to continuity of optimal transportation on positively curved Riemannian manifolds. *Int. Math. Res. Not. IMRN 2008,* Art. ID rnn120, 15 pp.

[23] Y. H. Kim and R. J. McCann: Continuity, curvature, and the general covariance of optimal transportation. *J. Eur. Math. Soc.*, to appear.

[24] G. Loeper: On the regularity of solutions of optimal transportation problems. *Acta Math., 202 (2009), no. 2,* 241–283.

[25] G. Loeper: Regularity of optimal maps on the sphere: The quadratic cost and the reflector antenna. *Arch. Ration. Mech. Anal.,* to appear.

[26] G. Loeper and C. Villani: Regularity of optimal transport in curved geometry: the nonfocal case. *Duke Matk. J., 151 (2010), no. 3,* 431–485.

[27] J. Lott: Optimal transport and Perelman's reduced volume. *Calc. Var. Partial Differential Equations 36 (2009), no. 1,* 49–84.

[28] J. Lott and C. Villani: Ricci curvature via optimal transport. *Ann. of Math. 169 (2009),* 903–991

[29] J. Lott and C. Villani: Weak curvature conditions and functional inequalities. *J. Funct. Anal. 245 (2007), no. 1,* 311–333.

[30] X. N. Ma, N. S. Trudinger and X. J. Wang: Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal., 177 (2005), no. 2,* 151–183.

[31] R. J. McCann and P. Topping: Ricci flow, entropy and optimal transportation. *Amer. J. Math.,* to appear.

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

[33] S.-I. Ohta: Finsler interpolation inequalities. *Calc. Var. Partial Differential Equations 36 (2009),* 211–249.

[34] Y. Ollivier: Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal. 256 (2009), no. 3,* 810–864.

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

[36] 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.

[37] M.-K. von Renesse and K.-T. Sturm: Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math. 58 (2005), no. 7,* 923–940.

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

[39] K.-T. Sturm: On the geometry of metric measure spaces. II. *Acta Math. 196 (2006), no. 1,* 133–177.

[40] P. Topping: L-optimal transportation for Ricci flow. *J. Reine Angew. Math.,* to appear.

[41] N. Trudinger and X. J. Wang: On the second boundary value problem for Monge-Ampère type equations and optimal transportation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) 8 (2009), no. 1,* 143–174.

[42] N. S. Trudinger and X. J. Wang: On strict convexity and continuous differentiability of potential functions in optimal transportation. *Arch. Ration. Mech. Anal. 192 (2009), no. 3,* 403–418.

[43] J. I. E. Urbas: Regularity of generalized solutions of Monge–Ampére equations. *Math. Z., 197 (1988), no. 3,* 365–393.

[44] J. I. E. Urbas: On the second boundary value problem for equations of Monge–Ampère type. *J. Reine Angew. Math., 487 (1997),* 115–124.

[45] C. Villani: Optimal transport, old and new. *Grundlehren des mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 338,* Springer-Verlag, Berlin-New York, 2009.

[46] C. Villani: Stability of a 4th-order curvature condition arising in optimal transport theory. *J. Funct. Anal. 255 (2008), no. 9,* 2683–2708.