Optimal transport and nonsmooth geometry

John Lott

http://www.math.lsa.umich.edu/~lott

Joint work with Cédric Villani.

Related work done independently by K.-T. Sturm.

Ideas from optimal transport $\Rightarrow$ Nonsmooth geometry

Ideas from nonsmooth geometry $\Rightarrow$ Optimal transport

Some basics of differential geometry

M a smooth n-dimensional manifold with a Riemannian metric g

m a point in M

T_mM the tangent space at m.

Sectional curvature : To each 2-plane $P \subset T_mM$, one assigns a number $K(P)$, its sectional curvature.

Ricci curvature : an averaging of sectional curvature.

Fix a unit-length vector $\mathbf{v} \in T_m M$.

Definition :

Ric$(\mathbf{v}, \mathbf{v})$ = (n−1)·(the average sectional curvature

of the 2-planes P containing $\mathbf{v}$).

Fact : Ric$(\mathbf{v}, \mathbf{v})$ extends to a symmetric bilinear form on $T_m M$, called the Ricci tensor.

Given $K \in \mathbb{R}$, we say that M has **Ricci curvature bounded below by K** if for all $m \in M$ and all $\mathbf{v} \in T_m M$,

$$\text{Ric}(\mathbf{v}, \mathbf{v}) \geq \text{K}\, \text{g}(\mathbf{v}, \mathbf{v}).$$

Question : Does it make sense to say that a metric space (X, d) has "Ricci curvature bounded below by K"?

1. For simplicity, take $K = 0$.
2. For more simplicity, assume X is compact.

To get started, assume that X is a **length space**, meaning that for all $x_0, x_1 \in X$,

$$d(x_0, x_1) = \inf_{\gamma} L(\gamma), \text{ where}$$

$\gamma : [0, 1] \to X$ continuous , $\gamma(0) = x_0, \gamma(1) = x_1$ and

$$L(\gamma) = \sup_J \sup_{0=t_0 \le t_1 \le \ldots \le t_J = 1} \sum_{j=1}^{J} d(\gamma(t_{j-1}), \gamma(t_j)).$$

Note : Any Riemannian manifold (M, g) has a length space structure on the set of points M.

Empirical observation (Fukaya, Cheeger-Colding)
When dealing with Ricci curvature, it's better to consider "measured length spaces" (X, d, ν).

Here ν is a Borel probability measure on X.

If (M, g) is a compact Riemannian manifold, canonical choice is $\nu = \frac{\mathrm{dvol}}{\mathrm{vol(M)}}$.

Rephrased Question : Is there a good notion of a measured length space (X, d, ν) having "nonnegative Ricci curvature"?

Rules of the game :

1. If $(X, d, \nu) = \left(M, g, \frac{dvol}{vol(M)}\right)$, should get back classical notion.

2. If $\{(X_i, d_i, \nu_i)\}_{i=1}^{\infty}$ have "nonnegative Ricci curvature" and $\lim_{i\to\infty}(X_i, d_i, \nu_i) = (X, d, \nu)$ then (X, d, ν) should have "nonnegative Ricci curvature".

3. Want nontrivial consequences.

Gromov-Hausdorff (GH) topology :

"Two metric spaces are GH-close if Mr. Magoo can't tell them apart."

Definition : $\lim_{i\to\infty}(X_i, d_i) = (X, d)$ if there are maps $f_i : X_i \to X$ and a sequence $\epsilon_i \to 0$ such that

1. (Almost isometry) For all $x_i, x_i' \in X_i$,

$$|d_X(f_i(x_i), f_i(x_i')) - d_{X_i}(x_i, x_i')| \le \epsilon_i.$$

2. (Almost surjective) For all $x \in X$ and all i, there is some $x_i \in X_i$ such that

$$d_X(f_i(x_i), x) \le \epsilon_i.$$

Note : X_i and X don't have to look much alike.

Fact : If each X_i is a length space, so is X.

Measured Gromov-Hausdorff (MGH) topology :

Definition. $\lim_{i\to\infty}(X_i, d_i, \nu_i) = (X, d, \nu)$ *if*

1. $\lim_{i\to\infty}(X_i, d_i) = (X, d)$ *in the GH topology, by means of Borel approximants* $f_i : X_i \to X$,

and

2. $\lim_{i\to\infty}(f_i)_*\nu_i = \nu$ *in the weak-*$*$ *topology.*

Historical background :

For sectional curvature, there's a notion of a length space having "nonnegative Alexandrov curvature".

Properties :

1. If $(X, d) = (M, g)$, get back classical notion of nonnegative sectional curvature.

2. If $\{(X_i, d_i)\}_{i=1}^{\infty}$ have nonnegative Alexandrov curvature and $\lim_{i\to\infty}(X_i, d_i) = (X, d)$ in the GH topology then (X, d) has nonnegative Alexandrov curvature.

3. Nontrivial consequences.

Obvious question : Is there something like this for Ricci curvature?

Another motivation : Gromov precompactness theorem

Given $N \in \mathbb{Z}^+$ and $D > 0$, have precompactness of

$$\left\{ \left(M, g, \frac{dvol}{vol(M)} \right) \right\}$$

in the MGH topology, where M ranges over Riemannian manifolds with
1. $\dim(M) \leq N$,
2. diam(M) $\leq$ D and
3. Ric(M) $\geq$ 0.

What are the limit spaces?
Generally not manifolds, but should have "nonnegative Ricci curvature".
What are the smooth limit spaces? (Will answer)

Optimal transport

For Riemannian manifolds, Otto-Villani and Cordoro-Erausquin-McCann-Schmuckenschläger showed that nonnegative Ricci curvature has something to do with "displacement convexity" of certain functions on the Wasserstein space.

Idea of the sequel : To X is canonically associated its Wasserstein space. Instead of looking at the geometry of X directly, look at the properties of its Wasserstein space.

Plan :

1. Look at certain "entropy" functions on the Wasserstein space.

2. Consider optimal transport on general length spaces.

3. Show that "convexity" of these entropy functions on the Wasserstein space gives a good notion of "nonnegative Ricci curvature".

Notation

X a compact Hausdorff space.

$P(X)$ = Borel probability measures on X, with weak-$*$ topology. Also a compact Haudorff space.

$U : [0, \infty) \to \mathbb{R}$ a continuous convex function with $U(0) = 0$.

Fix a background measure $\nu \in P(X)$.

"Negative entropy" of μ with respect to ν :

$$U_\nu(\mu) = \int_X U(\rho(x))\, d\nu(x) + U'(\infty)\, \mu_s(X).$$

Here

$$\mu = \rho\, \nu + \mu_s$$

is the Lebesgue decomposition of μ w.r.t. ν and

$$U'(\infty) = \lim_{r \to \infty} \frac{U(r)}{r}.$$

$U_\nu(\mu)$ measures <u>nonuniformity</u> of μ w.r.t. ν. Minimized when $\mu = \nu$.

Proposition. *a. $U_\nu(\mu)$ is lower-semicontinuous with respect to $(\mu, \nu) \in P(X) \times P(X)$.*
b. $U_{f_*\nu}(f_*\mu) \leq U_\nu(\mu)$.

Effective dimension

$N \in [1, \infty]$ a new parameter (possibly infinite).

It turns out that there's not a single notion of "nonnegative Ricci curvature", but rather a 1-parameter family.

That is, for each N, there's a notion of a space having "nonnegative N-Ricci curvature".

Here N is an <u>effective dimension</u> of the space, and must be inputted.

Displacement convexity classes

Definition. *(McCann) If* $N < \infty$ *then* DC_N *is the set of such convex functions* U *so that the function*

$$\lambda \to \lambda^N \, U(\lambda^{-N})$$

is convex on $(0, \infty)$*.*

Definition. DC_∞ *is the set of such convex functions* U *so that the function*

$$\lambda \to e^\lambda \, U(e^{-\lambda})$$

is convex on $(-\infty, \infty)$*.*

Example

$$U_N(r) \; = \; \begin{cases} Nr(1 - r^{-1/N}) & \text{if } 1 < N < \infty, \\ r \log r & \text{if } N = \infty. \end{cases}$$

(If $U = U_\infty$ then corresponding functional is

$$U_\nu(\mu) = \begin{cases} \int_X \rho \, \log \rho \, d\nu & \text{if } \mu \text{ is a.c. w.r.t. } \nu, \\ \infty & \text{otherwise.)} \end{cases}$$

Notions from optimal transport

(X, d) a compact metric space.

$$W_2(\mu_0, \mu_1)^2 = \inf \left\{ \int_{X \times X} d(x_0, x_1)^2 \, d\pi(x_0, x_1) \right\},$$

where

$$\pi \in P(X \times X), (p_0)_* \pi = \mu_0, (p_1)_* \pi = \mu_1.$$

Then $(P(X), W_2)$ is a metric space called the **Wasserstein space**. The metric topology is the weak-$*$ topology.

Proposition. *If X is a length space then so is the Wasserstein space $P(X)$.*

Hence we can talk about its (minimizing) geodesics $\{\mu_t\}_{t \in [0,1]}$, called **Wasserstein geodesics**

Proposition. *Wasserstein geodesics $\leftrightarrow$ Optimal dynamical transference plans*
(i.e. dirt moves along geodesics in X.)

Convexity on Wasserstein space

ν background measure.

We want to talk about whether U_ν is a convex function on $P(X)$.

That is, given $\mu_0, \mu_1 \in P(X)$, whether U_ν restricts to a convex function along a Wasserstein geodesic $\{\mu_t\}_{t \in [0,1]}$ from μ_0 to μ_1.

Nonnegative N-Ricci curvature

Definition. *Given $N \in [1, \infty]$, we say that a compact measured length space (X, d, ν) has nonnegative N-Ricci curvature if :*

For all $\mu_0, \mu_1 \in P(X)$ with $\mathrm{supp}(\mu_0) \subset \mathrm{supp}(\nu)$ *and* $\mathrm{supp}(\mu_1) \subset \mathrm{supp}(\nu)$, *there is* some *Wasserstein geodesic $\{\mu_t\}_{t \in [0,1]}$ from μ_0 to μ_1 so that for all $U \in DC_N$ and all $t \in [0, 1]$,*

$$U_\nu(\mu_t) \leq t\, U_\nu(\mu_1) + (1-t)\, U_\nu(\mu_0).$$

Note : We only require convexity along *some* geodesic from μ_0 to μ_1, not all geodesics.

But the same geodesic has to work for all $U \in DC_N$.

Weak displacement convexity. Works better.

What does this have to do with curvature?

Look at optimal transport on the 2-sphere.

$\nu =$ normalized Riemannian density.
Take μ_0, μ_1 two disjoint congruent blobs. Then $U_\nu(\mu_0) = U_\nu(\mu_1)$.

Optimal transport from μ_0 to μ_1 goes along geodesics. **Positive** curvature gives **focusing** of geodesics. Take snapshot at time t.

Intermediate-time blob μ_t is more spread out, so it's *more* uniform w.r.t. ν.
Negative entropy U_ν measures *nonuniformity*.
So $U_\nu(\mu_t) \leq U_\nu(\mu_0) = U_\nu(\mu_1)$, i.e.

$$U_\nu(\mu_t) \leq t\, U_\nu(\mu_1) + (1-t)\, U_\nu(\mu_0).$$

Main result

Theorem. *Let $\{(X_i, d_i, \nu_i)\}_{i=1}^{\infty}$ be a sequence of compact measured length spaces with*

$$\lim_{i \to \infty} (X_i, d_i, \nu_i) = (X, d, \nu)$$

in the measured Gromov-Hausdorff topology.

For any $N \in [1, \infty]$, if each (X_i, d_i, ν_i) has nonnegative N-Ricci curvature then (X, d, ν) has nonnegative N-Ricci curvature.

The proof is a bit involved.

What does all this have to do with <u>Ricci</u> curvature?

Let (M, g) be a compact connected n-dimensional Riemannian manifold.

We could take the canonical measure, but let's be more general.

Say $\Psi \in C^\infty(M)$ has

$$\int_M e^{-\Psi} \, \mathrm{dvol}_M = 1.$$

Put $\nu = e^{-\Psi} \, \mathrm{dvol}_M$.

Any smooth positive probability measure on M can be written in this way.

Definition. *For $N \in [1, \infty]$, define the N-Ricci tensor* $\mathrm{Ric_N}$ *of* (M, g, ν) *by*

$$\begin{cases} \mathrm{Ric} + \mathrm{Hess}(\Psi) & \text{if } N = \infty, \\ \mathrm{Ric} + \mathrm{Hess}(\Psi) - \frac{1}{\mathrm{N-n}} \mathrm{d}\Psi \otimes \mathrm{d}\Psi & \text{if } n < N < \infty \\ \mathrm{Ric} + \mathrm{Hess}(\Psi) - \infty \, (\mathrm{d}\Psi \otimes \mathrm{d}\Psi) & \text{if } N = n, \\ -\infty & \text{if } N < n, \end{cases}$$

where by convention $\infty \cdot 0 = 0$.

$\mathrm{Ric_N}$ is a symmetric covariant 2-tensor field on M that depends on g and Ψ.

(If $N = n$ then $\mathrm{Ric_N}$ is $-\infty$ except where Ψ is locally constant. There, $\mathrm{Ric_N} = \mathrm{Ric}$.)

$\mathrm{Ric}_\infty =$ Bakry-Emery tensor.

Intuition : M has dimension n but pretends to have dimension N. (Identity theft)

$\mathrm{Ric_N}$ would be the "effective" Ricci tensor if M did have dimension N.

Abstract Ricci recovers classical Ricci

Recall that $\nu = e^{-\Psi}\,\mathrm{dvol}_M$.

Theorem. *For $N \in [1, \infty]$, the measured length space (M, g, ν) has nonnegative N-Ricci curvature if and only if $\mathrm{Ric}_N \geq 0$.*

(Related to earlier work of Cordoro-Erausquin-McCann-Schmuckenschläger and Sturm-von Renesse.)

Classical case : Ψ constant, so $\nu = \frac{\mathrm{dvol}}{\mathrm{vol(M)}}$.

Then (M, g, ν) has <u>abstract</u> nonnegative N-Ricci curvature if and only if it has <u>classical</u> nonnegative N-Ricci curvature, provided that $N \geq n$.

Nontrivial consequences of the definition

Had Gromov precompactness theorem. What are the limit spaces (X, d, ν)? *Suppose* that the limit space is a *smooth* measured length space, i.e.

$$(X, d, \nu) = (B, g_B, e^{-\Psi} \mathrm{dvol_B})$$

for some n-dimensional smooth Riemannian manifold (B, g_B) and some $\Psi \in C^\infty(B)$.

Corollary. *If* $(B, g_B, e^{-\Psi}\mathrm{dvol_B})$ *is a measured Gromov-Hausdorff limit of Riemannian manifolds with nonnegative Ricci curvature and dimension at most* N *then* $\mathrm{Ric_N(B)} \geq 0$*.*

Note : the dimension can drop on taking limits.

<u>Converse essentially true</u>
If $(B, g_B, e^{-\Psi}\mathrm{dvol_B})$ has $\mathrm{Ric_N(B)} \geq 0$ then it **is** a measured Gromov-Hausdorff limit of Riemannian manifolds with nonnegative Ricci curvature and dimension at most N, provided that $N \geq \mathrm{dim}(B) + 2$.

Proof of Corollary :

Suppose that $\left\{\left(M_i, g_i, \frac{\mathrm{dvol_{M_i}}}{\mathrm{vol(M_i)}}\right)\right\}_{i=1}^{\infty}$ is a sequence of Riemannian manifolds with

1. $\mathrm{dim}(M_i) \leq N$.
2. $\mathrm{Ric(M_i)} \geq 0$.
3. $\lim_{i \to \infty} \left(M_i, g_i, \frac{\mathrm{dvol_{M_i}}}{\mathrm{vol(M_i)}}\right) = (B, g_B, e^{-\Psi} \mathrm{dvol_B})$ in the measured Gromov-Hausdorff topology.

From the second theorem, each $\left(M_i, g_i, \frac{\mathrm{dvol_{M_i}}}{\mathrm{vol(M_i)}}\right)$ has nonnegative N-Ricci curvature in the abstract sense.

From the first theorem, $(B, g_B, e^{-\Psi} \mathrm{dvol_B})$ has nonnegative N-Ricci curvature in the abstract sense.

From the second theorem, this means that $\mathrm{Ric_N} \geq 0$ on B (as a classical tensor).

More consequences of the definition

1. Bishop-Gromov-type inequality

Theorem. *If* (X, d, ν) *has nonnegative* N*-Ricci curvature and* $x \in \mathrm{supp}(\nu)$ *then* $r^{-N}\, \nu(B_r(x))$ *is nonincreasing in* r*.*

2. Sharp global Poincaré inequality

Theorem. *If* (X, d, ν) *has* N*-Ricci curvature bounded below by* $K > 0$ *and* f *is a Lipschitz function on* X *with* $\int_X f\, d\nu = 0$ *then*

$$\int_X f^2\, d\nu \leq \frac{N-1}{N} \frac{1}{K} \int_X |\nabla f|^2\, d\nu.$$

Here

$$|\nabla f|(x) = \limsup_{y \to x} \frac{|f(y) - f(x)|}{d(y, x)}.$$

<u>Special case</u> (Lichnerowicz' theorem) : If a connected N-dimensional Riemannian manifold has Ric $\geq$ Kg then $\lambda_1(-\triangle) \geq \frac{N}{N-1} K$.

3. Local Poincaré inequality :

Theorem. *If* (X, d, ν) *has nonnegative* N*-Ricci curvature and* f *is a Lipschitz function on* X *then for any ball* $B = B_r(x)$ *with* $\nu[B] > 0$*,*

$$\fint_B |f - \langle f \rangle_B| \, d\nu \;\le\; 2^{2N+1}\, r \fint_{2B} |\nabla f| \, d\nu,$$

provided that for almost all $(x_0, x_1) \in X \times X$*, there's a unique minimizing geodesic from* x_0 *to* x_1*.*

Here $2B = B_{2r}(x)$,

$$\fint_B \cdot \, d\nu = \frac{1}{\nu(B)} \int_B \cdot \, d\nu$$

and

$$\langle f \rangle_B = \fint_B f \, d\nu.$$

(Related work by von Renesse.)

4. Ricci O'Neill theorem

Open questions :

1. Take <u>any</u> result that you know about Riemannian manifolds with nonnegative Ricci curvature (or Ricci curvature bounded below).

Does it extend to measured length spaces (X, d, ν) with nonnegative N-Ricci curvature (or N-Ricci curvature bounded below)?

2. Take an interesting measured length space (X, d, ν). Does it have nonnegative N-Ricci curvature (or N-Ricci curvature bounded below)?

This almost always boils down to understanding the optimal transport on X.

Another topic :

Alexandrov geometry of Wasserstein space

Definition. *A compact length space X has nonnegative Alexandrov curvature if any geodesic triangle in X is "fatter" than the corresponding triangle in $\mathbb{R}^2$.*

Formal Riemannian geometry of Wasserstein space

Suppose that (M, g) is a compact connected Riemannian manifold. What does its Wasserstein space $P(M)$ look like?

Otto, Otto-Villani :

1. *Formally*, $P(M)$ is an infinite-dimensional manifold with a certain Riemannian metric.

(Note : an honest infinite-dimensional Hilbert manifold is never locally compact.)

2. *Formally*, the corresponding distance on $P(M)$ is W_2.

3. (Otto) *Formally*, the Riemannian metric on $P_2(\mathbb{R}^n)$ has nonnegative sectional curvature.

This started the whole story.

Theorem. (M, g) *has nonnegative* sectional curvature *if and only if* $P(M)$ *has nonnegative* Alexandrov curvature.

(Makes rigorous Otto's formal sectional curvature calculation.)

$P(M)$ is an interesting "Alexandrov" space : compact but infinite topological dimension.

<u>Gradient flows</u> : One has existence and uniqueness of the (downward) gradient flow for **any** semiconvex function on a complete Alexandrov space (Perelman-Petrunin).

Hereafter, suppose that M has nonnegative sectional curvature. (For example, the n-torus.)

How to make sense of the formal Riemannian metric on Wasserstein space

A compact length space X with nonnegative Alexandrov curvature has tangent cones (replacing tangent spaces).

Given $x \in X$, look at the space Σ' of minimal geodesics γ emanating from X.

Say $d_{\Sigma'}(\gamma_1, \gamma_2) =$ angle between γ_1 and γ_2.

Take the metric completion of $(\Sigma', d_{\Sigma'})$ to get the *space of directions* Σ.

Definition. *The* tangent cone K_x *is the metric cone over* Σ.

Example : If X is a Riemannian manifold (M, g) then $K_x = T_x M$, with the Euclidean metric on K_x coming from g.

Theorem. *If $\mu \in P(M)$ is absolutely continuous with respect to* $\mathrm{dvol_M}$ *then the tangent cone of $P(M)$ at μ is a Hilbert space. Its inner product is the same as Otto's formal Riemannian metric.*

More precisely, consider the quadratic form

$$Q(\phi) = \int_M |\nabla \phi|^2 \, d\mu$$

on Lip(M).

Quotient by the kernel to get Lip(M)/Ker(Q).

Then the tangent cone at μ is the metric completion of Lip(M)/Ker(Q).

Compare with the formal parametrization of the "tangent space" :

$$\delta\mu = -\nabla \cdot (\mu \nabla \phi).$$

<u>Note</u> : Tangent cones at non-a.c. measures need not be linear spaces. (Example : $\mu = \delta_m$)

Some apparently weird things about Wasserstein space

1. The formal exponential map $\exp_\mu : T_\mu P(M) \to P(M)$ doesn't cover a neighborhood of μ.

2. There is a formal Riemannian metric but no manifold structure.

Claim : This happens all the time for Alexandrov spaces.

Problems with the exponential map

Take a cone in $\mathbb{R}^3$ with cone angle less than 2π.

A geodesic that hits the vertex **cannot** be extended as a minimal geodesic beyond the vertex.

Now take a tetrahedron in $\mathbb{R}^3$. Add conical bumps with a small defect angle.

Add more bumps with smaller defect angle.

Continue and take limit in $\mathbb{R}^3$.

Get a 2-dimensional space X with nonnegative Alexandrov curvature. But for **no** point of X is there an exponential map from the tangent cone onto a neighborhood of the point.

The way out : Use Lipschitz coordinates instead of normal coordinates.

Theorem. *(Otsu-Shioya, Perelman) Any finite-dimensional Alexandrov space X has a Lipschitz-manifold structure almost everywhere. On the "regular" part of X there are*

1. A continuous Riemannian metric.

2. Measurable Christoffel symbols.

3. Jacobi fields.

A simple infinite-dimensional Alexandrov space

$$X = S^1 \times S^1 \times S^1 \times \dots$$

with the "Pythagorean" metric :

$$d_X\left(\{e^{i\theta_j}\},\{e^{i\theta'_j}\}\right) = \sqrt{\sum_{j=1}^{\infty}\left(\frac{d_{S^1}\left(e^{i\theta_j},e^{i\theta'_j}\right)}{j}\right)^2}.$$

The metric topology on X is the product topology.

Formal (flat) Riemannian metric on X:

$$g = \sum_{j=1}^{\infty} j^{-2}\, d\theta_j^2.$$

All tangent cones of X are Hilbert spaces with this inner product. But X is **not** a Hilbert manifold (since it's compact).

Upshot

Alexandrov geometry may be relevant for understanding Wasserstein space.