Spaces with Ricci curvature bounded from below

Nicola Gigli

March 10, 2014

Lessons

- Basics of optimal transport
- Definition of spaces with Ricci curvature bounded from below
- Analysis on spaces with Ricci curvature bounded from below

Lessons

- Basics of optimal transport
- Definition of spaces with Ricci curvature bounded from below
- Analysis on spaces with Ricci curvature bounded from below

Basics of optimal transport

- The optimal transport problem
  - Formulation of the problem
  - First characterization of optimal plans
  - Dual problem and second characterization of optimal plans
- The distance W_2
  - Definition and topology
  - Geodesics

Basics of optimal transport

- The optimal transport problem
  - Formulation of the problem
  - First characterization of optimal plans
  - Dual problem and second characterization of optimal plans
- The distance W_2
  - Definition and topology
  - Geodesics

Push forward of a measure

Let

- X, Y compact metric spaces
- $T : X \to Y$ a Borel map
- $\mu \in \mathscr{P}(X)$

The measure $T_\sharp \mu \in \mathscr{P}(Y)$ is defined by

$$T_\sharp \mu(E) := \mu(T^{-1}(E)), \qquad \forall E \subset Y, \text{ Borel}$$

and satisfies

$$\int f \, \mathrm{d} T_\sharp \mu = \int f \circ T \, \mathrm{d}\mu$$

for any Borel function $f : X \to \mathbb{R}$

Monge's formulation of the optimal transport problem

Given:

- X, Y compact metric spaces
- $c : X \times Y \to \mathbb{R}$ continuous
- $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$

Minimize

$$\int c\big(x, T(x)\big)\, \mathrm{d}\mu(x)$$

among all $T : X \to Y$ Borel such that

$$T_\sharp \mu = \nu$$

Why this is a bad formulation

- Maybe there is no T such that $T_\sharp \mu = \nu$
- The infimum can be not attained
- The functional to minimize is not lower semicontinuous w.r.t. any reasonable weak topology on transport maps

Kantorovich formulation of the optimal transport problem

With the same data as in Monge's formulation, minimize

$$\int c(x, y) \, \mathrm{d}\gamma(x, y)$$

among all $\gamma \in \mathscr{P}_2(X \times X)$ such that

$$\pi^1_\sharp \gamma = \mu$$
$$\pi^2_\sharp \gamma = \nu$$

The set of such admissible transport plans will be denoted $\mathscr{Adm}(\mu, \nu)$

Why this is a good formulation

- There always exists at least one transport plan: $\mu \times \nu$
- Transport plans 'include' transport maps: if $T_{\#}\mu = \nu$, then $(Id, T)_{\#}\mu$ is a transport plan
- The set of transport plans is closed w.r.t. the weak topology of measures
- The map $\gamma \mapsto \int c(x, y) d\gamma(x, y)$ is linear and weakly continuous

Why this is a good formulation

- There always exists at least one transport plan: $\mu \times \nu$
- Transport plans ‘include’ transport maps: if $T_{\#}\mu = \nu$, then $(Id, T)_{\#}\mu$ is a transport plan
- The set of transport plans is closed w.r.t. the weak topology of measures
- The map $\gamma \mapsto \int c(x, y) d\gamma(x, y)$ is linear and weakly continuous

In particular, minima exist.

Basics of optimal transport

- The optimal transport problem
  - Formulation of the problem
  - First characterization of optimal plans
  - Dual problem and second characterization of optimal plans
- The distance W_2
  - Definition and topology
  - Geodesics

A key example

Let $\{x_i\}_i \subset X$, $\{y_i\}_i \subset Y$, $i = 1, \ldots, N$ be given points and

$$\mu := \frac{1}{N} \sum_i \delta_{x_i}$$

$$\nu := \frac{1}{N} \sum_i \delta_{y_i}$$

A key example

Let $\{x_i\}_i \subset X$, $\{y_i\}_i \subset Y$, $i = 1, \dots, N$ be given points and

$$\mu := \frac{1}{N} \sum_i \delta_{x_i}$$
$$\nu := \frac{1}{N} \sum_i \delta_{y_i}$$

Then a plan γ is optimal iff for any $n \in \mathbb{N}$, permutation σ of $\{1, \dots, n\}$ and any $\{(x_i, y_i)\}_{i=1,\dots,n} \subset \mathrm{supp}(\gamma)$ it holds

$$\sum_i c(x_i, y_i) \leq \sum_i c(x_i, y_{\sigma(i)})$$

The general definition

We say that a set $\Gamma \subset X \times Y$ is *c-cyclically monotone* if for any $n \in \mathbb{N}$, permutation σ of $\{1, \ldots, n\}$ and any $\{(x_i, y_i)\}_{i=1,\ldots,n} \subset \Gamma$ it holds

$$\sum_i c(x_i, y_i) \leq \sum_i c(x_i, y_{\sigma(i)})$$

First characterization of optimal plans

Theorem A transport plan γ is optimal if and only if its support $\mathrm{supp}(\gamma)$ is c-cyclically monotone.

First characterization of optimal plans

Theorem A transport plan γ is optimal if and only if its support $\mathrm{supp}(\gamma)$ is *c*-cyclically monotone.

In particular, being optimal depends only on the support of γ, and not on how the mass is distributed on the support (!).

Basics of optimal transport

- The optimal transport problem
  - Formulation of the problem
  - First characterization of optimal plans
  - Dual problem and second characterization of optimal plans
- The distance W_2
  - Definition and topology
  - Geodesics

The dual problem

Given the measures $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ and the cost function $c : X \times Y \to \mathbb{R}$, maximize

$$\int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu$$

among all couples of functions $\varphi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$ such that

$$\varphi(x) + \psi(y) \leq c(x, y), \qquad \forall x \in X, \; y \in Y$$

We call such a couple of functions *admissible potentials*

A simple inequality

Let γ be a transport plan from μ to ν and (φ, ψ) admissible potentials. Then

$$\begin{aligned}\int c(x,y)\,\mathrm{d}\gamma(x,y) &\geq \int \varphi(x) + \psi(y)\,\mathrm{d}\gamma(x,y) \\ &= \int \varphi(x)\,\mathrm{d}\mu(x) + \int \psi(y)\,\mathrm{d}\nu(y).\end{aligned}$$

Thus

$$\inf\{\text{transport problem}\} \geq \sup\{\text{dual problem}\}$$

c-transform

For given $\varphi : X \to \mathbb{R}$ define $\varphi^c : Y \to \mathbb{R}$ as

$$\varphi^c(y) := \inf_x c(x, y) - \varphi(x)$$

and similarly for given $\psi : Y \to \mathbb{R}$ define $\psi^c : X \to \mathbb{R}$ as

$$\psi^c(x) := \inf_y c(x, y) - \psi(y).$$

c-transform

For given $\varphi : X \to \mathbb{R}$ define $\varphi^c : Y \to \mathbb{R}$ as

$$\varphi^c(y) := \inf_x c(x, y) - \varphi(x)$$

and similarly for given $\psi : Y \to \mathbb{R}$ define $\psi^c : X \to \mathbb{R}$ as

$$\psi^c(x) := \inf_y c(x, y) - \psi(y).$$

Notice that:

- (φ, φ^c) is always admissible
- $\varphi^c \geq \psi$ for any ψ such that (φ, ψ) is admissible
- $\varphi^{ccc} = \varphi^c$ for any φ.

c-concavity and c-superdifferential

A function φ is c-concave if $\varphi = \psi^c$ for some function ψ.

c-concavity and c-superdifferential

A function φ is c-concave if $\varphi = \psi^c$ for some function ψ.

Recalling that

$$\varphi(x) + \varphi^c(y) \leq c(x, y) \qquad \forall x \in X,\ y \in Y$$

we define the c-superdifferential $\partial^c\varphi \subset X \times Y$ as the set of (x, y) such that

$$\varphi(x) + \varphi^c(y) = c(x, y).$$

c-concavity and c-superdifferential

A function φ is c-concave if $\varphi = \psi^c$ for some function ψ.

Recalling that

$$\varphi(x) + \varphi^c(y) \leq c(x, y) \qquad \forall x \in X,\ y \in Y$$

we define the c-superdifferential $\partial^c\varphi \subset X \times Y$ as the set of (x, y) such that

$$\varphi(x) + \varphi^c(y) = c(x, y).$$

Note that the c-superdifferential is always c-cyclically monotone:

$$\begin{aligned}
\sum_i c(x_i, y_i) &= \sum_i \varphi(x_i) + \varphi^c(y_i) \\
&= \sum_i \varphi(x_i) + \varphi^c(y_{\sigma(i)}) \leq \sum_i c(x_i, y_{\sigma(i)})
\end{aligned}$$

Second structural theorem

Theorem A set Γ is c-cyclically monotone iff $\Gamma \subset \partial^c \varphi$ for some φ c-concave.

To summarize

Given $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ and a cost function c, for an admissible plan γ the following three are equivalent:

- γ is optimal
- $\mathrm{supp}(\gamma)$ is c-cyclically monotone
- $\mathrm{supp}(\gamma) \subset \partial^c \varphi$ for some c-concave function φ

No duality gap

It holds

$$\inf\{\text{transport problem}\} = \sup\{\text{dual problem}\}$$

Indeed, if γ is optimal, then $\mathrm{supp}(\gamma) \subset \partial^c \varphi$ for some c-concave φ. Thus

$$\int c(x,y)\,\mathrm{d}\gamma(x,y) = \int \varphi(x) + \varphi^c(y)\,\mathrm{d}\gamma(x,y) = \int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu$$

Any such φ is called Kantorovich potential from μ to ν

What these results give on Riemannian manifolds

Let M be a compact smooth Riemannian manifold, $\mu, \nu \in \mathscr{P}(M)$ such that $\mu \ll \mathrm{vol}$ and $c = d^2$.

What these results give on Riemannian manifolds

Let M be a compact smooth Riemannian manifold, $\mu, \nu \in \mathscr{P}(M)$ such that $\mu \ll \mathrm{vol}$ and $c = d^2$.

Then (Brenier-McCann):

- there is only one optimal plan γ
- γ is induced by a map, i.e. there is $T : M \to M$ such that $\gamma = (\mathrm{Id}, T)_\sharp \mu$
- T is of the form $\exp(-\nabla \varphi)$, where φ is a Kantorovich potential from μ to ν.

Basics of optimal transport

- The optimal transport problem
  - Formulation of the problem
  - First characterization of optimal plans
  - Dual problem and second characterization of optimal plans
- The distance W_2
  - Definition and topology
  - Geodesics

Definition

Let (X, d) be a compact metric space and $\mu, \nu \in \mathscr{P}(X)$.

The distance $W_2(\mu, \nu)$ is defined as:

$$W_2^2(\mu, \nu) := \inf_{\gamma \in \mathcal{Adm}(\mu, \nu)} \int d^2(x, y) \, \mathrm{d}\gamma(x, y)$$

Definition

Let (X, d) be a compact metric space and $\mu, \nu \in \mathscr{P}(X)$.

The distance $W_2(\mu, \nu)$ is defined as:

$$W_2^2(\mu, \nu) := \inf_{\gamma \in \mathcal{A}dm(\mu,\nu)} \int d^2(x, y) \, \mathrm{d}\gamma(x, y)$$

It is easy to see that

- $W_2(\mu, \nu) \in [0, \infty)$ for every μ, ν
- $W_2(\mu, \nu) = 0$ if and only if $\mu = \nu$
- $W_2(\mu, \nu) = W_2(\nu, \mu)$ for every μ, ν

Definition

Let (X, d) be a compact metric space and $\mu, \nu \in \mathscr{P}(X)$.

The distance $W_2(\mu, \nu)$ is defined as:

$$W_2^2(\mu, \nu) := \inf_{\gamma \in \mathcal{Adm}(\mu,\nu)} \int d^2(x, y) \, \mathrm{d}\gamma(x, y)$$

It is easy to see that

- $W_2(\mu, \nu) \in [0, \infty)$ for every μ, ν
- $W_2(\mu, \nu) = 0$ if and only if $\mu = \nu$
- $W_2(\mu, \nu) = W_2(\nu, \mu)$ for every μ, ν

It is less clear that the triangle inequality holds

(almost) proof of triangle inequality

Let $\mu, \nu, \sigma \in \mathscr{P}(X)$ and $T, S : X \to X$ such that $T_\sharp \mu = \sigma$, $S_\sharp \sigma = \nu$ and

$$W_2^2(\mu, \sigma) = \int d^2(x, T(x))\,\mathrm{d}\mu(x) \qquad W_2^2(\sigma, \nu) = \int d^2(y, S(y))\,\mathrm{d}\sigma(y)$$

(almost) proof of triangle inequality

Let $\mu, \nu, \sigma \in \mathscr{P}(X)$ and $T, S : X \to X$ such that $T_\sharp \mu = \sigma$, $S_\sharp \sigma = \nu$ and

$$W_2^2(\mu, \sigma) = \int d^2(x, T(x))\, \mathrm{d}\mu(x) \qquad W_2^2(\sigma, \nu) = \int d^2(y, S(y))\, \mathrm{d}\sigma(y)$$

Then $(S \circ T)_\sharp \mu = \nu$ and thus

(almost) proof of triangle inequality

Let $\mu, \nu, \sigma \in \mathscr{P}(X)$ and $T, S : X \to X$ such that $T_\sharp \mu = \sigma$, $S_\sharp \sigma = \nu$ and

$$W_2^2(\mu, \sigma) = \int d^2(x, T(x))\,\mathrm{d}\mu(x) \qquad W_2^2(\sigma, \nu) = \int d^2(y, S(y))\,\mathrm{d}\sigma(y)$$

Then $(S \circ T)_\sharp \mu = \nu$ and thus

$$\begin{aligned}
W_2(\mu, \nu) &\leq \sqrt{\int d^2(x, S(T(x)))\,\mathrm{d}\mu(x)} \\
&\leq \sqrt{\int \big(d(x, T(x)) + d(T(x), S(T(x)))\big)^2\,\mathrm{d}\mu(x)} \\
&\leq \sqrt{\int d^2(x, T(x))\,\mathrm{d}\mu(x)} + \sqrt{\int d^2(T(x), S(T(x)))\,\mathrm{d}\mu(x)} \\
&= W_2(\mu, \sigma) + W_2(\sigma, \nu)
\end{aligned}$$

Embedding of X in $\mathscr{P}(X)$

The map

$$x \quad \mapsto \quad \delta_x$$

is an isometry from (X, d) to $(\mathscr{P}(X), W_2)$

Embedding of X in $\mathscr{P}(X)$

The map

$$x \quad \mapsto \quad \delta_x$$

is an isometry from (X, d) to $(\mathscr{P}(X), W_2)$

In particular, W_2 has little to do with the total variation distance and is more linked to the weak convergence.

The topology of $(\mathscr{P}(X), W_2)$

Theorem The topology induced by W_2 on $\mathscr{P}(X)$ is the same as the weak topology, i.e. $W_2(\mu_n, \mu) \to 0$ if and only if

$$\lim_{n\to\infty} \int \varphi \, \mathrm{d}\mu_n = \int \varphi \, \mathrm{d}\mu \qquad \forall \varphi \in C(X).$$

The topology of $(\mathscr{P}(X), W_2)$

Theorem The topology induced by W_2 on $\mathscr{P}(X)$ is the same as the weak topology, i.e. $W_2(\mu_n, \mu) \to 0$ if and only if

$$\lim_{n\to\infty} \int \varphi \, \mathrm{d}\mu_n = \int \varphi \, \mathrm{d}\mu \qquad \forall \varphi \in C(X).$$

In particular, $(\mathscr{P}(X), W_2)$ is compact.

Basics of optimal transport

- The optimal transport problem
  - Formulation of the problem
  - First characterization of optimal plans
  - Dual problem and second characterization of optimal plans
- The distance W_2
  - Definition and topology
  - Geodesics

Geodesic spaces

Let (X, d) be a metric space. A curve $\gamma : [0, 1] \to X$ is a geodesic provided:

$$d(\gamma_t, \gamma_s) = |s - t| d(\gamma_0, \gamma_1) \qquad \forall t, s \in [0, 1].$$

(X, d) is said geodesic provided for every $x, y \in X$ there exists a geodesic such that $\gamma_0 = x$ and $\gamma_1 = y$

A simple geodesic

Let (X, d) be a geodesic space and notice that since $x \mapsto \delta_x$ is an isometric embedding, if γ is a geodesic from x to y in (X, d), then

$$t \mapsto \delta_{\gamma_t}$$

is a geodesic from δ_x to δ_y in $(\mathscr{P}(X), W_2)$.

A simple geodesic

Let (X, d) be a geodesic space and notice that since $x \mapsto \delta_x$ is an isometric embedding, if γ is a geodesic from x to y in (X, d), then

$$t \mapsto \delta_{\gamma_t}$$

is a geodesic from δ_x to δ_y in $(\mathscr{P}(X), W_2)$.

Notice that instead the linear interpolation

$$t \mapsto (1 - t)\delta_x + t\delta_y$$

has infinite length as soon as $x \neq y$.

The space of geodesics

We denote by $\mathrm{Geo}(X)$ the space of all geodesics on X endowed with the sup distance.

For $t \in [0, 1]$ the evaluation map $\mathrm{e}_t : \mathrm{Geo}(X) \to X$ is defined as

$$\mathrm{e}_t(\gamma) := \gamma_t \qquad \forall \gamma \in \mathrm{Geo}(X)$$

Measures on $\mathrm{Geo}(X)$ and geodesics in $(\mathscr{P}(X), W_2)$

Let $\boldsymbol{\pi} \in \mathscr{P}(\mathrm{Geo}(X))$ such that

$$(\mathrm{e}_0, \mathrm{e}_1)_\sharp \boldsymbol{\pi} \text{ is an optimal plan}$$

Measures on $\mathrm{Geo}(X)$ and geodesics in $(\mathscr{P}(X), W_2)$

Let $\boldsymbol{\pi} \in \mathscr{P}(\mathrm{Geo}(X))$ such that

$$(\mathrm{e}_0, \mathrm{e}_1)_\sharp \boldsymbol{\pi} \text{ is an optimal plan}$$

Then

$$\begin{aligned}
W_2^2\big((\mathrm{e}_t)_\sharp \boldsymbol{\pi}, (\mathrm{e}_s)_\sharp \boldsymbol{\pi}\big) &\leq \int d^2(x, y)\, \mathrm{d}(\mathrm{e}_t, \mathrm{e}_s)_\sharp \boldsymbol{\pi}(x, y) \\
&= \int d^2(\gamma_t, \gamma_s)\, \mathrm{d}\boldsymbol{\pi}(\gamma) \\
&= |s - t|^2 \int d^2(\gamma_0, \gamma_1)\, \mathrm{d}\boldsymbol{\pi}(\gamma) \\
&= |s - t|^2 W_2^2\big((\mathrm{e}_0)_\sharp \boldsymbol{\pi}, (\mathrm{e}_1)_\sharp \boldsymbol{\pi}\big)
\end{aligned}$$

Measures on $\mathrm{Geo}(X)$ and geodesics in $(\mathscr{P}(X), W_2)$

Let $\boldsymbol{\pi} \in \mathscr{P}(\mathrm{Geo}(X))$ such that

$$(\mathrm{e}_0, \mathrm{e}_1)_\sharp \boldsymbol{\pi} \text{ is an optimal plan}$$

Then

$$\begin{aligned}
W_2^2\big((\mathrm{e}_t)_\sharp \boldsymbol{\pi}, (\mathrm{e}_s)_\sharp \boldsymbol{\pi}\big) &\leq \int d^2(x,y)\,\mathrm{d}(\mathrm{e}_t, \mathrm{e}_s)_\sharp \boldsymbol{\pi}(x,y) \\
&= \int d^2(\gamma_t, \gamma_s)\,\mathrm{d}\boldsymbol{\pi}(\gamma) \\
&= |s-t|^2 \int d^2(\gamma_0, \gamma_1)\,\mathrm{d}\boldsymbol{\pi}(\gamma) \\
&= |s-t|^2 W_2^2\big((\mathrm{e}_0)_\sharp \boldsymbol{\pi}, (\mathrm{e}_1)_\sharp \boldsymbol{\pi}\big)
\end{aligned}$$

In other words

$$t \quad \mapsto \quad (\mathrm{e}_t)_\sharp \boldsymbol{\pi}$$

is a geodesic in $(\mathscr{P}(X), W_2)$

Characterization of geodesics in $(\mathscr{P}(X), W_2)$

Theorem Let (X, d) be a compact geodesic space.
Then the following are equivalent:

- $t \mapsto \mu_t$ is a geodesic in $(\mathscr{P}(X), W_2)$
- there exists $\boldsymbol{\pi} \in \mathscr{P}(\mathrm{Geo}(X))$ such that

$$(\mathrm{e}_t)_\sharp \boldsymbol{\pi} = \mu_t \qquad \text{for every } t \in [0, 1]$$

$$(\mathrm{e}_0, \mathrm{e}_1)_\sharp \boldsymbol{\pi} \text{ is an optimal plan from } \mu_0 \text{ to } \mu_1$$

What this result gives on Riemannian manifolds

Let $\mu_0, \mu_1 \in \mathscr{P}(M)$ be such that $\mu_0 \ll \mathrm{vol}$.

Then there is only one geodesic (μ_t) from μ_0 to μ_1 and it satisfies

$$\partial_t \mu_t - \nabla \cdot (\nabla \varphi_t \mu_t) = 0$$

where φ_0 is a Kantorovich potential from μ_0 to μ_1 and (φ_t) solves

$$\partial_t \varphi_t + \frac{|\nabla \varphi_t|^2}{2} = 0$$

Thank you