Metric measure spaces with lower Ricci curvature bounds

Christian Ketterer

CHAPTER 1

Introduction

These are the notes of the lecture "Metric measure spaces with lower Ricci curvature bounds" given by the author in winter term 2014/15 at the mathematical institute of Freiburg. The goal of the lecture was to present an introduction for beginners to the theory of optimal transport for metric measure space focussing on spaces that admit a lower Ricci curvature bounds. The main result is that in context of smooth metric measure spaces one has equivalence of generalized lower Ricci curvature bounds in the sense of optimal transport with the classical notion that is known from Riemannian geometry.

1. Motivation

Definition 1.1. Let us consider an arbitrary function $f : X \subset \mathbb{R} \to \mathbb{R}$.

1. One says the f is convex if for all $x, y \in X$ and for all $t \in X$ we have $(1-t)x + ty \in X$ and
$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$
In particular, X is an interval of the form $[a, b]$.
2. If we assume that $f \in C^2[a, b]$, then f is convex if and only if $f'' \geq 0$.

Remark 1.2. We summarize some aspects of the previous definition.

(i) (1) is a geometric property, that one can visualize easily.
(ii) (1) is stable. More precisely, assume X_n converges to X in the Hausdorff sense and $f_n : X_n \to \mathbb{R}$ converges to $f : X \to \mathbb{R}$ in the way that for any $\epsilon > 0$ there exists $N_0 \in \mathbb{N}$ and $\delta > 0$ such that $|f_n(x) - f(y)| < \epsilon$ for $n \geq N_0$ and $|x - y| < \delta$. Then, if f_n is convex, f is convex.
(iii) On the other hand, (2) is easy to check for C^2-functions.
(iv) But (2) is not stable w.r.t. to pointwise or uniform convergence.

(2) yields (1). Additionally, there is a theorem by A.D. Alexandrov that tells us that any convex function is twice differentiable almost everywhere.

Theorem 1.3. *Assume $f : [a, b] \to \mathbb{R}$ is a convex function. Then $f''(t)$ exists for $\mathcal{L}^1$-almost every $t \in [a, b]$, and $f'' \geq 0$ in the distributional sense.*

Riemannian geometry. Let (M, g) be a complete, connected Riemannian manifold. ∇ denotes its Levi-Civita connection. The Riemannian curvature tensor of the Riemannian metric $g = \langle \cdot, \cdot \rangle$ is defined as follows. Consider vector field X, Y on M.
$$R(X, Y)V = \nabla_X \nabla_Y V - \nabla_Y \nabla_X V - \nabla_{[X,Y]} V.$$

CHAPTER 2

General optimal transport theory

Notations. A Polish space X is a topological space where the topology comes from a metric that is complete and separable. The space of probability measures on X is denoted by $\mathcal{P}(X)$. It is equipped with the topology of weak convergence. For a sequence of probability measures (μ_n) we say that (μ_n) converges weakly to a probability measure μ if

$$\int_X h d\mu_n \to \int_X h d\mu \text{ for any } f \in C_b(X).$$

$C_b(X)$ denotes the set of bounded and continuous functions on X. Let $Z = X \times Y$ that is a Polish space as well. The topology is the product topology inherited by X and Y. $p_X : Z \to X$ and $p_Y : Z \to Y$ will be the projection maps.

Definition 0.4 (Monge problem). Let X and Y be Polish spaces. Let $c : X \times Y \to \mathbb{R} \cup \{\infty\}$ be a Borel measurable *cost function*. $\mathbb{R} \cup \{\pm\infty\}$ is equipped with the extended Borel σ-field. For probability measure $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ the Monge problem with respect to the cost function c is to find a measurable map $T : X \to Y$ such that $T_*\mu = \nu$ - we say that T is a transport map - and T is a minimizer of

$$S \longmapsto \int_X c(x, S(x)) d\mu(x) = \mathrm{Cost}(S) \tag{1}$$

where S is any transport map between μ and ν. If T is a transport map and a minimizer of (1), we call T an optimal map.

Example 0.5. Consider $X = Y = \mathbb{R}$, $\mu = \delta_x$ and $\nu = \sum_{i=1}^2 \frac{1}{2}\delta_{y_i}$ and $c(x, y) = |x - y|$. Then, there is no transport map between μ and ν. Hence, in general there is no solution for the Monge problem. But we can generalize the concept of transport map in an appropriate way.

Definition 0.6. Consider $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. A probability measure $\pi \in \mathcal{P}(Z)$ is a coupling of μ and ν if $(p_X)_*\pi = \mu$ and $(p_Y)_* = \nu$, or equivalently

$\pi(A \times Y) = \mu(A)$ for any Borel set $A \subset X$, and
$\pi(X \times B) = \nu(B)$ for any Borel set $B \subset Y$.

The set of all couplings between μ and ν is denoted by $\mathrm{Cpl}(\mu, \nu)$. The *total transportation cost* of a coupling $\pi \in \mathrm{Cpl}(\mu, \nu)$ is

$$\int_Z c(x, y) d\pi(x, y) =: \mathrm{Cost}(\pi).$$

A probabilistic formulation is: a coupling between μ and ν is a random variable $W : (\Omega, \mathcal{A}, \mathbb{P}) \to Z$ on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $(p_X \circ W)_*\mathbb{P} = \mu$ and $(p_Y \circ W)_*\mathbb{P} = \nu$. Then $\pi \in \mathcal{P}(Z)$ is obtained by $W_*\mathbb{P}$.

Example 0.7. (i) $\pi := \mu \otimes \nu$ is a coupling. Therefore, in the probabilistic formulation we choose independent random variable U and V with distributions μ and ν respectively and set $W = (U, V)$.
(ii) If $T : X \to Y$ is transport map between μ and ν then $(\mathrm{id}_X, T)_*\mu$ is a coupling of μ and ν.

DEFINITION 0.8 (Kantorovich problem). Find a coupling $\pi \in \mathrm{Cpl}(\mu,\nu)$ such that

$$\int_Z c(x,y)d\pi(x,y) = \inf_{\bar\pi\in \mathrm{Cpl}(\mu,\nu)} \int_Z c(x,y)d\bar\pi(x,y) = \inf_{\bar\pi\in \mathrm{Cpl}(\mu,\nu)} \mathrm{Cost}(\bar\pi) =: \mathrm{Cost}(\mu,\nu). \tag{2}$$

$\mathrm{Cost}(\mu,\nu)$ is the *minimal transportation cost* between μ and ν. If $\pi \in \mathrm{Cpl}(\mu,\nu)$ solves (2), we say π is an optimal coupling of μ and ν. The set of optimal couplings of μ and ν is denoted by $\mathrm{OptCpl}(\mu,\nu)$.

EXAMPLE 0.9 (Optimal couplings are not unique). Consider $X = Y = \mathbb{R}^2$ and probability measures $\mu = \frac{1}{2}\delta_{(-1,0)} + \frac{1}{2}\delta_{(1,0)}$ and $\nu = \frac{1}{2}\delta_{(0,-1)} + \frac{1}{2}\delta_{(0,1)}$. Then $\pi = \frac{1}{2}\delta_{((-1,0),(0,1))} + \frac{1}{2}\delta_{((1,0),(0,-1))}$ and $\widetilde\pi = \frac{1}{2}\delta_{((1,0),(0,-1))} + \frac{1}{2}\delta_{((-1,0),(0,1))}$ are optimal couplings of μ and ν.

REMARK 0.10. Since $(\mathrm{id}_X, T)_*\mu \in \mathrm{Cpl}(\mu,\nu)$ for any transport map $T : X \to Y$ between μ and ν, we have $\mathrm{Cost}(\mu,\nu) \le \mathrm{Cost}(T)$.

The next theorem states that optimal couplings exists under mild assumptions on the cost function.

THEOREM 0.11 (Kantorovich). *If $c : X \times Y \to \mathbb{R} \cup \{\infty\}$ is lower semi-continuous, and there are upper semi-continuous functions $a, b : X, Y \to \mathbb{R} \cup \{-\infty\}$, then for any $\mu \in \mathcal{P}(X)$ and for any $\nu \in \mathcal{P}(Y)$ there exists an optimal coupling. More precisely, there exists $\pi \in \mathcal{P}(X \times Y)$ such that*

$$\mathrm{Cost}(\pi) = \int_{X\times Y} c(x,y)d\pi(x,y) = \mathrm{Cost}(\mu,\nu).$$

The proof of the theorem is based on the following criterion for relative compactness in $\mathcal{P}(X \times Y)$ with respect to weak convergence. A subset $\mathcal{K} \subset \mathcal{P}(X)$ is relatively compact if any sequence in $\mathcal{K}$ has a subsequence that converges in $\mathcal{P}(X)$ with respect to weak convergence.

THEOREM 0.12 (Prohorov). *Let X be a Polish space, and $\mathcal{K} \subset \mathcal{P}(X)$. $\mathcal{K}$ is relatively compact with respect to weak convergence if and only if $\mathcal{K}$ is tight. Tight means that for all $\epsilon > 0$ there exists a compact subset $K_\epsilon \subset X$ such that $\mu(X\backslash K_\epsilon) < \epsilon$ for all $\mu \in \mathcal{K}$.*

PROOF. Billingsley ☐

LEMMA 0.13. *Let $\mathcal{P}_1 \subset \mathcal{P}(X)$ and $\mathcal{P}_2 \subset \mathcal{P}(Y)$ be relatively compact subsets. Then*

$$\mathcal{P} = \{\pi \in \mathcal{P}(X \times Y) : (p_1)_*\pi \in \mathcal{P}_1 \text{ and } (p_2)_*\pi \in \mathcal{P}_2\}$$

is relatively compact in $\mathcal{P}(Z)$.

PROOF. By Prohorov's criterion $\mathcal{P}_1$ and $\mathcal{P}_2$ are tight. Therefore, we can find compact subsets $K^1_\epsilon \subset X$ and $K^2_\epsilon \subset Y$ with respect to $\epsilon/2 > 0$ like in Prohorov's theorem. Then, we consider $K_\epsilon = K^1_\epsilon \times K^2_\epsilon$. If $\pi \in \mathcal{P}$, we have the following estimate.

$$\pi(Z\backslash K_\epsilon) \le \pi\left((X\backslash K_\epsilon) \times Y\right) + \pi(\,X \times (Y\backslash K^2_\epsilon)) < \epsilon/2 + \epsilon/2.$$

Hence, $\mathcal{P}$ is tight, and by Prohorov's theorem it is relatively compact. ☐

COROLLARY 0.14. *For any $\mu \in \mathcal{P}(X)$ and for any $\nu \in \mathcal{P}(Y)$ the set of couplings $\mathrm{Cpl}(\mu,\nu) \subset \mathcal{P}(Z)$ is compact with respect to weak convergence.*

PROOF. It is trivial that the sets $\{\mu\} = \mathcal{P}_1$ and $\{\nu\} = \mathcal{P}_2$ are relatively compact, and by the previous lemma $\mathrm{Cpl}(\mu,\nu)$ is relatively compact as well. Therefore for any sequence $(\pi_i) \subset \mathrm{Cpl}(\mu,\nu)$ there is a subsequence $(\widetilde\pi_i)$ that converges weakly in $\mathcal{P}(Z)$ to some probability measure π. If we consider an arbitrary function $f \in C_b(X)$, then $f \circ p_1 \in C_b(Z)$, and we see that

$$const = \int_X f d\mu = \int_Z f \circ p_1 d\widetilde\pi_i \to \int_Z f \circ p_1 d\pi = \int_Z f d(p_1)_*\pi$$

We conclude that $(p_1)_*\pi = \mu$, and similar $(p_2)_*\pi = \nu$. Hence, π is a coupling of μ and ν and $\mathrm{Cpl}(\mu,\nu)$ is compact. ☐

LEMMA 0.15. *Consider Polish spaces X and Y and $c : X \times Y \to \mathbb{R} \cup \{\infty\}$ like in Theorem 0.11. Then*

$$\pi \in \mathcal{P}(Z) \to \mathrm{Cost}(\pi)$$

is lower semi-continuous with respect to weak convergence.

PROOF. We can assume that the cost function is non-negative by replacing c with $c - a - b$. If the cost function c is continuous and bounded, then the assertion follows directly from the definition of weak convergence. For the general case we assume that the topologies of X and Y are induced by finite metrics d_X and d_Y respectively. Otherwise, we can replace d_X for instance by $\widetilde{\mathrm{d}}_X = \frac{\mathrm{d}_X}{1+\mathrm{d}_X}$ that is finite and induces the same topology as d_X. Consider

$$c_k(x, y) = \inf_{(\widetilde{x},\widetilde{y})\in Z} \{\min\{c(x, y), k\} + k [\mathrm{d}_X(x, \widetilde{x}) + \mathrm{d}_Y(y, \widetilde{y})]\}.$$

(Exercise.) The sequence (c_k) satisfies

(i) For each $k \in \mathbb{N}$ $c_k : Z \to \mathbb{R}$ is bounded and continuous with respect to $\mathrm{d}_X + \mathrm{d}_Y$.
(ii) $c_k \leq c$ and $c_k \to c$ pointwise everywhere.

Consider a weakly converging sequence $\pi_i \to \pi$. Then the theorem of monotone convergence implies

$$\int_Z c(x, y) d\pi(x, y) = \lim_{k\to\infty} \lim_{i\to\infty} \int_Z \underbrace{c_k(x, y)}_{\leq c(x,y)} d\pi_i(x, y) \leq \liminf_{i\to\infty} \int_Z c(x, y) d\pi_i(x, y)$$

□

PROOF OF THEOREM 0.11. Choose a sequence π_i sucht $\mathrm{Cost}(\pi_i)$ converges to $\mathrm{Cost}(\mu, \nu)$. By Corollary 0.14 $\mathrm{Cpl}(\mu, \nu)$ is compact. Hence, a subsequence of π_i converges weakly to $\pi \in \mathrm{Cpl}(\mu, \nu)$. The sub-sequence is also denoted with π_i. Since the total cost Cost is lower semi-continuous, it follows

$$\mathrm{Cost}(\pi) \leq \liminf_{i\to\infty} \int_Z c(x, y) d\pi_i(x, y) = \mathrm{Cost}(\mu, \nu).$$

Hence, π is an optimal coupling for μ and ν. □

THEOREM 0.16. *Let X, Y, μ, ν and c be as in Theorem 0.11 such that $\mathrm{Cost}(\mu, \nu) < \infty$. Consider $\pi \in \mathrm{OptCpl}(\mu, \nu)$. If $\widetilde{\pi} \leq \pi$ (i.e. $\widetilde{\pi}(C) \leq \pi(C)$ for any Borel set $C \subset X \times Y$) and if $\widetilde{\pi}(Z) > 0$, then $\pi' = \widetilde{\pi}(Z)^{-1}\widetilde{\pi}$ is an optimal plan between $(p_1)_*\pi' =: \mu'$ and $(p_2)_*\pi' =: \nu'$.*

PROOF. Let us assume that π' is not optimal. Then there exists $\pi'' \in \mathrm{Cpl}(\mu', \nu')$ such that $\mathrm{Cost}(\pi'') < \mathrm{Cost}(\pi')$. Consider

$$\hat{\pi} := \underbrace{(\pi - \widetilde{\pi})}_{\geq 0} + \underbrace{\widetilde{\pi}(Z)}_{>0} \cdot \pi''.$$

It is immediate that $\hat{\pi}(Z) = 1$, and therefore $\hat{\pi} \in \mathcal{P}(Z)$. $\hat{\pi}$ is also a coupling for μ and ν since

$$\hat{\pi}(A \times Y) = (\underbrace{\pi(A \times X)}_{\mu(A)} - \widetilde{\pi}(A \times X)) + \widetilde{\pi}(Z)\pi''(A \times X) = \mu(A).$$

Similar $\hat{\pi}(X \times B) = \nu(B)$. We can compute the following.

$$\int_Z cd\hat{\pi} - \int_Z cd\pi = -\int_Z cd\widetilde{\pi} + \widetilde{\pi}(Z) \underbrace{\int_Z cd\pi''}_{<\int_Z cd\pi'} - \int_Z cd\widetilde{\pi} + \int_Z cd\widetilde{\pi} = 0.$$

Hence $\mathrm{Cost}(\hat{\pi}) < \mathrm{Cost}(\pi)$. This contradicts the optimality of π. □

Question: How can we improve a plan such that it becomes optimal?

DEFINITION 0.17. Let X and Y be sets, and let c be a measurable cost function. We say that a subset $\Gamma \subset X \times Y$ is c-monotone if

$$\sum_{i=1}^{N} c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^{N} c(x_i, y_i)$$

whenever $\{(x_1, y_1), \ldots, (x_N, y_N)\} \subset \Gamma$ and $\sigma \in S_N$. We say that a coupling is c-montone if $\operatorname{supp} \pi \subset X \times Y$ is c-montone. The support of a Borel measure μ is the intersection of all closed sets C such that $\mu(C^c) = 0$.

EXAMPLE 0.18. Let $X \subset \mathbb{R}$ and $Y = \mathbb{R}^n$. Let $V : X \to Y$ be a C^1-vector field, and consider

$$\Gamma = \{(x, y) \in X \times Y : y = V(x) \text{ for some } x \in X\} \subset X \times Y.$$

We consider the cost function $c(x, y) = -x \cdot y$. Then Γ is c-monotone if and only if

$$\int_\gamma V \geq 0$$

for any closed curve γ in X. Hence, there exists $u \in C^2(\mathbb{R}^n)$ such that $\nabla u = V$.

REMARK 0.19. If $\pi \in \operatorname{OptCpl}(\mu, \nu)$ with respect to some continuous cost function c, then $\operatorname{supp} \pi$ is c-monotone.

DEFINITION 0.20. Let X and Y be Polisch spaces, and let $c : X \times Y \to \mathbb{R} \cup \{\pm\infty\}$ be measurable cost function. Consider $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$.

(1) We say a pair of functions $(\varphi, \psi) : X \times Y \to (\mathbb{R} \cup \{\pm\infty\})^2$ is competitive if $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$ such that $\varphi(x) + \psi(y) \leq c(x, y)$ for any $(x, y) \in Z$.

(2) The dual Kantorovich problem is to find a competitive pair (φ, ψ) that maximizes

$$D(\widetilde{\varphi}, \widetilde{\psi}) := \int \widetilde{\varphi} d\mu + \int \widetilde{\psi} d\nu$$

where $(\widetilde{\varphi}, \widetilde{\psi})$ is a competitive pair.

REMARK 0.21. If $\pi \in \operatorname{Cpl}(\mu, \nu)$ and (φ, ψ) is a competitive pair and $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$ then

$$\int \varphi d\mu + \int \psi d\nu \leq \int c d\pi.$$

DEFINITION 0.22 (Rüschendorf). Let X and Y be sets, and let $c : X \times Y \to \mathbb{R} \cup \{\infty\}$ be a function.

(1) A function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is *c-concave relative to X and Y* if it is not identically $-\infty$ and there exists $\xi : Y \to \mathbb{R} \cup \{-\infty\}$ such that

$$\varphi(x) = \inf_y [c(x, y) - \xi(y)]$$

(2) For a c-concave function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ its *c-transform* $\varphi^c : Y \to \mathbb{R} \cup \{-\infty\}$ relative to X and Y is given by

$$\varphi^c(y) = \inf_{x \in X} [c(x, y) - \varphi(x)].$$

Similar, we define the c-transform $\psi^c : X \to \mathbb{R} \cup \{-\infty\}$ of $\psi : Y \to \mathbb{R} \cup \{-\infty\}$.

(3) The *c-subdifferential* $\partial_c \varphi$ of a c-concave function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ relative to X and Y is the set

$$\partial_c \varphi = \{(x, y) \in X \times Y : \varphi(x) + \varphi^c(y) = c(x, y)\}.$$

In the following we often omit the phrase *relative to* if the meaning is clear from the context.

REMARK 0.23. (i) The pair (φ, φ^c) satisfies $\varphi(x) + \varphi^c(y) \leq c(x, y)$.

(ii) The c-subdifferential $\partial_c u \subset X \times Y$ of a function u is c-monotone.
(iii) If φ is c-concave, then $(\varphi^c)^c = \varphi$, or

$$\varphi(x) = \inf_{y \in Y} \left[c(x,y) - \varphi^c(y)\right].$$

EXAMPLE 0.24. (i) Consider $X = Y = \mathbb{R}^n$ and let $c(x,y) = |x-y|$. Then a function $\varphi : X \to \mathbb{R}$ is c-concave if and only if it is 1-Lipschitz, and $\varphi^c = -\varphi$.
(ii) Consider $X = Y = \mathbb{R}^n$ and let $c(x,y) = -x \cdot y$. Then a function $\varphi : X \to \mathbb{R}$ is c-concave if and only if it is concave.

Exercise: Proof the statements in the previous remark and the previous example.

THEOREM 0.25 (Varadarajan). *Let Y be a Polish space, and let $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to Y$ be i.i.d. random variables such that $(X_i)_*\mathbb{P} = \mu$. Then $\mathbb{P}$-almost surely*

$$\frac{1}{n}\sum_{i=1}^n \delta_{X_i(\omega)} =: \mu_\omega \to \mu \text{ weakly in } \mathcal{P}(X).$$

PROOF. From the law of large number we get that for fixed $f \in C_b(X)$

$$\int_X f d\mu_n = \frac{1}{n}\sum_{i=1}^n f(X_i) \to \mathbb{E}[f(X_i)] = \int f \circ X_i d\mathbb{P} = \int f d\mu \text{ almost surely.}$$

Since X is separable, this implies weak convergence of μ_n almost surely. □

THEOREM 0.26 (Kantorovich). *Assume the cost function c is real-valued and continuous, and there exist $c_X \in L^1(\mu)$ and $c_Y \in L^1(\nu)$ such that $c(x,y) \le c_X(x) + c_Y(y)$. Then*

(1) *There exists a c-concave function φ such that $\varphi \in L^1(\mu)$ and $\varphi^c \in L^1(\nu)$ and*

$$\int_X \varphi d\mu + \int_Y \varphi^c d\nu = \sup_{(\varphi,\psi) \in \mathcal{I}(c)} D(\varphi, \psi) = \mathrm{Cost}(\mu, \nu).$$

(2) *For some coupling $\pi \in \mathrm{Cpl}(\mu,\nu)$ the following statements are equivalent.*
  (i) $\pi \in \mathrm{OptCpl}(\mu,\nu)$,
  (ii) *π is c-monotone,*
  (iii) *$\mathsf{spt}\pi \subset \partial_c\varphi$ for any c-concave φ such that (φ, φ^c) is an optimal pair.*

PROOF. **1.** First, we assume that $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ and $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$.
Then $\pi \in \mathrm{Cpl}(\mu,\nu)$ if $\pi = \sum_{i,j=1}^n a_{i,j}\delta_{(x_i,y_j)}$ such that $\sum_{i=1}^n a_{i,j} = \sum_{j=1}^n a_{i,j} = \frac{1}{n}$. We have

$$\mathsf{spt}\pi = \{(x_i, y_i) : a_{i,j} > 0\}$$

and $\mathrm{Cost}(\pi) = \sum_{i,j=1}^n c(x_i,y_j)a_{i,j}$. Assume, that π is optimal but not c-monotone. Then there exist points $(x_{i_1}, y_{j_1}), \dots, (x_{i_N}, y_{j_N}) \in \mathsf{spt}\pi$ such that

$$\sum_{k=1}^N c(x_{i_k}, y_{i_{\sigma(k)}}) < \sum_{k=1}^N c(x_{i_k}, y_{i_k})$$

for some permutation σ. Then we can modify π in the following way.

$$\widetilde{\pi} = \pi + a\sum_{k=1}^N \left[\delta_{(x_{i_k}, y_{i_{\sigma(k)}})} - \delta_{(x_{i_k}, y_{i_k})}\right]$$

where $a := \min_k \{a_{i_k, j_k}\}$. One can check that $\widetilde{\pi}$ is a coupling of μ and ν such that $\mathrm{Cost}(\widetilde{\pi}) < \mathrm{Cost}(\pi)$ (Exercise).

2. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. By the previous theorem we can find probability measures of the form $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ which converge weakly to μ, and in the same way it holds for ν. A straightforward application of Prohorov's theorem implies that there exists a converging sequence

$\pi_i \in \mathrm{Cpl}(\mu_{k_i}, \nu_{k_i})$, and the limit π is a coupling of μ and ν. From the first part of the proof we know that π_i is c-monotone for all i. Therefore, and since the cost function is continuous, the set

$$\mathcal{C}(N) := \left\{ (x_i, y_i)_{i=1,\dots,N} : \sum_{i=1}^N c(x_i, y_i) \le \sum_{i=1}^N c(x_i, y_{\sigma(i)}) \text{ for any } \sigma \in S_N \right\} \subset (X \times Y)^N$$

is closed and satisfies $\pi_i^{\otimes N}(\mathcal{C}(N)) = 1$. Hence, the weak convergence $\pi_i \to \pi$ implies

$$\limsup_{i\to\infty} \pi_i^{\otimes N}(\mathcal{C}(N)) \le \pi^{\otimes N}(\mathcal{C}(N)).$$

It follows that $\mathsf{spt}\pi^{\otimes N} \subset \mathcal{C}(N)$ for any $N \in \mathbb{N}$, and therefore π is c-monotone.

3. Consider $\pi \in \mathrm{Cpl}(\mu, \nu)$ that is c-monotone. We will construct a c-concave function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ such that $\mathsf{spt}[\pi] =: \Gamma \subset \partial_c\varphi$. Pick a point $(x_0, y_0) \in \Gamma$ and define

$$\varphi(x) := \inf_{m\in\mathbb{N}} \Big\{ \inf \big\{ [c(x_1, y_0) - c(x_0, y_0)] + [c(x_2, y_1) - c(x_1, y_1)] + \dots \\ + [c(x, y_m) - c(x_m, y_m)] : \{(x_i, y_i)\}_{i=1}^m \subset \Gamma \big\} \Big\} \in \mathbb{R} \cup \{-\infty\}$$

If we choose $m = 1$ and $(x_1, y_1) = (x_0, y_0)$, we get $\varphi(x_0) \le 0$. On the other hand, the c-monotonicity implies $\varphi(x_0) \ge 0$. In particular, φ is not just identically $-\infty$. This is the only point where we use the c-monotonicity. By renaming $y_m = y$ the definition of φ immediately gives

$$\varphi(x) = \inf_{y\in Y} \{c(x, y) - \xi(y)\}$$

where $\xi : X \to \mathbb{R} \cup \{-\infty\}$ is defined via

$$\xi(y) := \inf_{m\in\mathbb{N}} \Big\{ \inf \big\{ [c(x_1, y_0) - c(x_0, y_0)] + [c(x_2, y_1) - c(x_1, y_1)] + \dots \\ + [c(x_m, y_{m-1}) - c(x_{m-1}, y_{m-1})] - c(x_m, y) : (x_1, y_1), \dots, (x_{m-1}, y_{m-1}), (x_m, y) \in \Gamma \big\} \Big\}$$

Hence, φ is c-concave. φ is also measurable since it is the supremum of a family of continuous functions.

If $(\bar{x}, \bar{y}) \in \Gamma$ is arbitrary, and if we set $(x_m, y_m) = (\bar{x}, \bar{y})$, then we see that

$$\varphi(x) \le \varphi(\bar{x}) + c(x, \bar{y}) - c(\bar{x}, \bar{y})$$

by the definition of φ. Hence

$$c(\bar{x}, \bar{y}) - \varphi(\bar{x}) \le c(x, \bar{y}) - \varphi(x) \implies c(\bar{x}, \bar{y}) - \varphi(\bar{x}) \le \varphi^c(\bar{y}).$$

Since the converse inequality is always true we obtain $(\bar{x}, \bar{y}) \in \partial_c\varphi$ and $\mathsf{spt}\pi \subset \partial_c\varphi$.

4. We show $\varphi \in L^1(\mu)$ and $\varphi^c \in L^1(\nu)$. This follows from

$$\varphi(x) = \inf [c(x, y) - \xi(y)] \le c_X(x) + \underbrace{c_Y(y_0) - \xi(y_0)}_{<\infty}$$

for $y_0 \in Y$ as in **3.** and since c_X is μ-integrable. Similar we can prove $\varphi^c \in L^1(\nu)$. Finally, we can integrate $\varphi + \varphi^c$ with respect to the c-monotone coupling π between μ and ν. But since the support of π is contained in the c-subdifferential $\partial_c\varphi$ of φ, we get

$$\int \varphi d\mu + \int \varphi^c d\nu = \int c d\pi.$$

From this identity we can draw several consequences. First, the coupling π has to be optimal. Therefore, c-monotonicity of π implies $\mathsf{spt}\pi \subset \partial_c\varphi$ that implies optimality of π. Second, the pair (φ, φ^c) is a maximizer of the dual Kantorovich problem. On the other hand, an optimal coupling has to be c-monotone if the cost function is continuous. This can be seen as follows. Assume the contrary. There are points $(x_i, y_i)_{i=1}^N \subset \operatorname{supp}\pi$ that violate c-monotonicity. This completes the proof of the theorem. $\square$

CHAPTER 3

Optimal transport on Riemannian manifolds

Let (M,g) be connected, complete Riemannian manifold without boundary. Let d_M be the induced intrinsic distance of (M,g_M). Let $X,Y\subset M$ be compact subsets.

LEMMA 0.27. *Let (Z,d_Z) be a metric space such that $\operatorname{diam}_Z<\infty$, and let $X,Y\subset Z$ be compact. Consider the cost function $c=\frac{1}{2}\mathrm{d}_Z: X\times Y\to\mathbb{R}$ and a c-concave function $\varphi: X\to\mathbb{R}\cup\{-\infty\}$ (relative to X and Y). Then, φ is real-valued and Lipschitz continuous.*

PROOF. We observe that c is bounded on $X\times Y$. If $\varphi(x)=-\infty$ for some $x\in X$, then φ^c must be unbounded from above, and consequently we already have $\varphi=-\infty$. Hence, we can assume that $|\varphi|<\infty$. Let $z\in X$. Then for any $\epsilon>0$ there is $y\in Y$ such that $\varphi(z)+\epsilon\geq c(z,y)-\varphi^c(y)$. Since for any $x\in X$ we have $\varphi(x)\leq c(x,y)-\varphi^c(y)$, we obtain

$$\varphi(x)-\varphi(z)=c(x,y)-c(z,y)+\epsilon\leq\frac{1}{2}\operatorname{diam}_Z\mathrm{d}_Z(x,z)+\epsilon.$$

Since $\epsilon>0$ was arbitrary and since we can switch the role of x and z, we obtain the result. $\square$

DEFINITION 0.28. Let $U\subset X$ be an open set.

(i) A function $\varphi: U\to\mathbb{R}$ is *super-differentiable* at $x\in U$ if there exist $p\in TM_x$ such that

$$\varphi(\exp_x(v))\leq\varphi(x)+\langle p,v\rangle_x+o(|v|) \tag{3}$$

for all $v\in TM_x$ such that $\varphi(\exp_x(v))\in U$. In the same way, we can define *sub-differentiability.*

(ii) If (3) holds for $p\in TM_x$, p is a super-gradient of φ at x, and we denote with $\nabla^+\varphi|_x$ the set of super-gradients of φ at x. In the same way, we define sub-differentiability and the set of subgradients $\nabla^-\varphi|_x$ of φ at x.

It is obvious, that if $\nabla^+\varphi|_x\cap\nabla^-\varphi|_x\neq\emptyset$, then φ is differentiable at x.

LEMMA 0.29 (Chain rule). *Let U be open, and let $\varphi: U\to\mathbb{R}$ and $h:\mathbb{R}\to\mathbb{R}$ have supergradients $p\in\nabla^+\varphi|_x$ and $\tau\in\nabla^+h|_{\varphi(x)}$ at $x\in U$ and $\varphi(x)$ respectively. If h is non-decreasing then $h\circ\varphi$ is superdifferentiable at x and has super-gradient $\tau\cdot p\in\nabla^+(h\circ\varphi)|_x$.*

PROOF. Since h non-decreasing,

$$h(\varphi(\exp_x(v))\leq h(\varphi(x)+\underbrace{\langle p,v\rangle_x+o(|v|)}_{=:\delta\sim|v|}))\leq h\circ\varphi(x)+\langle\tau p,v\rangle_x+\underbrace{\tau o(|v|)+o(\delta)}_{o(|v|)}$$

which is the statement. $\square$

PROPOSITION 0.30. *Consider $\varphi: U\to\mathbb{R}$ with $\varphi(x)=\frac{1}{2}\mathrm{d}_M(x,y)=c(x,y)$ for some $y\in M$ and $U\subset M$ open. Then φ is super-differentiable at any $x\in U$, and if $\sigma:[0,1]\to M$ is a geodesic between y and x, then $\dot\sigma(1)\in\nabla^+\varphi|_x$.*

PROOF. Let $x\in U$. There exists $\epsilon>0$ and a neighbourhood W of x such that for all $z\in W$ the map $\exp_z|_{B_\epsilon(0)}$ is a diffeomorphismus onto a open set U_z where $W\subset U_z$. (see Milnor Morse theory, 10.3).

1. Now choose $y\in W$ and a minimal geodesic σ between y and x. We show that φ is differentiable at x and $\nabla\varphi|_x=\dot\sigma(1)$.

Let $\exp_y|_{B_\epsilon(0)} =: F : B_\epsilon(0) \to U_y$ and $v \in TM_x$ such that $|v| < \epsilon$. Then

$$\begin{aligned}
\varphi(\exp_x(v) &= \frac{1}{2}\,\mathrm{d}_M(F(v),y)^2 = \frac{1}{2}\,\mathrm{d}_M(F\circ \underbrace{F^{-1}(\exp_x v)}_{w\in TM_y}, y)^2 = \frac{1}{2}|F^{-1}(\exp_x(v))|_y^2 \\
&= \frac{1}{2}|\underbrace{F^{-1}(\exp_x(0))}_{\dot\sigma(0)} + DF_x^{-1}\underbrace{(DF)_0 v}_{=v} + o(|v|)|_y^2 \\
&= \frac{1}{2}|\dot\sigma(0)|_y^2 + \langle \dot\sigma(0), DF_x^{-1}v\rangle_y + o(|v|_y) \\
&= \frac{1}{2}|\dot\sigma(0)|_y^2 + \langle DF_x^{-1}\underbrace{(DF)_0\dot\sigma(0)}_{=\dot\sigma(1)}, DF_x^{-1}v\rangle_y + o(|v|_y) \\
&= \frac{1}{2}|\dot\sigma(0)|_y^2 + \langle \dot\sigma(1), v\rangle_y + o(|v|_y)
\end{aligned}$$

where we used the Gauss Lemma in the last equality, and the standard identification $T(TM_x)_0 = TM_x$ in the second equality. Application of the chain rule with $h(t) = \sqrt{2t}$ implies that $\mathrm{d}_M(\cdot,y)$ is differentiable in x and $\nabla_x \mathrm{d}_M(x,y) = |\dot\sigma(1)|^{-1}\dot\sigma(1)$ is its gradient at x.

2. Now, we choose $y \in M$ arbitrarily. Consider again a minimal geodesic σ between y and x, and choose z on σ such that $z \in W$. Apply the previous step to $\mathrm{d}_M(\cdot,z)$. It follows

$$\begin{aligned}
\mathrm{d}_M(y,\exp_x v) &\le \mathrm{d}_M(y,z) + \mathrm{d}_M(z,\exp_x v) \\
&\le \mathrm{d}_M(y,z) + \mathrm{d}_M(z,x) + \langle \dot\sigma(1)/|\dot\sigma(1)|, v\rangle + o(|v|).
\end{aligned}$$

Hence $\mathrm{d}_M(y,\cdot)$ is super-differentiable. Again by application of the chain rule to $h(t) = \frac{1}{2}t^2$ and $\varphi = h\circ \mathrm{d}_M(\cdot,y)$, we obtain that φ is super-differentiable with super-gradient $\dot\sigma(1) \in \nabla^+\varphi|_x$. ☐

Lemma 0.31. *Consider $U \subset M$ open such that $\overline{U} =: X$ is compact, and let $Y \subset M$ be compact. Let ψ be c-concave relative to X and Y. Assume ψ is differentiable at $x \in U$. Then $(x,y) \in \partial_c\psi$ relative to X and Y if and only if $y = \exp_x(-\nabla\psi|_x)$.*

Proof. "$\Longrightarrow$": Let $y \in Y$ such that $(x,y) \in \partial_c\varphi$. It follows that

$$c(x,y) - \psi(x) - \psi^c(y) = 0 \le c(z,y) - \psi(z) - \psi^c(y)$$

for all $z \in X$ by definition of c-concavity relative to X and Y. For $z \in X$ we can choose $v \in TM_x$ such that $\exp_x(v) = z$. We define $\varphi(z) = \frac{1}{2}\,\mathrm{d}_M(z,y)^2 = c(z,y)$. Then

$$\begin{aligned}
\varphi(\exp_x v) = c(z,y) &\ge c(x,y) + \psi(z) - \psi(x) \\
&= \varphi(x) - \psi(x) + \psi(x) + \langle \nabla\psi_x, v\rangle + o(|v|)
\end{aligned}$$

Therefore, φ is sub-differentiable at x with sub-gradient $\nabla\psi|_x \in \nabla^-\varphi|_x$. On the other hand, the Hopf-Rinow theorem asserts that there is a minimal geodesic between y and x. Thus, the previous lemma tells us that $\dot\sigma(1) \in \nabla^+\varphi|_x$. Hence, φ is actually differentiable at x with

$$\nabla_x \mathrm{d}_M(\cdot,y)^2|_x = \nabla\psi|_x = \dot\sigma(1)$$

and $y = \sigma(0) = \exp_x(-\nabla\psi|_x)$.

"$\Longleftarrow$": Since ψ^c is Lipschitz by Lemma 0.27 and Y is compact, there exists a $y \in Y$ such that $\inf_y[c(x,y) - \psi^c(y)] = \psi(x)$, and following the previous step we obtain $y = \exp_x(-\nabla\psi_x)$. ☐

Theorem 0.32 (Brenier, McCann). *Let $X = \overline{U}$ and Y be as in the previous proposition. Let $\mu \in \mathcal{P}_2(X)$ be absolutely continuous with respect to vol_g such that $\mathsf{spt}\mu \subset U$, and let ψ be c-concave relative to X and Y. Then $\psi : X \to \mathbb{R}$ is differentiable vol_g-almost everywhere, $x \in X \mapsto \nabla\psi_x$ is Borel measurable, and $T : X \to Y$ with $T(x) = \exp_x(-\nabla\psi_x)$ is a solution of the Monge problem relative to X and Y, for $c = \frac{1}{2}\,\mathrm{d}_M^2$, μ and $\nu := T_*\mu$. We have $\mathrm{Cost}(\mu, T_*\mu) < \infty$. If T' is another solution, then $T' = T$ μ-a.e. .*

PROOF. **1.** By Lemma 0.27 ψ and ψ^c are Lipschitz on X and Y respectively, and therefore integrable with respect to probability measures on X and Y respectively. Rademacher's theorem yields ψ is differentiable vol_g-a.e. on U and T is well-defined and Borel measurable. Hence $T_*\mu \in \mathcal{P}_2(Y)$ is well-defined. (ψ, ψ^c) is a competitive pair

$$\psi(x) + \psi^c(y) \le c(x,y) \tag{4}$$

Integration of (4) with respect to $(\mathrm{id}_X, T)_*\mu \in \mathrm{Cpl}(\mu, T_*\mu)$ yields

$$\int_X \psi d\mu + \int_Y \psi^c dT_*\mu \le \mathrm{Cost}(\mu, T_*\mu) \le \int_X c(x, T(x)) d\mu(x)$$

The previous proposition yields that for all $x \in X$ there exist $y(x) \in Y$ such that $\psi(x) + \psi(y(x)) = c(x, y(x))$, and if ψ is differentiable at x, then $y(x) = \exp_x(-\nabla\psi_x) = T(x)$. Hence, for vol_g-a.e. $x \in X$ (and therefore for μ-a.e. $x \in X$) $\psi(x) + \psi(T(x)) = c(x, T(x))$. Integration with respect to μ yields

$$\infty > \int_X \psi d\mu + \int_X \psi^c dT_*\mu = \int_X c(x, T(x)) d\mu = \mathrm{Cpl}(\mu, T_*\mu)$$

and $T : X \to Y$ is an optimal map between μ and $T_*\mu$.
2. If $S : X \to Y$ is another optimal transportation map ($S_*\mu = T_*\mu$), then $\int_X c(x, S(x)) d\mu(x) = \int_X c(x, T(x)) d\mu(x)$. It follows

$$\psi(x) + \psi^c(S(x)) = c(x, S(x)).$$

But the previous proposition tells us that if ψ is differentiable at x, we get that $S(x) = T(x)$. Hence, $S = T$ μ-a.e. . $\square$

THEOREM 0.33. *Let $\mu \in \mathcal{P}(M)$ be absolutely continuous such that $\mathsf{spt}\mu \subset X$, and let $\nu \in \mathcal{P}(M)$ with $\mathsf{spt}\nu \in Y$. Then, there exists a c-concave function $\psi : X \to \mathbb{R}$ such that $T(x) = \exp_x(\nabla\psi_x)$ is an optimal transportation map between μ and ν, and $\mathrm{Cost}(\mu, \nu) < \infty$. T is unique up to a set of μ-measure zero.*

PROOF. **1.** Clearly, the assumptions of Theorem 0.26 are satisfied. Therefore, it tells us that there is a c-concave function ψ such that the pair (ψ, ψ^c) is optimal in the dual Kantorovich problem. By Lemma 0.27 we know that ψ and ψ^c are real-valued and Lipschitz. It is also clear that ψ and ψ^c are integrable with respect to μ and ν respectively. In particular

$$\mathrm{Cost}(\mu, \nu) = \int_X \psi d\mu + \int_X \psi^c d\nu < \infty.$$

The previous theorem yields that $T(x) = \exp_x(-\nabla\psi_x)$ is an optimal transportation map between μ and $T_*\mu$. We still have to show that $T_*\mu = \nu$.
2. Consider $h \in C_b(Y)$ and $|\epsilon| < 1$. We define $\varphi_\epsilon(y) = \psi^c(y) + \epsilon h(y)$ ($\varphi_0 = \psi^c$), and

$$\psi_\epsilon(x) = \inf_y \left[c(x,y) - \varphi(y) - \epsilon h(y)\right] \tag{5}$$

($\psi_0 = \psi$). Let us fix $x \in U$ where ψ is differentiable. Compactness of Y yields that (5) attains its minimum in y_ϵ where $y_0 = T(x)$. The map $\epsilon \mapsto y_\epsilon$ is continuous. Therefore

$$c(x,y) - \varphi(y) - \epsilon h(y) \ge \psi_\epsilon(x) = c(x, y_\epsilon) - \varphi(y_\epsilon) - \epsilon h(y_\epsilon) \ge c(x, T(x)) - \varphi(T(x)) - \epsilon h(y_\epsilon).$$

If we choose $y = T(x)$, then

$$\psi_\epsilon(x) = c(x, T(x)) - \varphi(T(x)) - \epsilon h(T(x)) + o(|\epsilon|)$$

where $o(|\epsilon|) \to 0$ uniform in x. The pair $(\psi_\epsilon, \varphi_\epsilon)$ is competitive by construction, and $D(\psi_\epsilon, \varphi_\epsilon)$ attains its maximum in $D(\psi, \varphi)$. Hence

$$0 = \frac{d}{d\epsilon} D(\psi_\epsilon, \varphi_\epsilon)|_{\epsilon=0} = \int_X \frac{d}{d\epsilon}\psi_\epsilon|_{\epsilon=0} d\mu + \int_Y h(y) d\nu = -\int_X h(T(x)) d\mu(x) + \int_Y h(y) d\nu(y)$$

Hence, $T_*\mu = \nu$. The uniqueness statement follows precisely as in the previous theorem. $\square$

COROLLARY 0.34 (Stability).

CHAPTER 4

Wasserstein distance

Let X be a Polish space.

DEFINITION 0.35. (i) L^p-Wasserstein space:

$$\mathcal{P}^2(X) = \left\{ \mu \in \mathcal{P}(X) : \int_X \mathrm{d}_X(x_0, x)^p d\mu(x) = M^p_{x_0}(\mu) < \infty \text{ for some } x_0 \in X \right\}.$$

(ii) L^p-Wasserstein distance between $\mu, \nu \in \mathcal{P}^p(X)$:

$$W_p(\mu, \nu) = \left[\inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int_{X^2} \mathrm{d}_X(x, y)^p d\pi(x, y) \right]^{\frac{1}{p}}.$$

REMARKS 0.36. (i) If $M^p_{x_0}(\mu) < \infty$ for some x_0, then $M^p_y(\mu) < \infty$ for all $y \in X$.

(ii) $W_p(\mu, \nu) < \infty$ for all $\mu, \nu \in \mathcal{P}^p(X)$, since

$$W_p(\mu, \nu)^p \leq M^p_{x_0}(\mu) + M^p_{x_0}(\nu).$$

(iii) $W_p \leq W_q$ if $p \leq q \in (0, \infty]$.

(iv) $W_p(\delta_x, \delta_y) = \mathrm{d}_X(x, y)$ for all $x, y \in X$.

(v) Kantorovich duality:

$$W_1(\mu, \nu) = \sup_{\mathrm{Lip}\, \varphi \leq 1} \left(\int \varphi d\mu - \int \varphi d\nu \right).$$

PROPOSITION 0.37. *$(\mathcal{P}^p(X), W_p)$ is a metric space.*

PROOF. We already have that $W_p(\mu, \nu) \in [0, \infty)$ for all $\mu, \nu \in \mathcal{P}^p(X)$.

1. $\mu = \nu \iff W_p(\mu, \nu) = 0$. Assume $W_p(\mu, \nu) = 0$. Consider $h \in C_b(X)$. Then $(h, -h)$ is a competitive pair in the dual Kantorovich problem with cost function $\mathrm{d}^p_X : X^2 \to [0, \infty)$. By Theorem 0.26 it follows that

$$0 = W_p(\mu, \nu) \geq \int h d\mu - \int h d\nu.$$

Hence, $\int h d\nu \geq \int h d\mu$. Replacing h by $-h$ yields equality, and therefore $\mu = \nu$.
On the other hand, if $\mu = \nu$, $\pi = (\mathrm{id}_X, \mathrm{id}_X)_* \mu$ is a coupling. Since $\mathrm{d}_X(x, x) = 0$, it yields $W_p(\mu, \mu) = 0$.

2. $\triangle$-inequality. *Claim: If $\pi \in \mathrm{Cpl}(\mu, \nu)$, then there exist Markov kernels $Q, Q' : X \times \mathcal{B}_X \to [0, 1]$ such that*

$$\pi(A \times B) = \int_A \int_B Q(x, dy) \mu(dx) = \int_B \int_A Q'(y, dx) \nu(dy).$$

A Markov kernel is a map $Q : X \times \mathcal{B}_X \to [0, 1]$ such that $x \mapsto Q(x, B)$ is measurable for any $B \in \mathcal{B}_X$, and such that $Q(x, dy)$ is a probability measure for any $x \in X$.

Now, consider $\mu_1, \mu_2, \mu_3 \in \mathcal{P}^2(X)$, optimal couplings $\pi_{1,2} \in \mathrm{OptCpl}(\mu_1, \mu_2)$ and $\pi_{2,3} \in \mathrm{OptCpl}(\mu_2, \mu_3)$, and Markov kernels $Q_{1,2}, Q'_{1,2}$ and $Q_{2,3}, Q'_{2,3}$ with respect to $\pi_{1,2}$ and $\pi_{2,3}$ as described above. We define a probability measure $\pi_{1,2,3}$ on X^3 as follows:

$$\pi_{1,2,3}(A \times B \times C) = \int_A \int_B \int_C Q_{2,3}(y, dz) Q_{1,2}(x, dy) \mu_1(dx).$$

One can easily check that $(p_{12})_*\pi_{123} = \pi_{12}$, $(p_{23})_*\pi_{123} = \pi_{23}$ and $(p_{13})_*\pi_{123} \in \mathrm{Cpl}(\mu_1, \mu_3)$ where $p_{ij} : X^3 \to X^2$ are the projections to the product of the ith and jth marginal. For instance, we have

$$\pi_{123}(A \times B \times X) = \int_A \int_B \underbrace{\int_X Q_{23}(y, dz)}_{=1} Q_{12}(x, dy)\mu_1(dx) = \int_A Q_{12}(x, B)\mu(dx) = \pi_{12}(A \times B).$$

Then

$$\begin{aligned} W_p(\mu_1, \mu_3) &\leq \left[\int \mathrm{d}_X(x, z)^p d\pi_{13}(x, z)\right]^{\frac{1}{p}} = \left[\int_{X^3} \mathrm{d}_X(x, z)^p d\pi_{123}(x, y, z)\right]^{\frac{1}{p}} \\ &\leq \left[\int_{X^3} (\mathrm{d}_X(x, y) + \mathrm{d}(y, z))^p d\pi_{123}(x, y, z)\right]^{\frac{1}{p}} \\ &\leq \left[\int \mathrm{d}_X(x, y)^p d\pi_{123}\right]^{\frac{1}{p}} + \left[\int \mathrm{d}_X(y, z)^p d\pi_{123}\right]^{\frac{1}{p}} \\ &\leq \left[\int \mathrm{d}_X(x, y)^p d\pi_{12}\right]^{\frac{1}{p}} + \left[\int \mathrm{d}_X(y, z)^p d\pi_{23}\right]^{\frac{1}{p}} = W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3). \end{aligned}$$

3. $W_p(\mu, \nu) = W_p(\nu, \mu)$ comes from the symmetry of $\mathrm{d}_X(x, y)$. □

LEMMA 0.38. *If $(\mu_k)_{k\in\mathbb{N}} \subset \mathcal{P}^p(X)$ is a Cauchy sequence w.r.t. W_p, then $\{\mu_k\}_{k\in\mathbb{N}}$ is tight.*

PROOF. Since $W_p \geq W_1$ for $p \geq 1$, we assume (μ_k) is Cauchy w.r.t. W_1. It is clear that for all $\epsilon > 0$ there is $N \in \mathbb{N}$ such that $\{\mu_k\} \subset \bigcup_{j=1}^N B^{W_1}_{\epsilon^2}(\mu_j)$. $\{\mu_j\}_{j=1,\dots N}$ is tight since finite. Hence, there exists a compact set $K \subset X$ such that $\mu_j(K^c) < \epsilon$ for all $j = 1, \dots N$. We can also find points $x_1, \dots, x_m \in X$ s.t. $K \subset \bigcup_{i=1}^m B_\epsilon(x_i) =: U$.. Consider $U_\epsilon = B_\epsilon(U)$ and $\varphi(x) = \max(1 - \frac{1}{\epsilon}\mathrm{d}_X(x, U), 0)$. φ is $\frac{1}{\epsilon}$-Lipschitz. It follows

$$\begin{aligned} \mu_k(U_\epsilon) &\geq \int \varphi d\mu_k \\ &= \underbrace{\int \varphi d\mu_j}_{\geq \mu_j(K)} + \underbrace{\int \varphi d\mu_k - \int \varphi d\mu_j}_{\geq -\frac{1}{\epsilon}W_1(\mu_k, \mu_j)} \geq 1 - \epsilon - \epsilon \end{aligned}$$

for any $k \in \mathbb{N}$ and $j \in \{1, \dots, N\}$.

If choose $\epsilon = 2^{-p}\epsilon$ in the previous proof with $p \in \mathbb{N}_0$, we obtain U^p open such that

$$\mu_k(U^p) \geq 1 - 2^{-p+1}\epsilon \text{ for all } k \in \mathbb{N}.$$

Now, we define $S = \bigcap_{p=0}^\infty \overline{U^p}$ that is closed and satisfies

$$\mu_k(X \backslash S) = \mu_k(\bigcup_{p=0}^\infty \overline{U^p}^c) \leq \sum_{p=0}^\infty \mu_k(\overline{U^p}^c) \leq \sum_{p=0}^\infty 2^{-p+1}\epsilon = \epsilon.$$

S is totally bounded (Exercise), closed subset of complete metric space and therefore compact. This proves that $\{\mu_k\}$ is tight. □

THEOREM 0.39. *$(\mu_k)_{k\in\mathbb{N}} \subset \mathcal{P}^2(X) \to \mu$ w.r.t. W_p if and only if $(\mu_k)_{k\in\mathbb{N}} \to \mu$ weakly and*

$$\limsup_{k\to\infty} \int_X \mathrm{d}_X(x, x_0)^p d\mu_k \leq \int_X \mathrm{d}_X(x, x_0)^p d\mu$$

for some $x_0 \in X$.

PROOF. ”$\Rightarrow$“ The W_p is lower semi-continuous w.r.t. weak convergence. To see that we consider $(\mu_k), \mu, \nu \in \mathcal{P}^2(X)$ such that $(\mu_k) \to \mu$ weakly. Let π_k, π be optimal couplings between μ_k, μ and ν, respectively. Then $(\pi_k) \to \pi$ weakly. The claim follows since the total cost Cost is lower semi-continuous on $\mathcal{P}(X)$.

Let (μ_k) converge to μ w.r.t. W_p. (μ_k) is tight by the previous lemma. Hence, a subsequence (μ_{k^i}) converges weakly to $\widetilde{\mu}$. Hence

$$W_p(\widetilde{\mu}, \mu) \le \liminf_{i \to \infty} W_p(\mu_{k^i}, \mu) = 0$$

and consequently $\widetilde{\mu} = \mu$. We remind on the following elementary estimate. For all $\epsilon > 0$ we can find $C_\epsilon > 0$ such that

$$(a+b)^p \le (1+\epsilon)a^p + C_\epsilon b^p$$

Let $\pi_k \in \mathrm{OptCpl}(\mu, \mu_k)$. Hence

$$\underbrace{\int \mathrm{d}_X(x_0, x)^p d\pi_k}_{M^p_{x_0}(\mu_k)} \le (1+\epsilon) \underbrace{\int \mathrm{d}_X(x_0, y)^p d\pi_k}_{M^p_{x_0}(\mu)} + C_\epsilon \underbrace{\int \mathrm{d}_X(x, y) d\pi_k}_{W_p(\mu, \mu_k)^p}.$$

If $k \to \infty$, we obtain

$$\limsup_{k \to \infty} M^p_{x_0}(\mu_k) \le (1+\epsilon) M^p_{x_0}(\mu).$$

for $\epsilon > 0$ arbitrary.

”$\Leftarrow$“ (Sketch): Assume (μ_k) converges weakly to μ, and $M^p_{x_0}(\mu) \ge \limsup M^p_{x_0}(\mu_k)$. Let π_k be an optimal coupling between μ_k and μ. By Prohorov's theorem $\{\mu_k : k \in \mathbb{N}\}$ and μ are tight. Then by a previous lemma $\{\pi_k : k \in \mathbb{N}\}$ is also tight. Then, up to extraction of sub π_k converges weakly to an optimal coupling π between μ with itself. Hence, π concentrated on $\{(x, x)\}$. Then we can show

$$W_p(\mu, \mu_k) = \int \mathrm{d}_X(x, y) \wedge R d\pi_k + \int |\, \mathrm{d}_X(x, y)^p - R^p| d\pi_k.$$

... $\square$

THEOREM 0.40. *$(\mathcal{P}_p(X), W_p)$ is complete and separable.*

PROOF. **1.** Consider a Cauchy sequence (μ_k). By the previous lemma (μ_k) is tight. Hence, there is a subsequence (μ_{k^i}) converging to $\mu \in \mathcal{P}(X)$ weakly. As before, we check that

$$M^p_{x_0}(\mu) \le \liminf M^p_{x_0}(\mu_{k^i}) < \infty.$$

Since W_p is lower semi-continuous w.r.t. weak convergergence and since (μ_k) is a Cauchy sequence, we have

$$W_p(\mu, \mu_k) \le \liminf_{i \to \infty} W_p(\mu_{k^i}, \mu_k) \to 0 \text{ if } k \to \infty.$$

Henc $\mu \in \mathcal{P}^2(X)$ and $\mu_k \to \mu$ w.r.t. W_p.

2. Let $\mathcal{D} \subset X$ be dense and countable. We will show that

$$\left\{ \nu \in \mathcal{P}^p(X) : \nu = \sum_{i=1}^N a_i \delta_{x_i} \text{ where } x_i \in \mathcal{D}, a_i \in \mathbb{Q}, N \in \mathbb{N} \right\} = \mathcal{M}.$$

Since $\mu \in \mathcal{P}^p(X)$, there is $K \subset X$ compact such that

$$\int_{X \setminus K} \mathrm{d}_X(x_0, x)^p d\mu < \epsilon^p.$$

This follows since we can apply Prohorov's theorem to the finite measure $\mathrm{d}_X(x_0,\cdot)^p d\mu$. We can find finitely many points $\{x_i\}_{i=1,\dots,N} \subset \mathcal{D}$ such that $K \subset \bigcup_{i=1}^N B_\epsilon(x_i)$. Define $U_k = B_\epsilon(x_k)\backslash \bigcup_{i=1}^{k-1} B_\epsilon(x_i)$ and $f : X \to X$ via

$$f(x) = \begin{cases} x_k & \text{if } x \in U_k \\ x_0 & \text{if } x \in X\backslash K. \end{cases}$$

f is measurable and a transport map between $\mu \in \mathcal{P}^2(X)$ and $f_*\mu = \sum_{k=1}^N \mu(\widetilde{B}_k \cap K)\delta_{x_k}$. Additionally

$$W_p(\mu, f_*\mu)^p \le \int_K \underbrace{\mathrm{d}(x,f(x))^p}_{<\epsilon^p} d\mu + \int_{X\backslash K} \mathrm{d}_X(x,f(x))^p d\mu \le 2\epsilon^p.$$

Finally, we can approximate the elements of $\left\{\mu(\widetilde{B}_k \cap K)\right\}$ by rational numbers arbitrarily close to obtain a $\nu \in \mathcal{M}$ that is close to μ w.r.t. W_p. □

1. Geometric properties

First, let (Y, d_Y) some arbitrary metric space.

Definition 1.1. (i) We call a continuous map $\gamma : [a,b] \to Y$ a parametrized curve. The length of a curve γ is

$$\mathrm{L}(\gamma) = \sup_{a=t_0\le\dots\le t_N=b} \sum_{k=1}^N \mathrm{d}_X(\gamma(t_k), \gamma(t_{k+1})) \in [0,\infty].$$

We say that a curve γ is rectifiable if $\mathrm{L}(\gamma) < \infty$.

(ii) A metric space (Y, d_Y) is a length space if for all $x, y \in Y$

$$\mathrm{d}_X(x,y) = \inf\{\mathrm{L}(\gamma) : \gamma : [a,b] \to Y \text{ such that } \gamma(a) = x, \gamma(b) = y\}.$$

(iii) A length space (Y, d_Y) is a geodesic space if for all $x, y \in Y$ there exist a curve $\gamma : [a,b] \to Y$ such that $\mathrm{L}(\gamma) = \mathrm{d}_X(x,y)$ and each such minimizer is called geodesic between x and y. The set of all geodesics is denoted by $\mathcal{G}(Y)$.

Remark 1.2. Each rectifiable curve admits a "constant speed reparametrization". More precisely, there exist $\varphi : [a,b] \to [0, \mathrm{L}(\gamma)]$ continuous and non-decreasing, and $\widetilde{\gamma} : [0, \mathrm{L}(\gamma)] \to Y$ such that $\gamma = \widetilde{\gamma} \circ \varphi$ and $\mathrm{L}(\widetilde{\gamma}|_{[a,t]}) = (t-a)c$ for some $c > 0$.

Proof (Sketch). Consider $\varphi(t+a) = \mathrm{L}(\gamma|_{[a,t]})$. □

In the following we will always assume $[a,b] = [0,1]$, and curves are parametrized by constant speed.

Proposition 1.3. *Let (Y, d_Y) be a complete metric space. Then, (Y, d_Y) is a length (geodesic) space if and only for all $x, y \in Y$ and for all $\epsilon > 0$ (for $\epsilon = 0$) there is $z(\epsilon) \in Y$ such that*

$$\mathrm{d}_Y(x,z), \mathrm{d}_Y(z,y) \le \frac{1}{2}\mathrm{d}_Y(x,y) + \epsilon.$$

We say z is an ϵ-midpoint between x and y.

Theorem 1.4. *If (Y, d_Y) is a length space that is complete and locally compact, then (Y, d_Y) is geodesic.*

Proof (Sketch). From the assumptions it follows that $\overline{B}_R(x) = \{y : \mathrm{d}_Y(x,y) \le R\}$ compact for all $x \in Y$ and all $R > 0$. Then the theorem of Arzela-Ascoli implies the result. □

Theorem 1.5. *Let $(X, \mathrm{d}_X, \mathrm{m}_X)$ be a complete metric measure space. Then, X is a length space if and only if $\mathcal{P}^p(X)$ with W_p is a length space.*

PROOF. "$\Leftarrow$": We already know that X isometrically embed in $\mathcal{P}^p(X)$. Then, if $x, y \in X$, we can find $\mu \in \mathcal{P}^p(X)$ such that

$$\frac{1}{2}\,\mathrm{d}_X(x,y) + \epsilon \geq \left[\int_X \mathrm{d}_X(x,z)^p d\mu(z)\right]^{\frac{1}{p}}, \left[\int_X \mathrm{d}_X(z,y)^p d\mu(z)\right]^{\frac{1}{p}}.$$

Then, there has to be $z_0 \in \mathsf{spt}\mu$ that is ϵ-midpoint between x and y. This can be seen as follows. Imagine there is no such point. Then there is $\epsilon > 0$ such that

$$\frac{1}{2}\,\mathrm{d}_X(x,y) + \epsilon < \mathrm{d}_X(x,z) \text{ and } \frac{1}{2}\,\mathrm{d}_X(x,y) + \epsilon < \mathrm{d}_X(z,y)$$

for all $z \in X$. Integration with respect to μ yields a contradiction.
"$\Rightarrow$": Let $\mu, \nu \in \mathcal{P}^p(X)$ and $\epsilon > 0$. Choose $\widetilde{\mu} = \sum_{i=1}^N a_i \delta_{x_i}, \widetilde{\nu} = \sum_{j=1}^K b_j \delta_{y_j} \in \mathcal{D}$ that are ϵ-close to μ and ν w.r.t. W_p. Let $\pi = \sum_{i,j=1} \pi_{i,j}\delta_{(x_i,y_j)}$ be an optimal coupling between $\widetilde{\mu}$ and $\widetilde{\nu}$. Since X is a length space, we can find for all i, j an $\widetilde{\epsilon}$-midpoint $z_{i,j}$ of x_i and y_j for $\epsilon > 0$. Then, we define

$$\widetilde{\mu}_{\frac{1}{2}} = \sum_{i,j} \pi_{i,j}\delta_{z_{i,j}} \in \mathcal{P}^p(X).$$

We can check that

$$\widetilde{\pi}_{\frac{1}{2}} = \sum_{i,j} \pi_{i,j}\delta_{(x_i,z_{i,j})} \in \mathrm{Cpl}(\widetilde{\mu}, \widetilde{\mu}_{\frac{1}{2}}).$$

It follows that

$$W_p(\widetilde{\mu}, \widetilde{\mu}_{\frac{1}{2}}) \leq \left[\sum_{i,j} \mathrm{d}_X(x_i, z_{i,j})^p \pi_{i,j}\right]^{\frac{1}{2}} \leq \left[\sum_{i,j} \underbrace{(\frac{1}{2}\,\mathrm{d}_X(x_i,y_j) + \epsilon)^p}_{\frac{1}{2}\,\mathrm{d}_X(x,y)^p + c\epsilon^p} \pi_{i,j}\right]^{\frac{1}{2}} \leq \frac{1}{2} W_p(\widetilde{\mu}, \widetilde{\nu}) + C\epsilon$$

Similar for $\widetilde{\nu}$. □

THEOREM 1.6. (i) *X compact if and only if $\mathcal{P}^p(X)$ is compact.*
(ii) *X locally compact if and only if $\mathcal{P}^p(X)$ is compact.*

PROOF. **(i)** "$\Rightarrow$" This follows, since X embeds isometrically into $\mathcal{P}^p(X)$, and X is closed.
"$\Leftarrow$" Apply Prohorov's Theorem.
(ii) Ambrosio, Gigli, Savaré: Gradient flows, Remark 7.1.3. □

COROLLARY 1.7. *If X is a compact length space, then $\mathcal{P}^p(X)$ is a geodesic space.*

More generally:

THEOREM 1.8. *X is a geodesic space then $\mathcal{P}^p(X)$ is a geodesic space.*

PROOF. First, we remark that there is a measurable map $\varphi : (x,y) \mapsto \gamma_{x,y} \in \mathcal{G}(X)$ such that $\gamma(0) = x$ and $\gamma(1) = y$. This follows from a measurable selection theorem (references will be given). Then, $(e_{\frac{1}{2}}) \circ \varphi$ is measurable as well, and we can consider $(e_{\frac{1}{2}} \circ \varphi)_* \pi = \mu_{\frac{1}{2}} \in \mathcal{P}(X)$ where π is an optimal coupling between $\mu_0, \mu_1 \in \mathcal{P}^p(X)$. $e_t : \mathcal{G}(X) \to X$ denotes the evaluation map $e_t(\gamma) = \gamma(t)$. Then, we can define $((e_i, e_{\frac{1}{2}}) \circ \varphi)_* \pi = \pi_i$ for $i = 0, 1$ that are coupling between μ_i and $\mu_{\frac{1}{2}}$. We compute

$$\begin{aligned}
W_p(\mu_i, \mu_{\frac{1}{2}})^p &\leq \int \mathrm{d}_X(x,y)^p d\pi_i(x,y) \\
&= \int \mathrm{d}_X(x,y)^p d((e_i, e_{\frac{1}{2}}) \circ \varphi)_* \pi(x,y) \\
&= \int \mathrm{d}_X(e_i \circ \varphi(x,y), e_{\frac{1}{2}} \circ \varphi(x,y))^p d\pi(x,y) \\
&= \int 2^{-p}\,\mathrm{d}_X(x,y)^p d\pi(x,y) = 2^{-p} W_p(\mu_0, \mu_1)^p < \infty.
\end{aligned}$$

In particular, it follows that $\mu_{\frac{1}{2}} \in \mathcal{P}^p(X)$, and $\mu_{\frac{1}{2}}$ is a midpoint. $\square$

CHAPTER 5

Ricci curvature bounds for metric measure spaces

DEFINITION 0.9. Let (Y, d_Y) be a general metric space. We consider $f : Y \to \mathbb{R} \cup \{\pm\infty\}$ and $D(f) = \{y \in Y : f(y) < \infty\}$.

(i) We say f is k-convex if $f > -\infty$ and for all $\gamma \in \mathcal{G}(D(f))$ with constant speed parametrization

$$f \circ \gamma(t) \leq (1-t) f \circ \gamma(0) + t f \circ \gamma(1) - \frac{1}{2}(1-t)tk \, \mathrm{d}_X(\gamma(0), \gamma(1))^2. \tag{6}$$

(ii) We say f is weakly k-convex if for all $x, y \in D(f)$ there exists $\gamma \in \mathcal{G}(D(f))$ such that $f \circ \gamma > -\infty$, $\gamma(0) = x$ and $\gamma(1) = y$ and (6) holds. In particular, $D(f)$ is a geodesic space.

DEFINITION 0.10 (Metric measure space). (i) Let (X, d_X) be complete, separable metric space, and let m_X be a locally finite Borel measure. Locally finite means that for all $x \in X$ there exists $\epsilon > 0$ such that $\mathrm{m}(B_\epsilon(x)) < \infty$. Then, we say the triple $(X, \mathrm{d}_X, \mathrm{m}_X)$ is a metric measure space.

(ii) $\mathcal{P}^p(X, \mathrm{m}) = \{\mu \in \mathcal{P}^p(X) \text{ such that } \mu = \rho m \text{ with } \rho : X \to [0, \infty)\}$.

REMARK 0.11. Let (M, g_M) be a Riemannian manifold. Consider the Riemannian distance d_M and the Riemannian volume vol_g. Then $(M, \mathrm{d}_g, \mathrm{vol}_g)$ is a mm space.

DEFINITION 0.12 (Relative entropy). The *relative entropy functional* of a metric measure space $(X, \mathrm{d}_X, \mathrm{m}_X)$ is given by $\mathrm{Ent}_{\mathrm{m}_X} : \mathcal{P}^2(X) \to \mathbb{R} \cup \{\pm\infty\}$, where

$$\mathrm{Ent}_{\mathrm{m}_X}(\mu) = \begin{cases} \lim_{\delta \downarrow 0} \int_{\{\rho > \delta\}} \log \rho d\mu & \text{if } d\mu = \rho d \, \mathrm{m}_X \\ \infty & \text{otherwise .} \end{cases}$$

If $\int_{\{\rho \geq 1\}} \log \rho d\mu < \infty$, then

$$\lim_{\delta \downarrow 0} \int_{\{\rho > \delta\}} \log \rho d\mu = \int \log \rho d\mu \in [-\infty, \infty)$$

or if not $= \infty$.

REMARK 0.13. 1. If $\mathrm{m}_X \in \mathcal{P}^2(X) \implies \mathrm{Ent}_{\mathrm{m}_X} \geq 0$. This follows since by Jensen's inequality

$$\mathrm{Ent}_{\mathrm{m}_X}(\mu) = \int \log(\rho) \rho d\mathrm{m}_X \geq \int \rho d \, \mathrm{m}_X \log \int \rho d \, \mathrm{m}_X = 0.$$

2. If m_X satisfies the growth condition $\int e^{-c \, \mathrm{d}_X(x_0, x)^2} d \, \mathrm{m}_X < \infty$ (GC) then $\mathrm{Ent}_{\mathrm{m}_X} > -\infty$. Consider $\widetilde{\mathrm{m}} = \frac{1}{Z} e^{-C \, \mathrm{d}_X(x_0, x)^2} \, \mathrm{m}_X \in \mathcal{P}^2(X)$. Then

$$\begin{aligned} 0 \leq \mathrm{Ent}_{\widetilde{m}}(\mu) &= \int_X \widetilde{\rho} \log \widetilde{\rho} d\widetilde{\mathrm{m}} = \int \log \widetilde{\rho} d\mu \\ &= \underbrace{\int_X \log \rho d\mu}_{\mathrm{Ent}_{\mathrm{m}_X}(\mu)} - \underbrace{\int \log Z d\mu}_{const} + c \underbrace{\int_X \mathrm{d}_X(x_0, x)^2 d\mu}_{M^2_{x_0}(\mu) < \infty} \end{aligned}$$

DEFINITION 0.14 (Lott, Sturm, Villani). A metric measure space $(X, \mathrm{d}_X, \mathrm{m}_X)$ satisfies the curvature dimension $CD(K, \infty)$ if the relative Entropy $\mathrm{Ent}_{\mathrm{m}_X}$ is weakly K-convex.

REMARK 0.15. If $\mu, \nu \in \mathcal{P}^2(X, \mathrm{m}_X)$ then the condition $CD(K, \infty)$ implies that there is geodesic μ_t in $\mathcal{P}^2(X, \mathrm{m}_X)$ between μ and ν.

THEOREM 0.16. *Let (M, g_M) be a Riemannian manifold. Then*

$$\mathrm{ric}_M \geq K \iff (M, \mathrm{d}_M, \mathrm{vol}_g) \textit{ satisfies } CD(K, \infty).$$

For simplicity we assume M is compact. It follows that we can always assume $X = Y = M$.

REMARK 0.17. Consider $\mu_0, \mu_1 \in \mathcal{P}^2(M, \mathrm{m}_X)$ (otherwise there is nothing to prove) and the corresponding Brenier-McCann map $T_t(x) = \exp_x(-t\nabla\varphi_x)$ that induces the unique L^2-Wasserstein geodesic $(T_t)_*\mu_0 = \mu_t$ in $\mathcal{P}^2(M)$. φ is c-concave function.

THEOREM 0.18. *Let $\varphi : U \to \mathbb{R}$ be a semi-concave function on some open set $U \subset M$, then it admits a Hessian vol_g-almost everywhere in the following sense. More precisely, φ admits a Hessian at $x \in U$ if it is differentiable at x and there is a self-adjoint operator $A : TM_x \to TM_x$ such that*

$$\sup_{v \in \nabla^+ \varphi_{\exp_x u}} \|\mathbb{P}^\gamma_{1\to 0} v - \nabla\varphi_x - Au\| = o(|u|). \tag{7}$$

In particular, one gets

$$\varphi(\exp_x(u)) = \varphi(x) + \langle \nabla\varphi_x, v\rangle + \langle Au, u\rangle + o(|u|). \tag{8}$$

Conversely, (8) also implies (7) again.

PROOF. [**CEMS01**, Theorem 3.10]

PROPOSITION 0.19. *A c-concave function φ is semi-concave on M, and hence admits a Hessian vol_g-almost everywhere in M.*

PROOF. [**CEMS01**, Proposition 3.14]

PROPOSITION 0.20. *Consider $\varphi : M \to \mathbb{R}$ c-concave. Let $x \in M$ be a point such that φ admits a Hessian at x. Then the optimal map T that is induced by φ is differentiable at x. More precisely, there is a map $dT : TM_x \to TM_{T(x)}$ such that*

$$\sup \left\{ \|v - dT(u)\| : (\exp_y v, exp_x u) \in \partial\varphi, |v| = \mathrm{d}_M(y, \exp_y v) \right\} = o(|u|).$$

dT is non-singular for all t.

PROOF. [**CEMS01**, Proposition 4.1]

THEOREM 0.21. *Consider $\mu_0, \mu_1 \in \mathcal{P}^2(M, \mathrm{vol}_g)$ and the induced optimal map T and a corresponding d^2-concave function φ. Set $\mu_i = \rho_i d\,\mathrm{vol}_g$. Then the following Monge-Ampére-type equation holds μ_0-a.e.:*

$$\rho_0(x) = \rho_1(T(x)) \det dT|_x \neq 0.$$

PROOF. [**CEMS01**, Theorem 4.2]

THEOREM 0.22. *Consider μ_0 and μ_1 as in the previous theorem, and the W_2-geodesic μ_t that is induced by the one parameter family of optimal maps T_t between μ_0 and μ_1. Then $\{\mu_t\}_{t\in[0,1]} \subset \mathcal{P}^2(X, \mathrm{vol}_g)$.*

Let $x \in M$ be a point such that the φ admits a Hessian at x. Let $(e_i) \in TM_x$ be an orthonormal frame. The Hessian $A =: \nabla^2\varphi$ of φ induces a bilinear form on TM_x via

$$b(u, v) = \langle Au, v\rangle$$

and we set $A_{i,j} = \langle Ae_i, e_j\rangle$.

REMARK 0.23. We add some general considerations.
Consider x where T is differentiable. Then T_t is differentiable at x for all $t \in (0,1]$. Set $\gamma(t) = T_t(x)$. Let $(e_i)_{i=1,\dots,n}$ be an orthonormal frame in TM_x, and $E_i(t) = \mathbb{P}^{\gamma}_{0\to t} e_i$ where $\mathbb{P}^{\gamma}_{0\to t}$ is the parallel transport along γ. If η such that $\dot\eta(0) = e_i$ we can calculate $DT_t e_i$ explicitely. Let $W : (-\delta,\delta) \to TM$ be a vector field along η such that $W(0) = \nabla\varphi_x$ and $W'(0) = Ae_i$.

$$DT_t|_x e_i = \frac{d}{ds} T_t \circ \eta_i|_{s=0} = \frac{d}{ds}\exp_{\eta_i(s)}(t\nabla\varphi|_{\eta_i(s)})|_{s=0} = \frac{d}{ds}\exp_{\eta_i(s)}(tW_{\eta_i(s)})|_{s=0} =: J_i(t).$$

By construction

$$J_i(t) = \sum_{j=1}^n \langle J_i, E_j\rangle(t) E_j(t) = \sum_{j=1}^n J_{i,j}(t)E_j(t)$$

is Jacobi field. The matrix $J(t) = (J_{i,j}(t))_{i,j=1,\dots,n}$ represents the differential of T_t w.r.t. the orthonormal frame (E_i) along γ. Each J_i satisfies

$$J_i'' + R(J_i,\dot\gamma)\dot\gamma = 0 \text{ with } J_i(0) = e_i \text{ and } J_i'(0) = Ae_i = \sum_j A_{i,j}E_j.$$

Therefore, $J = (J_{i,j})_{i,j=1,\dots,n}$ satisfies

$$J'' + R\cdot J = 0 \text{ with } J(0) = E_n \text{ and } J'(0) = (A_{i,j})_{i,j=1,\dots,n}.$$

where $R = (R_{i,j}(t))_{i,j=1,\dots,n}$ and $R_{i,j}(t) = \langle R(E_i,\dot\gamma)\dot\gamma, E_j\rangle$.

PROPOSITION 0.24. *Consider $x \in M$ as in the previous remark. We set $DT_t|_x = J(t)$ and $\det J(t) = \mathcal{J}(t)$ and $\log\mathcal{J}(t) =: y(t)$. Then*

$$y''(t) + \frac{1}{n}(y'(t))^2 + \mathrm{ric}(\dot\gamma(t)) \le 0$$

where $\gamma(t) = T_t(x)$ is the geodesic between x and $T(x)$. A simple reformulation of that is

$$\left[\mathcal{J}^{\frac{1}{N}}\right]''(t) \le -\frac{K}{N}|\dot\gamma|^2\mathcal{J}^{\frac{1}{N}}$$

where $\mathrm{ric}(\dot\gamma(t)) \ge K$ and $\dim_M \le N$.

PROOF. The differentiation rule for det yields

$$\mathcal{J}'(t) = \mathcal{J}(t)\,\mathrm{tr}(\underbrace{\dot J(t)\cdot J(t)^{-1}}_{=:U(t)})$$

and differentiation of U yields

$$U'(t) = J''(t)J(t)^{-1} - (J'(t)J(t)^{-1})^2 = -R(t) - (U(t))^2.$$

Taking the trace gives us

$$(\mathrm{tr}\,U)'(t) + \mathrm{tr}(U(t)^2) + \mathrm{ric}(t) = 0.$$

Consider $U(0) = J'(0)J(0)^{-1} = A$. Since A is symmetric, the identity $\frac{d}{dt}\left(\langle J_i', J_j\rangle - \langle J_i, J_j'\rangle\right)$ yield that $U(t)$ is symmetric for any $t \in [0,1]$. Hence, we can consider the Hilber-Schmidt inner product for symmetric matrices and the corresponding Cauchy-Schwarz inequality for U and E_n:

$$\mathrm{tr}\,U\cdot E_n = \mathrm{tr}\,U = \langle U, E_n\rangle_{HS} \le \sqrt{n}\,\|U\|_{HS} = \sqrt{\mathrm{tr}\,U^2\,\mathrm{tr}\,E_n^2}.$$

It yields

$$(\mathrm{tr}\,U)'(t) + \frac{1}{n}\,\mathrm{tr}(U(t))^2 + \mathrm{ric}(\dot\gamma(t)) \le 0.$$

Since $(\log\mathcal{J})' = \mathcal{J}'/\mathcal{J} = \mathrm{tr}\,U$, we obtain the first result. An easy computation yields the second one. $\square$

Proof of Theorem 0.16. "$\Longrightarrow$": As before we consider μ_0, μ_1 and the unique W_2-Wasserstein geodesic μ_t that is induced by d^2-concave function φ. For all t μ_t is absolutely continuous w.r.t. vol_g. Hence, we can write $\mu_t = \rho_t d\mu$ for some measurable function $\rho_t \geq 0$. We consider the entropy.

$$
\begin{aligned}
&\mathrm{Ent}(\mu_t) = \int \log \rho_t d\mu_t = \int \log \rho_t d(T_t)_* \mu_0 \\
&= \int \log \rho_t(T_t(x)) d\mu_0(x) \\
&= \int \log\left(\rho_0(x)/\det \mathcal{J}_x(t)\right) d\mu_0(x) \\
&= \int \log \rho_0(x) d\mu_0(x) - \int y_x(t) d\mu_0 \\
&\leq \int \log \rho_0(x) d\mu_0(x) - \underbrace{(1-t)\int y_x(0) d\mu_0(x)}_{=0} + t\int y_x(1) d\mu_0(x) - \frac{1}{2}K(1-t)t \underbrace{\int \mathrm{d}_X(x, T(x)) d\mu_0}_{=W_2(\mu_0,\mu_1)^2} \\
&= (1-t)\int \log \rho_0(x) d\mu_0(x) + t \underbrace{\int \log\left(\rho_0 y_x(1)\right) d\mu_0(x)}_{=\int \log \rho_t d\mu_t} - \frac{1}{2}K(1-t)tW_2(\mu_0,\mu_1)^2 \\
&= (1-t)\,\mathrm{Ent}(\mu_0) + t\,\mathrm{Ent}(\mu_1) - \frac{1}{2}K(1-t)tW_2(\mu_0,\mu_1)^2.
\end{aligned}
$$

Therefore $\mathrm{Ent} = \mathrm{Ent}_{\mathrm{vol}_g}$ is weakly K-convex. $\square$

Remark 0.25. Consider the metric measure space $(\mathbb{R}^n, |\cdot|_2, e^{-|\cdot|_2^2} d\mathcal{L}^n)$. It satisfies $CD(1,\infty)$ for all $n \in \mathbb{N}$.

1. Curvature-dimension condition

Definition 1.1. We define the so-called distortion coefficients for $K \in \mathbb{R}$, $\theta \geq 0$ and $t \in [0,1]$.

$$
\sigma_K^{(t)}(\theta) = \begin{cases} \frac{\sin_K(\theta t)}{\sin_K(\theta)} & \text{if } \theta^2 K < \pi \\ \infty & \text{otherwise,} \end{cases}
$$

where the generalized sin-function $\sin_K$ are given by

$$
\sin_K(s) = \begin{cases} \sin(\sqrt{K}s) & \text{if } K > 0 \\ s & \text{if } K = 0 \\ \sinh(\sqrt{-K}s) & \text{if } K < 0. \end{cases}
$$

We also set $\cos_K = \sin_K'$.

Lemma 1.2. *Let $f : [a,b] \to [0,\infty)$ be continuous, and let $K \in \mathbb{R}$. Then the following statements are equivalent.*

(1) *$f'' \leq -Kf$ in the distributional sense. More precisely, for all $\varphi \in C^\infty((a,b))$*

$$
\int_a^b f\varphi'' dx \leq -K \int_a^b f\varphi dx.
$$

(2) *For all $x, y \in [a,b]$*

$$
f((1-t)x + ty) \geq \sigma_K^{(1-t)}(|x-y|)f(x) + \sigma_K^{(t)}(|x-y|)f(y).
$$

In particular, the case $K = 0$ implies convexity of f.

Definition 1.3. We set $\sigma^{(t)}_{K/N}(\theta) = \sigma^{(t)}_{K,N}(\theta)$. The modified distortion coefficients are defined as follows.

$$\tau^{(t)}_{K,N}(\theta) = t^{1/N}\left[\sigma^{(t)}_{K,N-1}(\theta)\right]^{1-1/N}.$$

We have the conventions $r \cdot \infty$ for $r > 0$, and $(\infty)^\alpha = \infty$ for $\alpha \geq 0$. In particular, if $K > 0$, we have $\tau^{(t)}_{K,1}(\theta) < \infty$ iff $\theta = 0$, and $\tau^{(t)}_{K,1}(\theta) = t$ if $K \leq 0$.

Definition 1.4 (N-Renyi entropy). Given a metric measure space $(X, \mathrm{d}_X, \mathrm{m}_X)$ and $N \in [1, \infty)$, we define the N-Renyi entropy functional with respect to m_X as

$$S_N : \mathcal{P}^2(X) \to [-\infty, 0], \quad S_N(\mu) = -\int_X \rho^{-1/N} d\mu$$

where ρ denotes the density of the absolutely continuous part of μ in the Lebesgue decomposition of with respect m_X.

Definition 1.5 (Curvature-dimension condition). A metric measure space $(X, \mathrm{d}_X, \mathrm{m}_X)$ satisfies the curvature-dimension condition $CD(K,N)$ for $K \in \mathbb{R}$ and $N \in [1, \infty)$ if $S_N > -\infty$ and for all $\mu_0, \mu_1 \in \mathcal{P}^2(X, \mathrm{m}_X)$ there exists a W_2-geodesic $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}^2(X, \mathrm{m}_X)$ and an optimal coupling $\pi \in \mathrm{Cpl}(\mu_0, \mu_1)$ such that for all $N' \geq N$

$$S_{N'}(\mu_t) \leq -\int_{X\times X}\left[\tau^{(t)}_{K,N'}(\mathrm{d}_X)\rho_0(x)^{-\frac{1}{N'}} + \tau^{(t)}_{K,N'}(\mathrm{d}_X)\rho_1(y)^{-\frac{1}{N'}}\right] d\pi(x,y)$$

where $\mu_i = \rho_i d\,\mathrm{m}_X$ for $i = 0, 1$ and $\mathrm{d}_X := \mathrm{d}_X(x,y)$. We say $(X, \mathrm{d}_X, \mathrm{m}_X)$ satisfies the reduced curvature-dimension condition $CD^*(K,N)$ if we replace $\tau^{(t)}_{K,N}(\mathrm{d}_X)$ by $\sigma^{(t)}_{K,N}(\mathrm{d}_X)$.

Remark 1.6. Since $\sigma^{(t)}_{K,N}(\mathrm{d}_X) \leq \tau^{(t)}_{K,N}(\mathrm{d}_X)$, we have $CD(K,N)$ implies $CD^*(K,N)$.

Remark 1.7. The definition implies that the support of μ_i for $i = 0, 1$ is contained in the support of m_X.

Remark 1.8. By definition of $\tau^{(t)}_{K,1}(\theta)$ a connected metric measure space satisfies the condition $CD(K,1)$ for $K > 0$ if and only if it consists of only one point.

Theorem 1.9 (Sturm). *Let (M,g) be a complete Riemannian manifold. Then*

$$(M, \mathrm{d}_M, \mathrm{vol}_g) \text{ satisfies } CD(K,N) \iff \mathrm{ric}_g \geq K \text{ and } \dim_M \leq N.$$

Proof. For simplicity assume (M,g) is compact. We remind on the following facts. For $\mu_0, \mu_1 \in \mathcal{P}^2(X, \mathrm{vol}_g)$ there exist a c-concave function φ such that $T_t(x) = \exp_x(-t\nabla\varphi_x)$ is the optimal map between μ_0 and $(T_t)_*\mu_0 = \mu_t$, and μ_t is the unique geodesic between μ_0 and μ_1 in $\mathcal{P}^2(M)$. T_t and φ have the following properties.

(i) φ admits a Hessiann μ_0-almost everywhere.
(ii) DT_t exists μ_0-a.e. for all t, and $DT_t|_x$ is regular for all t.
(iii) μ_t is absolutely continuous w.r.t. vol_g for all t.
(iv) $f(x) = \det DT_t|x f_t(T_t(x))$ m_0-a.e.

We know that

$$U'(t) + U^2(t) + R(t) = 0$$

where $U(t) = J'(t)J(t)^{-1}$ and $R_{i,j}(t) = \langle R(E_i, \dot\gamma)\dot\gamma, E_j\rangle$. The matrix R has the following form

$$R(t) = \begin{pmatrix} 0 & 0 \\ 0 & \bar{R}(t)\end{pmatrix}$$

where $\bar{R}$ is a $(n-1)\times(n-1)$-matrix. Hence

$$u'_{11} + \sum_{i=1}^n u^2_{1,i} = 0 \implies u'_{11} + u^2_{11} \leq 0. \tag{9}$$

Taking the trace yields

$$\operatorname{tr} U' + \operatorname{tr} U^2 + \operatorname{ric} = 0.$$

where $\operatorname{ric}(t) = \operatorname{ric}(\dot\gamma)$. The Jacobi determinant $\mathcal{J}$ satisfies $\mathcal{J}'/\mathcal{J} = \operatorname{tr} U = u_{11} + \sum_{i=2}^n u_{ii}$. Therefore, if we set $\lambda(t) = \int_0^t u_{11}(s)ds$, we see

$$\mathcal{J}(t) = e^{\lambda(t)} e^{\int_0^t \left[\sum_{i=2}^n u_{ii}(s)\right]ds}.$$

The factor $e^{\int_0^t u_{11}} = L(t)$ describes volume distortion in direction of the transport geodesics. From (9) follows $L''(t) \le 0$. We remove this part and consider

$$\bar{\mathcal{J}}(t) = \mathcal{J}(t)/L(t).$$

We study $\bar{\mathcal{J}}$ in more detail. A straightforward computation yields $\bar{\mathcal{J}}' = \bar{\mathcal{J}} \operatorname{tr} \bar U$ where $\bar U = (U_{ij})_{i,j=2,\dots n}$. Set $\bar y = \log \bar{\mathcal{J}}$. Then

$$\begin{aligned}\bar y''(t) = y''(t) - \lambda''(t) &= -\operatorname{tr} U(t)^2 - \operatorname{ric}(t) - u_{11}'(t) = -\sum_{i,j} u_{ij}^2(t) - \operatorname{ric}(t) + \sum_{i=1}^n u_{1i}(t)^2 \\ &\le -\sum_{i,j=2}^n u_{ij}(t)^2 - \operatorname{ric}(t) = \operatorname{tr} \bar U(t)^2 - \operatorname{ric}(t) \le -\frac{1}{n}\left(\operatorname{tr} U(t)\right)^2 - \operatorname{ric}(t)\end{aligned}$$

Corollary 1.10.

$$\left[\bar{\mathcal{J}}^{\frac{1}{N-1}}\right]'' \le -\frac{K}{N-1}|\dot\gamma|^2 \bar{\mathcal{J}}^{\frac{1}{N-1}}$$

where $\operatorname{ric}(\dot\gamma(t)) \ge K$ *and* $\dim_M \le N$. *By Lemma 1.2 we obtain an integrated inequality for* $\bar{\mathcal{J}}$ *of the form*

$$\bar{\mathcal{J}}(t)^{\frac{1}{N}} \ge \sigma_{K,K}^{(1-t)}(|\dot\gamma|)\bar{\mathcal{J}}(0)^{\frac{1}{N-1}} + \sigma_{K,K}^{(t)}(|\dot\gamma|)\bar{\mathcal{J}}(1)^{\frac{1}{N-1}}$$

Corollary 1.11.

$$\mathcal{J}(t)^{\frac{1}{N}} \ge \tau_{K,N}^{(1-t)}(|\dot\gamma|)\mathcal{J}(0)^{\frac{1}{N}} + \tau_{K,N}^{(t)}(|\dot\gamma|)\mathcal{J}(1)^{\frac{1}{N}}$$

where $\operatorname{ric}(\dot\gamma(t)) \ge K$ *and* $\dim_M \le N$.

Proof.

$$\begin{aligned}\mathcal{J}(t)^{\frac{1}{N}} &= \left(\bar{\mathcal{J}}(t)L(t)\right)^{\frac{1}{N}} = \left(\bar{\mathcal{J}}(t)^{\frac{1}{N-1}}\right)^{\frac{N-1}{N}} \left(L(t)\right)^{\frac{1}{N}} \\ &\ge \left(\sigma_{K,K}^{(1-t)}(|\dot\gamma|)\bar{\mathcal{J}}(0)^{\frac{1}{N-1}} + \sigma_{K,K}^{(t)}(|\dot\gamma|)\bar{\mathcal{J}}(1)^{\frac{1}{N-1}}\right)^{\frac{N-1}{N}} \Big((1-t)L(0) + tL(1)\Big)^{\frac{1}{N}} \\ &\ge \left(\sigma_{K,K}^{(1-t)}(|\dot\gamma|)\bar{\mathcal{J}}(0)^{\frac{1}{N-1}}\right)^{\frac{N-1}{N}} ((1-t)L(0))^{\frac{1}{N}} + \left(\sigma_{K,K}^{(t)}(|\dot\gamma|)\bar{\mathcal{J}}(1)^{\frac{1}{N-1}}\right)^{\frac{N-1}{N}} ((t)L(1))^{\frac{1}{N}} \\ &= \tau_{K,N}^{(1-t)}(|\dot\gamma|)\mathcal{J}(0)^{\frac{1}{N}} + \tau_{K,N}^{(t)}(|\dot\gamma|)\mathcal{J}(1)^{\frac{1}{N}}.\end{aligned}$$

In the second inequality we use Hölder's ineqality for $p = \frac{N}{N-1}$ and $q = N$. □

Now, we can complete the proof of the Theorem. Consider μ_0, μ_1, φ, T_t and $\mu_t = \rho_t d\,\mathrm{vol}_g$ as before. Then

$$\begin{aligned} S_N(\mu_t) &= -\int \rho_t(x)^{-\frac{1}{N}} d\mu_t(x) = -\int \rho_t(T_t(x))^{-\frac{1}{N}} d\mu_0(x) \\ &= -\int (\rho_0/\mathcal{J}_x(t))^{-\frac{1}{N}}\, d\mu_0 = -\int \rho(x)^{-\frac{1}{N}} \mathcal{J}_x(t)^{\frac{1}{N}} d\mu_0 \\ &\le -\int \left[\tau_{K,N}^{(1-t)}(|\dot\gamma|)(\rho_0/\mathcal{J}_x(0))^{-\frac{1}{N}} + \tau_{K,N}^{(t)}(|\dot\gamma|)(\rho_0(x)/\mathcal{J}_x(1))^{-\frac{1}{N}}\right] d\mu_0(x) \\ &= -\int \left[\tau_{K,N}^{(1-t)}(\mathrm{d}_X(x,T_1(x)))(\rho_0(x))^{-\frac{1}{N}} + \tau_{K,N}^{(t)}(\mathrm{d}_X(x,T_1(x)))(\rho_1(T_1(x)))^{-\frac{1}{N}}\right] d\mu_0(x) \\ &= -\int \left[\tau_{K,N}^{(1-t)}(\mathrm{d}_X(x,y))(\rho_0(x))^{-\frac{1}{N}} + \tau_{K,N}^{(t)}(\mathrm{d}_X(x,y))(\rho_1(y))^{-\frac{1}{N}}\right] d\,\underbrace{(\mathrm{id},T_1)_*\mu_0}_{=:q}(x,y) \end{aligned}$$

Since T_1 is an optimal map between μ_0 and μ_1, $q = (\mathrm{id}, T_t)_*\mu_0$ is an optimal coupling. Hence, we verified the curvature-dimension condition $CD(K,N)$. ☐

2. Geometric consequences of the curvature-dimension condition

LEMMA 2.1. *For all $K, K' \in \mathbb{R}$, all $N, N' \in (0,\infty)$, all $t \in [0,1]$ and all $\theta \in (0,\infty)$, it holds that*

$$\sigma_{K,N}^{(t)}(\theta)^N \cdot \sigma_{K',N'}^{(t)}(\theta)^{N'} \ge \sigma_{K+K',N+N'}^{(t)}(\theta)^{N+N'}$$

and, if $N \ge 1$,

$$\tau_{K,N}^{(t)}(\theta)^N \cdot \tau_{K',N'}^{(t)}(\theta)^{N'} \ge \tau_{K+K',N+N'}^{(t)}(\theta)^{N+N'}.$$

PROPOSITION 2.2 (Curved Brunn-Minkowski inequality). *Assume the metric measure space $(X, \mathrm{d}_X, \mathrm{m}_X)$ satifies the condition $CD(K,N)$ for $K \in \mathbb{R}$ and $N \ge [1,\infty)$. Then for all measurable sets $A_0, A_1 \subset X$ with positive mass, we have*

$$\mathrm{m}(A_t)^{\frac{1}{N'}} \ge \tau_{K,N}^{(1-t)}(\Theta)\,\mathrm{m}(A_0)^{\frac{1}{N'}} + \tau_{K,N}^{(t)}(\Theta)\,\mathrm{m}(A_1)^{\frac{1}{N'}},$$

for all $t \in [0,1]$ and $N' \ge N$, where $A_t = \{x \in X : \gamma(t) = x, \gamma \in \mathcal{G}(X), t \in [0,1], \gamma(i) \in A_i$ for $i = 0,1\}$ and

$$\Theta := \begin{cases} \inf_{x\in A_0, y \in A_1} \mathrm{d}_X(x,y) & \text{if } K \ge 0 \\ \sup_{x\in A_0, y\in A_1} \mathrm{d}_X(x,y) & \text{if } K < 0. \end{cases}$$

In particular, if $K \ge 0$,

$$\mathrm{m}(A_t)^{\frac{1}{N'}} \ge (1-t)\,\mathrm{m}(A_0)^{\frac{1}{N'}} + t\,\mathrm{m}(A_1)^{\frac{1}{N'}}.$$

PROOF. First, assume $\mathrm{m}(A_0), \mathrm{m}(A_1) < \infty$ and set $\mu_i = \mathrm{m}(A_i)^{-1}\,\mathrm{m}|_{A_i}$ for $i = 0,1$. The curvature-dimension yields

$$\int_{A_t} \rho_t^{\frac{1}{N'}} d\mu_t \ge \tau_{K,N}^{(1-t)}(\Theta)\,\mathrm{m}(A_0)^{\frac{1}{N'}} + \tau_{K,N}^{(t)}(\Theta)\,\mathrm{m}(A_1)^{\frac{1}{N'}}$$

where $(\mu_t = \rho_t d\,\mathrm{m}_X)_t$ denotes the absolutely continuous geodesic that connects μ_0 and μ_1. By Jensen's inequality the left hand side of the previous inequality is smaller than $\mathrm{m}_X(A_t)^{\frac{1}{N'}}$. The general case follows by approximation of A_i by sets of finite measure. ☐

DEFINITION 2.3 (Minkowski content). Consider $x_0 \in X$ and $B_r(x_0) \subset X$. Set $v(r) = \mathrm{m}_X(\bar B_r(x_0))$. The Minkowski content of $\partial B_r(x_0)$ (the r-sphere around x_0) is defined as

$$s(r) := \limsup_{\delta\to 0} \frac{1}{\delta}\,\mathrm{m}_X(\bar B_{r+\delta}(x_0)\backslash B_r(x_0)).$$

Theorem 2.4 (Bishop-Gromov volume growth inequality). *Assume* $(X, \mathrm{d}_X, \mathrm{m}_X)$ *satisfies* $CD(K,N)$ *for* $K \in \mathbb{R}$ *and* $N \in [1,\infty)$. *Then, each bounded set has finite measure and either* m_X *is by one point or all points and all sphere have mass* 0.

More precisely, if $N > 1$ *then for each* $x_0 \in \operatorname{supp} \mathrm{m}_X$ *and for all* $0 < r < R \le \pi\sqrt{\frac{N-1}{\max\{K,0\}}}$, *we have*

$$\frac{s(r)}{s(R)} \ge \frac{\sin^{N-1}_{K/(N-1)} r}{\sin^{N-1}_{K/(N-1)} R}$$

and

$$\frac{v(r)}{v(R)} \ge \frac{\int_0^r \sin^{N-1}_{K/(N-1)} t dt}{\int_0^R \sin^{N-1}_{K/(N-1)} t dt}.$$

If $N = 1$ *and* $K \le 0$, *then*

$$\frac{s(r)}{s(R)} \ge 1, \qquad \frac{v(r)}{v(R)} \ge \frac{r}{R}.$$

Proof. If $K > 0$ and $N = 1$, then the definition of $CD(K,N)$ implies that the support of m_X consist of just one point, and nothing is to prove.

Let us prove the other cases. Fix a point $x_0 \in \operatorname{supp} \mathrm{m}_X$, assume $\mathrm{m}_X(x_0)$ and put $t = \frac{r}{R} \in (0,1)$. Choose $\epsilon > 0$ and $\delta > 0$. We apply the curved Brunn-Minkowski inequality to $A_0 = B_\epsilon(x_0)$ and $A_1 = \bar{B}_{R+\delta R}(x_0) \backslash B_R(x_0)$. One verifies easily that

$$A_t \subset \bar{B}_{r+\delta r+\epsilon r/R}(x_0) \backslash B_{r-\epsilon r/R}(x_0) \text{ and } R - \epsilon \le \Theta \le R + \delta R + \epsilon.$$

Hence, the curved Brunn-Minkowski inequality implies that

$$\begin{aligned}\mathrm{m}_X(\bar{B}_{r+\delta r+\epsilon r/R}(x_0) \backslash B_{r-\epsilon r/R}(x_0))^{\frac{1}{N}} \ge &\tau^{(1-r/R)}_{K,N}(\Theta)\, \mathrm{m}_X(B_\epsilon(x_0))^{\frac{1}{N}} + \\ &\tau^{(r/R)}_{K,N}(\Theta)\, \mathrm{m}_X(\bar{B}_{R+\delta R}(x_0) \backslash B_R(x_0))^{\frac{1}{N}}.\end{aligned}$$

If $\epsilon \to 0$, it yields

$$\mathrm{m}_X(\bar{B}_{r+\delta r}(x_0) \backslash B_{r-}(x_0))^{\frac{1}{N}} \ge \tau^{(r/R)}_{K,N}(R \pm \delta R)\, \mathrm{m}_X(\bar{B}_{R+\delta R}(x_0) \backslash B_R(x_0))^{\frac{1}{N}}.$$

or equivalently

$$v(r+\delta r) - v(r) \ge \tau^{(r/R)}_{K,N}(R \pm \delta R)^N \left(v(R+\delta R) - v(R)\right). \tag{10}$$

Since we assume that m_X is locally finite, the left hand side (and therefore also the right hand side) of the previous inequality is finite for r sufficiently small. Then, $v(R)$ is finite for all $R > 0$, and $v(R) = v(R^*)$ for $R \ge R^* = \pi\sqrt{(N-1)/\max\{K,0\}}$ since $\tau^{(t)}_{K,N}(\theta) = \infty$ if $\theta \ge R^*$ by definition. Moreover, v is right continuous by construction. It is non-decreasing, and therefore it has only countably many discontinuities. In particular, there will be arbitrarily small $r > 0$ and $\delta > 0$ such that v is continuous on the intervall $[r, (1+\delta)r])$. Hence, (10) implies that v is continuous on $(0,\infty)$. Therefore, $\mathrm{m}_X(\partial B_r(x_0)) = 0$ for all $r > 0$, and also $\mathrm{m}_X(\{x\}) = 0$ for all $x \neq x_0$. (10) can be restated as

$$\frac{1}{\delta r}\left(v(r+\delta r) - v(r)\right) \ge \frac{1}{\delta R}\left(v(R+\delta R) - v(R)\right) \frac{\sin_{K/(N-1)}((1\pm\delta)r)^{N-1}}{\sin_{K/(N-1)}((1\pm\delta)R)^{N-1}}.$$

In the limit this yields the first claim.

We will show that v is locally Lipschitz contiuous in R. Let $r, \delta > 0$ as before. If we consider a diadic subdiffision of $[r, r+\delta)$, one can show that for any $n \in \mathbb{N}$ there is $r_n \in [r, r+\delta r)$ such that

$$0 < \frac{2^n}{\delta r}\left(v(r_n + 2^{-n} r) - v(r_n)\right) \le \frac{1}{\delta r}\left(v(r+\delta r) - v(r)\right) =: C.$$

It follows with (10)

$$v(R+2^{-n}\delta R)-V(R)\le \underbrace{\frac{R}{r_n}\left(\frac{\sin_{K/(N-1)}((1\pm\delta)R)}{\sin_{K/(N-1)}((1\pm\delta)r_n)}\right)^{N-1}}_{\le\tau_{K,N}^{(r/R)}((1\pm\delta)R)^{-(N-1)}} v(r_n+2^{-n}\delta r)-v(r_n)\le C2^{-n}\delta R$$

This implies that v is locally Lipschitz. For instance, consider $r<R<R'$ such that $R'-R=\epsilon$ is small. Then we can choose n such that $v(R+2^{-n}\delta R)-v(R')<\frac{1}{k}$. Hence, v is differentiable almost everywhere in R. Hence, for a.e. R the limit $s(R)$ exists and s is the weak differential of v. The fundamental theorem of calculus implies that

$$\int_0^R s(r)dr=v(R).$$

By Gromov's lemma this implies the volume estimate. $\frac{\sin_{K/(N-1)}((1\pm\delta)r)^{N-1}}{\sin_{K/(N-1)}((1\pm\delta)R)^{N-1}}$

Now, we treat case where $\mathrm{m}_X(\{x_0\})>0$. If $\operatorname{supp}\mathrm{m}_X\setminus\{x_0\}\neq\emptyset$, there has to be $x_1\in\operatorname{supp}\mathrm{m}_X\setminus\{x_0\}$ with $\mathrm{m}_X(x_1)=0$. Otherwise, pick any $x_1\neq x_0$. The curvature dimension condition implies the existence of a geodesic $\gamma_t\subset\operatorname{supp}\mathrm{m}_X$ between x_0 and x_1. Since m_X is locally finite, this yields a contradiction. Hence, we repeat the previous proof for x_1 instead of x_0. This implies $\mathrm{m}_X(\{x_0\})=0$ what is a contradiction again. Hence, there was no $x_1\neq x_0$ and $X=\{x_0\}$. $\square$

Bibliography

[CEMS01] Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Invent. Math. **146** (2001), no. 2, 219–257. MR 1865396 (2002k:58038)