$CD^e(K,N)$ SPACES AND STABILITY UNDER GROMOV-HAUSDORFF CONVERGENCE.

These notes were born out of the Spring 2022 course "Optimal Transportation, geometry, and dynamics" held at the Fields Institute and taught by Robert McCann.

Contents

1. The Wasserstein Space. 2
  1.1. The Wasserstein metric. 2
    1.1.1. Disintegration of measures. 3
  1.2. The topology induced by d_p. 5
  1.3. Metric Geometry. 8
  1.4. Displacement convexity. 10
2. Ricci Curvature Bounds in general Polish (Geodesic) Spaces. 11
  2.1. Review of Riemannian Geometry. 11
  2.2. (K,N)-convexity. 13
  2.3. $CD^e(K,N)$ spaces. 16
    2.3.1. Examples of $CD^e(K,N)$ spaces. 16
3. Applications of $CD^e(K,N)$ Spaces. 20
  3.1. The Generalized Brunn-Minkowski Inequality. 20
    3.1.1. Corollaries of the Generalized Brunn-Minkowski Inequality. 22
  3.2. The Bishop-Gromov Inequality. 23
    3.2.1. Corollaries of the Bishop-Gromov Inequality. 25
4. Convergence of Metric Spaces. 27
  4.1. The Hausdorff distance. 27
  4.2. The Gromov-Hausdorff distance. 28
  4.3. Alternative representations of the Gromov-Hausdorff distance. 29
    4.3.1. Metric extensions. 29
    4.3.2. Correspondences. 30
  4.4. Gromov-Hausdorff and isometry classes. 32
  4.5. Reductions to finite sets. 33
  4.6. Stability of uniformly totally bounded classes of compact metric spaces. 34
  4.7. Pointed Gromov-Hausdorff convergence and stability. 35
5. Stability of $CD^e(K,N)$ Spaces. 36
  5.1. Pre-compactness of $CD^e(K,N)$. 36
  5.2. Approximate inverses and limits of geodesics. 37
  5.3. Lifting results to probability measures. 38
  5.4. Behavior of the entropy along limits. 39
  5.5. Measured Gromov-Hausdorff and stability for compact $CD^e(K,N)$ spaces. 41
  5.6. Regularizing kernels and approximating measures with continuous densities. 43
  5.7. Stability for boundedly compact $CD^e(K,N)$ spaces. 45
6. Additional Properties of $CD^e(K,N)$ spaces. 46
  6.1. Building new $CD^e(K,N)$ spaces from old ones. 46
  6.2. Local to global property of $CD^e(K,N)$ spaces. 46
  6.3. Tangent cones and $RCD(K,N)$ spaces. 47
Appendix A. Selected Homework Problems. 48
  A.1. Problem 1 [Riemannian geodesics and distance] 48
  A.2. Problem 2 [Converse of Theorem] 53
  A.3. Problem 3 [New $CD^e(K,N)$ spaces from old] 55
  A.4. Problem 4 [Γ-convergence] 58
References 58

1. The Wasserstein Space.

1.1. The Wasserstein metric.

Definition 1.1 (The p-Wasserstein distance)**.** Let (X,d) be a Radon space. Given $\mu^\pm \in \mathcal{P}(X)$ then

$$d_p(\mu^+,\mu^-) := \left(\min_{\gamma\in\Gamma(\mu^+,\mu^-)} \int_{X\times X} d(x,y)^p \, d\gamma(x,y) \right)^{1/p}$$

is a metric on $\mathcal{P}(X)$ if X is compact, and on the set of measures with finite p-th moment

$$\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X) \;\middle|\; \int_X d(x,x_0)^p \, d\mu(x) < \infty \right\}$$

otherwise.

Remark 1.2. The above appears to depend on x_0, but by the triangle inequality if $\tilde{x}_0 \in X$ then

$$\begin{aligned}\int_X d(x,\tilde{x}_0)^p \, d\mu(x) &\leq \int_X [d(x,x_0)+d(\tilde{x}_0,x_0)]^p \, d\mu(x)\\ &\leq 2^{p-1}\int_X d(x,x_0)^p \, d\mu(x) + 2^{p-1} d(\tilde{x}_0,x_0)^p \mu(X) < \infty.\end{aligned}$$

Proof. We verify that d_p is in fact a metric.

i) [Symmetry]: Let $\gamma \in \Gamma(\mu^+,\mu^-)$ and define $\sigma : X\times X \to X\times X$ by $\sigma(x,y) = \sigma(y,x)$. Setting $\overline{\gamma} = \sigma_\#\gamma$ we see that

$$(\pi_1)_\#\overline{\gamma} = (\pi_1\circ\sigma)_\#\gamma = (\pi_2)_\#\gamma = \mu^-$$

and similarly $(\pi_2)_\#\overline{\gamma} = \mu^+$. Thus $\overline{\gamma} \in \Gamma(\mu^+,\mu^-)$. In particular,

$$\begin{aligned} d_p(\mu^-,\mu^+) &\leq \left(\int_{X\times X} d(y,x)^p \, d\overline{\gamma}(y,x)\right)^{1/p} = \left(\int_{X\times X} d(\sigma(x,y))^p \, d\gamma(x,y)\right)^{1/p}\\ &= \left(\int_{X\times X} d(y,x)^p \, d\gamma(x,y)\right)^{1/p} = \left(\int_{X\times X} d(x,y)^p \, d\gamma(x,y)\right)^{1/p}\end{aligned}$$

where in the last line we have used the symmetry of d. By taking the infimum over $\gamma \in \Gamma(\mu^+,\mu^-)$ we see that $d_p(\mu^-,\mu^+) \leq d_p(\mu^+,\mu^-)$. Exchanging roles of $\mu^\pm$ we obtain equality.

ii) [Non-negativity]: Since $d(x,y)$ is non-negative and γ is a positive measure the integral

$$\int_{X\times X} d(x,y)^p \, d\gamma(x,y)$$

is non-negative.

iii) [Vanishing only on the diagonal]: If $\mu^\pm \in \mathcal{P}(X)$ are such that $d_p(\mu^+,\mu^-) = 0$, then there exists $\gamma \in \Gamma(\mu^+,\mu^-)$ such that

$$0 = d_p(\mu^+,\mu^-)^p = \int_{X\times X} d(x,y)^p \, d\gamma(x,y).$$

Since $d(x,y) \geq 0$, this implies that $d(x,y) = 0$ for γ-a.e. $(x,y) \in X\times X$. But $d(x,y) = 0$ precisely when $x = y$, so γ is supported on $\mathrm{Diag}(X\times X) = \mathrm{Graph}(\mathrm{Id})$. In particular since γ is supported on the graph of a map, we can write $\gamma = (\mathrm{Id}\times\mathrm{Id})_\#\mu^+$. Then,

$$\mu^- = (\pi_2)_\#\gamma = (\pi_2\circ(\mathrm{Id}\times\mathrm{Id}))_\#\mu^+ = (\mathrm{Id})_\#\mu^+ = \mu^+.$$

iv) [Triangle inequality]: To prove this we'll restrict our attention to compact X. In the generic case we'd have to worry about existence of minimizers and use the finite p-th moment condition to deduce.

The triangle inequality says: Given $\mu^\pm \in \mathcal{P}(X)$, for any $\mu \in \mathcal{P}(X)$ we have

$$d_p(\mu^+,\mu^-) \leq d_p(\mu^+,\mu) + d_p(\mu,\mu^-).$$

In particular, there exist $\gamma_{12} \in \Gamma(\mu^+, \mu)$ and $\gamma_{23} \in \Gamma(\mu, \mu^-)$ such that

$$d_p(\mu^+, \mu^-) \leq \left(\int_{X\times X} d(x,y)^p \; d\gamma_{12}(x,y) \right)^{1/p} + \left(\int_{X\times X} d(y,z)^p \; d\gamma_{23}(y,z) \right)^{1/p}.$$

Thus we need to combine γ_{12} and γ_{23} to construct a competitor $\tilde{\gamma}_{13} \in \Gamma(\mu^+, \mu^-)$ such that

$$\begin{aligned} d_p(\mu^+, \mu^-) &\leq \left(\int_{X\times X} d(x,z)^p \; d\tilde{\gamma}_{13}(x,z) \right)^{1/p} \\ &\leq \left(\int_{X\times X} d(x,y)^p \; d\gamma_{12}(x,y) \right)^{1/p} + \left(\int_{X\times X} d(y,z)^p \; d\gamma_{23}(y,z) \right)^{1/p} \end{aligned}$$

[Note: the $\sim$ here is used to denote the fact that $\tilde{\gamma}_{13}$ may not be the minimizer γ_{13} in the definition of $d_p(\mu^+, \mu^-)$]

To do this we "glue" γ_{12} with γ_{23}. That is, we search for $\tilde{\gamma} \in \mathcal{P}(X \times X \times X)$ such that $(\pi_{12})_\# \tilde{\gamma} = \gamma_{12}$ and $(\pi_{23})_\# \tilde{\gamma} = \gamma_{23}$. Then by defining $\tilde{\gamma}_{13} = (\pi_{13})_\# \tilde{\gamma}$,

$$\begin{aligned} \left(\int_{X\times X} d(x,z)^p \; d\tilde{\gamma}_{13}(x,z) \right)^{1/p} &= \left(\int_{X\times X} d(x,z)^p \; d\tilde{\gamma}(x,y,z) \right)^{1/p} \\ &\leq \left(\int_{X\times X} [d(x,y) + d(y,z)]^p \; d\tilde{\gamma}(x,y,z) \right)^{1/p} \\ &= \|d(x,y) + d(y,x)\|_{L^p(X\times X\times X, d\tilde{\gamma})} \\ &\leq \|d(x,y)\|_{L^p(X\times X\times X, d\tilde{\gamma})} + \|d(y,x)\|_{L^p(X\times X\times X, d\tilde{\gamma})} \\ &= \left(\int_{X\times X} d(x,y)^p \; d\tilde{\gamma}(x,y,z) \right)^{1/p} \\ &\quad + \left(\int_{X\times X} d(y,z)^p \; d\tilde{\gamma}(x,y,z) \right)^{1/p} \\ &= \left(\int_{X\times X} d(x,y)^p \; d\gamma_{12}(x,y) \right)^{1/p} \\ &\quad + \left(\int_{X\times X} d(y,z)^p \; d\gamma_{23}(y,z) \right)^{1/p} \end{aligned}$$

as desired. So, the triangle inequality is a combination of two facts: the ability to glue measures, and the triangle inequality for the L^p norm.

□

1.1.1. *Disintegration of measures.* We now need to talk about the following question, which we took for granted in the proof of Definition 1.1: If we have a marginal on the first two factors (γ_{12}) and a marginal on the third two factors (γ_{23}) with the compatibility condition that the projection onto the second factor is equal ($(\pi_2)_\# \gamma_{12} = (\pi_2)_\# \gamma_{23} = \mu$), does there exist a measure $\tilde{\gamma}$ on the triple space with the correct projections ($(\pi_{12})_\# \tilde{\gamma} = \gamma_{12}$ and $(\pi_{23})_\# \tilde{\gamma} = \gamma_{23}$)?

The answer is yes, and in Villani's first book ([Vil03]) it is the gluing lemma. There is a nice technique to prove this via the disintegration of measures (as known in the probability world, the existence of conditional probabilities).

Here is some intuition: On the unit square $Q = [0,1]^2$, we might have some area measure on it $d\,\mathrm{Area}$. Given a nice measurable set $E \subset Q$ we'd like to say that the "sum of the lengths in E" is just the area of E. In fact this is true, and we have

$$\mathrm{Area}(E) = \int_0^1 \mathrm{Length}(E \cap (\{x\} \times [0,1])) \; dx.$$

This works because we know a priori that Area = Length $\otimes$ Length. But this is not always the case, e.g. what if we decompose Q into radial segments? Then we need to add an additional weight, that is we have rdr instead.

The theory of disintegration of measures answers the following: How do we decompose a measure in a semidirect way, where we have measures on fibers (which may vary from fiber to fiber) and another measure on the space of fibers?

Theorem 1.3 (Disintegration of Measures)**.** *Suppose $(X^{\pm}, d^{\pm})$ are Polish spaces (Radon is enough). Given $\mu^+ \in \mathcal{P}(X^+)$ and $q : X^+ \to X^-$ Borel there exists a unique set of measures $\{\mu_y^+\}_{y\in X^-} \subset \mathcal{P}(X^+)$ such that*

i) If $\mu^- := q_\#\mu^+$ then μ^- a.e. y satisfies $\mu_y^+(q^{-1}(y)) = 1$. That is, the fibers discussed earlier are the level sets of q, and μ_y^+ is supported on the appropriate level set.
ii) For every $B \in \mathscr{B}(X^+)$ the map $y \in X^- \mapsto \mu_y^+(B)$ is μ^--measureable.
iii) For all $f : X^+ \to \mathbb{R}$ Borel,

$$\int_{X^+} f(x)\ d\mu^+(x) = \int_{X^-}\int_{q^{-1}(y)} f(x)\ d\mu_y^+(x)d\mu^-(y).$$

So, $d\mu^+ = d\mu_y^+ \otimes d\mu^-$.

Sketch proof. A full proof is contained in an article by Chang and Pollard. The idea is to look at $q_\#(fd\mu^+)$ where f is arbitrary. Looking at the Radon-Nikodym derivative with respect to μ^- yields

$$d(q_\#(fd\mu^+))(y) = g_f(y)\ d\mu^-(y).$$

The left hand-side is linear in f, so g_f is also linear in f (for every y). The Riesz representation theorem guarantees for μ^--a.e. $y \in X^-$,

$$g_f(y) = \int_{X^+} f(x)\ d\mu_y^+(x).$$

After obtaining these μ_y^+, need to check the above three properties. □

To answer the question earlier, we disintegrate γ_{12} using $q_1(x,y) = y$ to get

$$\gamma_{12}(E) = \int_X \gamma_{12}^y(E)\ d\mu(y)$$

and similarly disintegrate γ_{23} using $q_2(y,z) = y$ to get

$$\gamma_{23}(E) = \int_X \gamma_{23}^y(E)\ d\mu(y).$$

Define $\tilde{\gamma}$ (through the Riesz representation theorem) via the following bounded linear functional

$$\int_{X\times X\times X} f(x,y,z)\ d\tilde{\gamma}(x,y,z) = \int_X \int_{\{y\}\times X}\int_{X\times\{y\}} f(x,y,z)\ d\gamma_{12}^y(x)d\gamma_{23}^y(z)d\mu(y).$$

First note that γ_{12}^y and γ_{23}^y are measures on the product space $X \times X$, but because they are supported on a fiber we can disregard this. Observe now that if $f(x,y,z) = g(x,y)$ then, since γ_{23}^y is a probability measure for μ-a.e. $y \in X$,

$$\int_{X\times X\times X} f(x,y,z)\ d\tilde{\gamma}(x,y,z) = \int_X\int_{X\times\{y\}} g(x,y)\ d\gamma_{12}^y(x)d\mu(y) = \int_{X\times X} g(x,y)\ d\gamma_{12}(x,y).$$

That is, $(\pi_{12})_\#\tilde{\gamma} = \gamma_{12}$. The same argument works for the third two factors.

1.2. **The topology induced by d_p.** Given the definition of d_p, we wonder what topology it metrizes. It is easier to answer this when X is compact, so we will work in this case. Before continuing we have the following easy remark

Remark 1.4. Given $\mu^\pm \in \Gamma(\mu^+,\mu^-)$ the map $p \in [1,\infty) \mapsto d_p(\mu^+,\mu^-)$ is non-decreasing. Essentially, the proof is by Jensen's inequality. E.g., let

$$\gamma_p \in \operatorname*{arg\,min}_{\gamma\in\Gamma(\mu^+,\mu^-)} \int_{X\times X} d(x,y)^p \; d\gamma(x,y).$$

Since γ_p is a valid competitor in the definition of d_1, we have

$$d_1(\mu^+,\mu^-)^p \le \left(\int_{X\times X} d(x,y)\; d\gamma_p(x,y)\right) \le \int_{X\times X} d(x,y)^p \; d\gamma_p(x,y) = d_p(\mu^+,\mu^-)^p.$$

Thus, $d_1(\mu^+,\mu^-) \le d_p(\mu^+,\mu^-)$. The same argument works to show $d_p \le d_{p+\epsilon}$ for all $p \in [1,\infty)$. On the other hand, if $\mathrm{Diam}(X) < \infty$ then

$$\begin{aligned} d_p(\mu^+,\mu^-)^p &\le \int_{X\times X} d(x,y)^p \; d\gamma_1(x,y) \\ &= \mathrm{Diam}(X)^{p-1}\int_{X\times X} d(x,y)\; d\gamma_1(x,y) \\ &= \mathrm{Diam}(X)^{p-1} d_1(\mu^+,\mu^-), \end{aligned}$$

that is $d_p(\mu^+,\mu^-) \le \mathrm{Diam}(X)^{1-1/p} d_1(\mu^+,\mu^-)^{1/p}$.

To summarize, convergence in d_p always implies convergence in d_1. If also X is bounded, then d_p metrizes the same topology as d_1. What topology is this? Recall the following

Definition 1.5 (Narrow convergence)**.** Given $\{\mu_k\}_{k=1}^\infty$ a sequence of Radon measures we say that μ_k *converges narrowly* to the Radon measure μ, denoted $\mu_k \overset{n}{\rightharpoonup} \mu$, if for every $\varphi \in C_b^0(X)$

$$\int_X \varphi \; d\mu_k \to \int_X \varphi \; d\mu$$

Definition 1.6 (Weak-$*$ (or wide[1]) convergence)**.** Given $\{\mu_k\}_{k=1}^\infty$ a sequence of Radon measures we say that μ_k *converges weak-$*$* to the Radon measure μ, denoted $\mu_k \overset{\star}{\rightharpoonup} \mu$, if for every $\varphi \in C_c^0(X)$

$$\int_X \varphi \; d\mu_k \to \int_X \varphi \; d\mu$$

So, the only difference between the two is the space of test functions. In general, narrow convergence is slightly stronger since the space of test functions is larger. However, in some cases they are equivalent, as demonstrated in the following

Theorem 1.7 (c.f. [San15] Theorem 5.8)**.** *If (X,d) is locally and σ-compact, then narrow and wide convergence coincide on $\mathcal{P}(X)$. I.e., convergence against $\varphi \in C_c(X)$ implies convergence against $\varphi \in C_b(X)$ as well.*

Proof. Let $\varphi \in C_b(X)$. Then $\varphi + c \ge 0$ for some $c \in \mathbb{R}$. By σ- and local compactness and Urysohn's lemma, there exist compactly supported cutoffs $0 \le \chi_1 ... \le \chi_n \le ...$ with $\lim_{n\to\infty}\chi_n = 1$. Set $\phi_n = (\varphi + c)\chi_n$. These are non-negative functions converging monotonically to $\varphi + c$. By monotone convergence, and supposing that $\mu_k \overset{\star}{\rightharpoonup} \mu$,

$$\begin{aligned} \int_X \varphi \; d\mu + c\mu(X) &= \int_X (\varphi + c)\; d\mu = \lim_{n\to\infty}\int_X \phi_n \; d\mu = \lim_{n\to\infty}\lim_{k\to\infty}\int_X \phi_n \; d\mu_k \\ &\le \liminf_{k\to\infty}\int_X (\varphi_n + c)\; d\mu_k = \liminf_{k\to\infty}\int_X \varphi\; d\mu_k + c\mu_k(X). \end{aligned}$$

Since all the μ_k and μ are probability measures, the constant c is irrelevant and we have shown

$$\int_X \varphi \; d\mu \le \liminf_{k\to\infty}\int_X \varphi \; d\mu_k.$$

[1] Possibly different, maybe with test functions those which vanish at infinity.

Applying the above argument to $-\varphi$ gives the reverse inequality with a limsup. Combining the two gives the required convergence. □

With this, we can show that d_1 is metrizing this topology.

Theorem 1.8 (Topology of d_1, c.f. [San15] Theorem 5.9)**.** *If $X \subseteq \mathbb{R}^d$ is compact then $\mu_n \overset{\star}{\rightharpoonup} \mu$ if and only* $\lim_{n\to\infty} d_1(\mu_n, \mu) = 0$.

Remark 1.9. In particular, since X is compact, Theorem 1.8 says that d_p metrizes the weak-$*$ topology for any $p \in [1, \infty)$.

Proof. Recall

$$d_1(\mu_k, \mu) = \inf_{\gamma \in \Gamma(\mu_k,\mu)} \int_{X\times X} d(x,y)\ d\gamma(x,y) = \sup_{\mathrm{Lip}(\phi)\leq 1} \left(\int_X \phi\ d\mu_k - \int_X \phi\ d\mu \right)$$

by Kantorovich duality. If $d_1(\mu_k, \mu) \to 0$, then given $\phi \in C_c(X)$ we can mollify it: $\phi_\epsilon = \phi * \eta_\epsilon$, where $\eta_\epsilon = 1/\epsilon^d \eta_1(x/\epsilon)$ and η_1 is some compactly supported smooth function integrating to 1. Then $\phi_\epsilon \to \phi$ uniformly as $\epsilon \to 0$. Since μ_n, μ are probability measures

$$\int_X (\phi_\epsilon - \phi)\ d(\mu_n - \mu) \leq 2\|\phi_\epsilon - \phi\|_{L^\infty(X)}.$$

For every $\delta > 0$ choose $\epsilon > 0$ such that $2\|\phi_\epsilon - \phi\|_{L^\infty(X)} \leq \delta$. Now since each $\phi_\epsilon \in C_c^\infty(X)$, they are in particular Lipschitz. By applying the above,

$$d_1(\mu_n, \mu) \geq \int_X \frac{\phi_\epsilon}{\|D\phi_\epsilon\|_{L^\infty(X)}}\ d(\mu_n - \mu) \to 0$$

as $n \to \infty$. Since the above holds for fixed ϵ, we also have

$$\lim_{n\to\infty} \int_X \phi_\epsilon\ d(\mu_n - \mu) = 0.$$

Also select n large so that, for the previously selected ϵ,

$$\left| \int_X \phi_\epsilon\ d(\mu_n - \mu) \right| < \delta.$$

Then,

$$\left| \int_X \phi\ d(\mu_n - \mu) \right| \leq \left| \int_X \phi\ d(\mu_n - \mu) \right| + 2\|\phi_\epsilon - \phi\|_{L^\infty(X)} \leq 2\delta$$

and since δ was arbitrary we see

$$\lim_{n\to\infty} \int_X \phi\ d(\mu_n - \mu) = 0.$$

Conversely if $\mu_n \overset{\star}{\rightharpoonup} \mu$ let

$$L = \limsup_{n\to\infty} d_1(\mu_n, \mu) = \lim_{k\to\infty} d_1(\mu_{n_k}, \mu)$$

for some subsequence $\{\mu_{n_k}\}_{k=1}^\infty$. We need to show that $L = 0$. For each k select a 1-Lipschitz function ϕ_k such that

$$d_1(\mu_{n_k}, \mu) = \int_X \phi_k d(\mu_{n_k} - \mu).$$

Fix $x_0 \in X$ and set $\tilde{\phi}_k(x) = \phi_k(x) - \phi_k(x_0)$ so that the $\tilde{\phi}_k$ are uniformly bounded. As the μ_{n_k} and μ are probability measures, we can replace ϕ_k in the above with $\tilde{\phi}_k$. Next since $\{\phi_k\}_{k=1}^\infty$ is a collection of equi-Lipschitz functions, by Arzela-Ascoli there exists a further subsequence (still indexed by k) such that $\tilde{\phi}_k \to \tilde{\phi}$ uniformly. Then,

$$d_1(\mu_{n_k}, \mu) = \int_X \tilde{\phi}_k d(\mu_{n_k} - \mu) = \int_X (\tilde{\phi}_k - \tilde{\phi})\ d(\mu_{n_k} - \mu) + \int_X \tilde{\phi}\ d(\mu_{n_k} - \mu).$$

As $\mu_{n_k} \overset{\star}{\rightharpoonup} \mu$ and $\tilde{\phi}$ is bounded, the latter integral converges to 0 as $k \to \infty$. On the other hand, $\tilde{\phi}_k$ converges uniformly to $\tilde{\phi}$, and because the $|\mu_{n_k} - \mu|(X) \leq 2$ the former integral also converges to 0 as $k \to \infty$. □

So, we've seen that d_p metrizes the narrow/weak-$*$ topology on $\mathcal{P}(X)$ whenever $X \subset \mathbb{R}^n$ is compact. It turns out we can also define a d_∞ metric which induces a finer topology.

Definition 1.10 (The d_∞ metric)**.** When X is compact, we can define

$$d_\infty(\mu^+, \mu^-) = \lim_{p\to\infty} d_p(\mu^+, \mu^-).$$

Example 1.11. Consider the curve $t \in [0,1] \mapsto (1-t)\delta_x + t\delta_y$ where $x \neq y \in X \subset \mathbb{R}^n$. This depends continuously on t for $p < \infty$ but not for $p = \infty$. Set $\mu_t = (1-t)\delta_x + t\delta_y$. Then,

$$d_p(\mu_t, \mu_s) = \left(\min_{\gamma\in\Gamma(\mu_t,\mu_s)} \int_{X\times X} |z-w|^p \, d\gamma_p(z,w) \right)^{1/p}$$

We examine the set $\Gamma(\mu_t, \mu_s)$. There are only four possible ways to transfer mass: from x to itself or to y, and from y to itself or x. That is

$$\gamma \in \Gamma(\mu_t,\mu_s) \quad \Leftrightarrow \quad \gamma = c_{x,y}\delta_{(x,y)} + c_{y,x}\delta_{(y,x)} + c_{x,x}\delta_{(x,x)} + c_{y,y}\delta_{(y,y)}.$$

Then,

$$\begin{aligned}(1-t)\delta_x + t\delta_y = \mu_t = (\pi_1)_\#\gamma = (c_{x,y} + c_{x,x})\delta_x + (c_{y,x} + c_{y,y})\delta_y \\ (1-s)\delta_x + s\delta_y = \mu_s = (\pi_s)_\#\gamma = (c_{y,x} + c_{x,x})\delta_x + (c_{x,y} + c_{y,y})\delta_y\end{aligned}$$

By equating coefficients,

$$\begin{aligned} c_{x,y} + c_{x,x} &= 1-t, \qquad c_{y,x} + c_{y,y} = t \\ c_{y,x} + c_{x,x} &= 1-s, \qquad c_{x,y} + c_{y,y} = s. \end{aligned}$$

Hence for $\gamma \in \Gamma(\mu_t,\mu_s)$ we have

$$\begin{aligned}\int_{X\times X} |z-w|^p \, d\gamma(z,w) &= c_{x,y}|x-y|^p + c_{y,x}|y-x|^p + c_{x,x}|x-x|^p + c_{y,y}|y-y|^p \\ &= (c_{x,y} + c_{y,x})|x-y|^p.\end{aligned}$$

There are now two cases. First suppose $s > t$. Then setting $c_{y,x} = \epsilon \in [0, \min\{1-s, t\}]$, we see that $c_{x,y} = \epsilon - t + s$. In particular, $c_{x,y} + c_{y,x} = \epsilon - t + s$ is minimized when $\epsilon = 0$, so that $c_{x,y} + c_{y,x} = s - t$. Similarly by reversing roles of s and t we see that

$$d_p(\mu_t, \mu_s) = |s-t|^{1/p}|x-y|.$$

However, as $p \to \infty$, since $s \neq t \in [0,1]$ it follows that $|s-t|^{1/p} \to 1$. That is,

$$d_\infty(\mu_t, \mu_s) = |x-y|.$$

Remark 1.12. Here is an example of the usefulness of the d_∞ topology, as given in Robert's thesis.[2]

When he was a Ph.D. student, he was studying a model for rotating stars. A star is a certain amount of gas made up of compressible fluid. And in the Newtonian model each particle of gas is attracted to each other by gravity, which wants to make the gas collapse to a point. On the other hand, what it means to be a gas is if you squeeze it then it pushes back. So it obeys some equation of state where the pressure is a function of density. You can integrate $P\,dV$ to get the total amount of internal energy, and ask how do the gravitational effects balance the pressure effects defining the steady-state of the gas.

Maybe you assume the total angular momentum is zero (so the center of mass stays fixed). Then you try and minimize the sum of the gravitational potential energy and the internal energy. Under an assumption of the form of the pressure (so if the pressure is strong enough to prevent gravitational collapse but not so strong that it overwhelms gravity and pushes everything to infinity) then maybe there will be a unique ground state and a unique minimum energy minimizing state and it will be spherically symmetric with densest material in the center. You use rearrangement inequalities (like the Riesz rearrangement) to show this. Depending on the pressure you assume, the density will be zero at some finite distance and the star will probably have fixed boundary.

[2]This problem is how he got into optimal mass transport in the first place!

The problem proposed to him by his advisor Elliot Lieb is the following: what if the star is moving, that is if there is some total angular momentum and the star is spinning around its fixed center of mass. Can you perform this energy minimization? It looks fairly bleak, because one way to satisfy the angular momentum constraint is to take a teaspoon of mass and put it into an orbit far away. If it moves slowly with little kinetic energy, then it will have large angular momentum. By doing that, you can approach the non-rotating minimum energy by letting the mass be smaller and smaller further away. But that doesn't sound so physical.

After grappling with this problem for a little while, the understanding he came to is that there is no global energy minimizing state but you can still have local minima. One example is if the angular momentum is small enough, you can have one connected component. But you can also prescribe a mass ratio and have a binary star. If the mass ratio is unequal enough (depending on the size of the angular momentum) you can construct the binary stars to have a large component and a small component satisfying the angular momentum constraint.

The punchline is the following: It would be favorable if you could evaporate mass from the small component and transfer it to the large component, but in the d_∞ topology that is a discontinuous action. So you can't have local energy minimizers in the d_p topology for $p < \infty$, but you can in the d_∞ sense.

From this, one might conclude that the d_∞ topology being finer is a "feature not a bug". On the other hand it is not a vector space topology.

We've seen that $(\mathcal{P}(X), d_p)$ metrizes narrow convergence. When $X \subset \mathbb{R}^n$ is compact, then $\mathcal{P}(X)$ is closed with respect to narrow convergence and complete with respect to d_p. Moreover, $(\mathcal{P}(X), d_p)$ is separable (hence Polish). Letting $\{x_i\}_{i=1}^\infty \subset X$ be countable and dense, then

$$\bigcup_{k=1}^\infty \left\{ \sum_{i=1}^k t_i \delta_{x_i} \;\middle|\; 0 \le t_i \in \mathbb{Q},\ \sum_{i=1}^\infty t_k = 1 \right\}$$

is countable and dense in $(\mathcal{P}(X), d_p)$. We can view this in light of the Krein-Milman theorem. By Riesz representation, we have that $\mathcal{P}(X) \subset (C_0(X), \|\cdot\|_\infty)^*$, where the topology is the weak-$*$ topology. As X is compact, $\mathcal{P}(X)$ is closed. Moreover $\mathcal{P}(X)$, being the space of probability measures, is a subset of the unit ball. Therefore by Banach-Alaoglu, $\mathcal{P}(X)$ is compact in the weak-$*$ topology. Finally $\mathcal{P}(X)$ is convex, so by Krein-Milman it is the closed convex hull of its extreme points. We now claim that the extreme points of $\mathcal{P}(X)$ are exactly the Dirac measures.

1.3. **Metric Geometry.**

We recall here some basics about metric geometry.

Definition 1.13 (Length and geodesic spaces)**.** Let (X, d) be a metric space. For $\sigma \in C([0,1], X)$ we define

$$\text{Length}(\sigma) = \sup_{k \in \mathbb{N}} \left\{ \sum_{i=1}^k d(\sigma(s_i), \sigma(s_{i-1})) \;\middle|\; 0 = s_0 \le s_1 \le ... \le s_k = 1 \right\}.$$

This induces a new distance

$$\tilde{d}(x_0, x_1) = \inf_{\sigma \in C([0,1],X)} \left\{ \text{Length}(\sigma) \mid \sigma(0) = x_0,\ \sigma(1) = x_1 \right\}.$$

By the triangle inequality we have that $d(x_0, x_1) \le \tilde{d}(x_0, x_1)$. If in fact equality holds we call X a *length space*. Furthermore, if the infimum is attained them $(X, \tilde{d})$ is a *geodesic space*.[3] Any minimizer of the length functional is called a *minimizing geodesic*.

Example 1.14. Let $X = [0,1]$ and $0 < \alpha \le 1$. Define $d^\alpha(x,y) = |x-y|^\alpha$. Then d^α is a metric, and $\tilde{d^1} = d^1$, $\tilde{d^\alpha} \equiv \infty$ for $x \ne y$ and $0 < \alpha < 1$. To see this, first note that since $x, y \in [0,1]$ then $|x-y| \le 1$. In particular, this implies that $d^\alpha(x,y) \ge d^1(x,y)$ and $\tilde{d^\alpha}(x,y) \ge d^1(x,y)$. Now, one can show that for any $x,y,z \in [0,1]$ with $x \le y \le z$ we have

$$d^\alpha(x,y) \le d^\alpha(x,z) + d^\alpha(y,z) \le d^\alpha\left(x, \frac{x+y}{2}\right) + d^\alpha\left(y, \frac{x+y}{2}\right) = 2^{1-\alpha}|x-y|^\alpha.$$

[3]In Robert's definition, geodesic spaces need not be length spaces. I find this definition better.

In particular, this implies that

$$\tilde{d^\alpha}(x_0,x_1) \geq \sum_{i=1}^k d(x_i,x_{i-1}) = k^{1-\alpha}|x_0-x_1|^\alpha, \qquad x_i = x_0 + \frac{i}{k}(x_1-x_0).$$

Taking $k \to \infty$ shows that $\tilde{d^\alpha} \to \infty$ unless $x_0 = x_1$ or $\alpha = 1$. So, (X, d^α) is not a length space unless $\alpha = 1$.

We claim that if $X \subset \mathbb{R}^n$ is compact and convex then $(\mathcal{P}(X), d_p)$ is a geodesic space. Assuming this, we'd like to determine the geodesics in $(\mathcal{P}(X), d_p)$.

Remark 1.15 (A characterization of geodesics). If there exists $\sigma : [0,1] \to X$ such that

$$d(\sigma(s),\sigma(t)) \leq |t-s| d(\sigma(0),\sigma(1)),$$

then σ attains the previous infimum (that is, σ is a geodesic). Since (X,d) is a length space,

$$\begin{aligned}
\mathrm{Length}(\sigma) &\leq \sup_{k\in\mathbb{N}} \left\{ \sum_{i=1}^k |s_i - s_{i-1}| d(\sigma(0),\sigma(1)) \;\middle|\; 0 = s_0 \leq s_1 \leq ... \leq s_k = 1 \right\} \\
&= d(\sigma(0),\sigma(1)) \sup_{k\in\mathbb{N}} \left\{ \sum_{i=1}^k (s_i - s_{i-1}) \;\middle|\; 0 = s_0 \leq s_1 \leq ... \leq s_k = 1 \right\} \\
&= d(\sigma(0),\sigma(1)) = \tilde{d}(\sigma(0),\sigma(1)) \leq \mathrm{Length}(\sigma).
\end{aligned}$$

See A.1 for more information about geodesics, especially in the case of Riemannian manifolds.

With this, we can show the following theorem concerning geodesics in $\mathcal{P}(X)$

Theorem 1.16 (Geodesics in $(\mathcal{P}(X), d_p)$)**.** *If $X \subset \mathbb{R}^n$ is compact and convex, and*

$$\gamma_p \in \operatorname*{arg\,min}_{\gamma\in\Gamma(\mu^+,\mu^-)} \int_{X\times X} |x-y|^p \; d\gamma(x,y)$$

and $\pi^s(x,y) = (1-s)x + sy$ then

$$\mu_s := (\pi^s)_\# \gamma_p$$

is a d_p-geodesic.

In fact, the converse is true too.

Proof. We want to estimate $d_p(\mu_s,\mu_t)$ and prove an inequality like above. To do this, select the competitor $\gamma_p^{s,t} := (\pi^s,\pi^t)_\#\gamma_p \in \Gamma(\mu_s,\mu_t)$. Then by definition,

$$\begin{aligned}
d_p(\mu_s,\mu_t)^p &\leq \int_{X\times X} |z-w|^p \; d\gamma_p^{s,t}(z,w) = \int_{X\times X} |\pi^s(x,y) - \pi^t(x,y)|^p \; d\gamma_p(x,y) \\
&= \int_{X\times X} |(t-s)x - (t-s)y|^p \; d\gamma_p(x,y) = |t-s|^p \int_{X\times X} |x-y|^p \; d\gamma_p(x,y) \\
&= |t-s|^p d_p(\mu^+,\mu^-).
\end{aligned}$$

Taking p-th roots completes the proof. ∎

In particular, we have the following special case

Theorem 1.17 ($p = 2$)**.** *For $\mu^+ \ll \mathcal{H}^n$ recall that $\gamma_2 = (\mathrm{Id} \times Du)_\#\mu^+$ for some convex function $u : X \to \mathbb{R}$. So,*

$$\mu_s = (\pi^s)_\#\gamma_2 = ((1-s)\,\mathrm{Id} + sDu)_\#\mu^+.$$

Then also $\mu_s \ll \mathcal{H}^n$ for all $s \in [0,1)$

1.4. **Displacement convexity.** In $\mathbb{R}^n$, we say that a function is convex if it obeys a certain interpolation inequality on straight line segments – these curves being geodesics. In general, we are interested at looking at the behavior of functions along geodesics. Here, we will discuss this in the space $(\mathcal{P}(X), d_2)$, where $X \subset \mathbb{R}^n$ is convex, and along d_2-geodesics.

Here is an example (the interacting gas model). Let $\mu \in \mathcal{P}(X)$ be the density of a gas contained in X. Mostly, we will be concerned when $d\mu(x) = \rho(x) d\mathcal{H}^n(x)$. Associated to this density is an energy functional:

$$E(\mu) = \mathcal{U}(\mu) + \mathcal{W}(\mu) + \mathcal{V}(\mu), \qquad \mathcal{U}(\mu) = \int_X U(\rho(x))\ dx,$$
$$\mathcal{W}(\mu) = \int_X \int_X W(x-y)\ d\mu(x) d\mu(y), \qquad \mathcal{V}(\mu) = \int_X V(x)\ d\mu(x)$$

where $\mathcal{U}$ measures the internal energy associated with the resistance of gas to compression, $\mathcal{W}$ measures the interaction between different gas particles with respect to a Newtonian potential, and $\mathcal{V}$ measures the interaction with the background. The corresponding integrands U, W, and V are the potentials. Both $\mathcal{U}$ and $\mathcal{V}$ are local while $\mathcal{W}$ is non-local, but $\mathcal{U}$ is nonlinear whereas $\mathcal{V}$ is. $\mathcal{W}$ is quadratic and linear on the product space.

If we want to try to minimize this energy among $\mu \in \mathcal{P}(X)$, in a physical model this would correspond to a stationary state of the gas. We're interested in determining if this is unique or not, and convexity of the energy provides a nice technique for this analysis. But first we need to determine what it means to be convex on $(\mathcal{P}(X), d_2)$.

One way to establish convexity is to look at two measures $\mu_0, \mu_1 \in \mathcal{P}(X)$ and evaporate mass from one to the other: $\mu_s = (1-s)\mu_0 + s\mu_1$. Another way is to look at convexity along geodesics: $\mu_s = (\pi^s)_\# \gamma$ where γ solves the Kantorovich problem between μ_0 and μ_1. The latter seems more natural from a transport point of view, since evaporating mass is a discontinuous process. However, both have utility. We'll focus on the former, and in this case wonder for which U, V, W does the energy E become convex along d_2-geodesics. This is easy to answer for V and W, but more subtle for U.

Lemma 1.18. *If $W : \mathbb{R}^n \to \mathbb{R}$ is (strictly) convex then $\mathcal{W}$ is (strictly) d_2-geodesically convex.*

Remark 1.19. For the strict case, it actually holds unless μ_0 is a translation of μ_1. Since E is translation invariant, if $X = \mathbb{R}^n$ and we take μ and a translate of μ, then the energy remains the same. Moreover, the d_2-geodesic from μ to one of its translate is just a translation. But, W is constant along translations so we cannot have strict convexity. This is a clue that when W is convex you cannot hope to have ordinary linear interpolation convexity for $\mathcal{W}$ because if you take a minimizer and a translate, and shrink one but grow the other, you probably won't have a minimizer anymore.

Proof. Let $\mu_0, \mu_1 \in \mathcal{P}^{ac}(X)$. Recall that a d_2-geodesic from μ_0 to μ_1 takes the form $\mu_s = ((1-s)\operatorname{Id} + sDu)_\# \mu_0$ with u a maximizer in the dual Kantorovich problem (hence convex). Then we compute $\mathcal{W}$ along this curve:

$$\begin{aligned}
\mathcal{W}(\mu_s) &= \int_X \int_X W(z-w)\ d\mu_s(z) d\mu_s(w) \\
&= \int_X \int_X W((1-s)(x-y) + s(Du(x) - Du(y)))\ d\mu_0(x) d\mu_0(y) \\
&\leq \int_X \int_X [(1-s)W(x-y) + sW(Du(x) - Du(y))]\ d\mu_0(x) d\mu_0(y) \\
&= (1-s) \int_X \int_X W(x-y)\ d\mu_0(x) d\mu_0(y) + s \int_X \int_X W(z-w)\ d\mu_1(z) d\mu_1(w) \\
&= (1-s)\mathcal{W}(\mu_0) + s\mathcal{W}(\mu_1).
\end{aligned}$$

The above works when W is convex. Under the strictly convex assumption, we may replace the $\leq$ with $<$ unless $x - y = Du(x) - Du(y)$. Rearranging this gives $x - Du(x) = y - Du(y)$ for $(\mu_0 \otimes \mu_1)$-a.e. $(x, y) \in X \times X$. Since the left hand side is independent of y, while the right hand side is independent of x, this means that it is in fact constant. That is to say, $Du = x + c$ for some $c \in \mathbb{R}$, a translation. ☐

A similar argument holds for $\mathcal{V}$, but without any concern for translations. What about $\mathcal{U}$?

Theorem 1.20. *If U vanishes at the origin and $\lambda \in (0,\infty) \mapsto \lambda^n U(x/\lambda^n)$ is convex and non-increasing, then $\mathcal{U}(\mu)$ is d_2-geodesically convex on $(\mathcal{P}^{ac}(X), d_2)$ when $X \subset \mathbb{R}^n$ is compact*[4] *and convex.*

Example 1.21. We typically take U to be a power. For example, the map $U(s) = s^m/(m-1)$ satisfies the above conditions if and only if $m \geq 1 - 1/n$. To see this, we have that

$$\lambda^n U\left(\frac{1}{\lambda^n}\right) = \lambda^n \left(\frac{1}{(m-1)\lambda^{nm}}\right) = \frac{\lambda^{n(1-m)}}{m-1},$$

which is convex when $n(1-m) \leq 1$, i.e. $m \geq 1 - 1/n$. $1 - m \leq 1 - (1 - 1/n) = 1/n$

Proof. First recall if $(\mu_t)_{t\in[0,1]}$ is a d_2-geodesic in $\mathcal{P}_2(X)$ with $\mu_0 \in \mathcal{P}_2^{ac}(X)$ then in fact $\mu_t \in \mathcal{P}_2^{ac}(X)$ for all $t \in [0,1)$ and $\rho_t := d\mu_t/dx$ satisfies

$$\rho_t(G_t(x)) = \frac{\rho_0(x)}{\det(DG_t(x))}$$

where $G_t(x) = (1-t)x + tDu(x)$. For a proof of this, see [Vil03]. Evaluating $\mathcal{U}(\mu_t)$,

$$\begin{aligned}\mathcal{U}(\mu_t) = \int_X U(\rho_t(y))\; dy &= \int_X U(\rho_t(G_t(x))) \det DG_t(x)\; dx \\ &= \int_X U\left(\frac{\rho_0(x)}{\det DG_t(x)}\right) \det DG_t(x)\; dx \\ &= \int_X f(\rho_0(x), g(t,x))\; dx\end{aligned}$$

where $f(\rho, \lambda) = \lambda^n U(\rho/\lambda^n)$ (for fixed ρ, by assumption this map is convex and non-increasing) and $g(t,x) = \det^{1/n}[(1-t)\,\mathrm{Id} + tD^2u(x)]$ (since $G_t(x) = (1-t)x + tDu$). Recall that $t \in [0,1] \mapsto g(\cdot, x)$ is concave.[5] Hence, $t \in [0,1] \mapsto f(\rho_0(x), g(\cdot,x))$ is convex. ☐

2. Ricci Curvature Bounds in general Polish (Geodesic) Spaces.

2.1. Review of Riemannian Geometry.

Recall in a Riemannian manifold (M,g) if $\sigma(s)$ and $\tau(t)$ are arc-length parametrized geodesics and $\sigma(0) = \tau(0)$ then

$$d_g^2(\sigma(s), \tau(t)) = st\langle\dot\sigma(0), \dot\tau(0)\rangle + Cs^2t^2 + O(|s|^5 + |t|^5)$$

where $d_g(\cdot,\cdot)$ is the distance on X induced by geodesics. The constant C is something like

$$C = \frac{1}{6}\frac{\partial^4}{\partial s^2 \partial t^2} d_g^2(\sigma(s), \tau(t))\bigg|_{s=t=0} = -\frac{1}{3} R_{ijkl}\dot\sigma^i(0)\dot\tau^j(0)\dot\sigma^k(0)\dot\tau^l(0).$$

The number R_{ijkl} is a component of the *Riemann curvature tensor* and is measuring, in some approximate sense, what is the deviation of the Pythagorean theorem for a triangle with side lengths in the direction $\sigma(s)$ and $\tau(t)$. That is, when $R_{ijkl} = 0$ the Pythagorean theorem holds (to leading order).

There are some easy properties we can derive about the R_{ijkl}. First since i,k and j,l are playing symmetric roles, we have $R_{ijkl} = R_{kjil} = R_{klij} = R_{ilkj}$. If $\dot\sigma(0) = \dot\tau(0)$ then $\sigma(s) = \tau(t)$ locally (since the exponential map is a local diffeomorphism, i.e. geodesics are unique for short time). Then, $R_{iiii} = 0$. Now if you try and change $\dot\sigma^i(0)$ with $\dot\tau^i(0)$, but keep the rest the same, something should change – indeed, you get $R_{ijkl} = -R_{jikl}$. These should be enough to recover all the symmetries of the Riemann curvature tensor.

---

[4] We can remove the compactness assumption by working with finite second moments.

[5] For a proof of this see Section 3.1.

Example 2.1. Taking $M^n = \partial B_r^{n+1}(0) \subset \mathbb{R}^{n+1}$,

$$R_{ijkl} = \frac{1}{r^2}(g_{ik}g_{jl} - g_{jk}g_{il})$$

The above property about geodesics with the same velocity being the same locally isn't true in general, so we have the following definition:

Definition 2.2 (Non-Branching)**.** A metric space (X, d) is *non-branching* unless there exist d-geodesics $(x(s))_{s\in[0,1]}$ and $(y(t))_{t\in[0,1]}$ such that $x(s) = y(s)$ for all $s \in [0, 1/2]$ but there exists $t \in (1/2, 1]$ such that $x(t) \neq y(t)$

Remark 2.3. The use of $1/2$ is mostly artificial, what's important is that the geodesics stay together for some amount of time but then drift apart.

Example 2.4. Riemannian manifolds without boundary are non-branching but if there is boundary, they need not be. For example, take $\mathrm{Cl}(\mathbb{R}^2 \setminus B_1(0))$ with the Euclidean metric. Consider $p = (-2, 0)$ and $q = (1, 1)$. In $\mathbb{R}^2$, the geodesic would just be a straight line. However this passes through $B_1(0)$. So, a geodesic in this new space goes in a straight line from p to S^1, then travels around an arc towards q, then in a straight line to reach q. If we choose another point $\tilde{q}$ near q (such that the straight line between p and q is also obstructed by $B_1(0)$) then we'll get the same phenomenon. Both these geodesics follow the same arc of S^1 for some time, but then diverge.

We can also take (partial) traces of the Riemann tensor:

$$\mathrm{Ric}_{ik} = g^{jl}R_{ijkl}$$

where we are using the Einstein summation notation. These are components of the *Ricci tensor*. Notice that the indices we sum over correspond to the same curve – in the previous, j, l both corresponded to $\tau(t)$. Visually, you sum all the contributions of curvatures coming from triangles containing a given direction (here, $\dot{\sigma}(0)$) and directions orthogonal to it (here, $\dot{\tau}(0)$, which we sum over).

What this also tells you is if you take small cone of geodesics around a given one, and you exponentiate (i.e., continue following the geodesics) it, how does the volume change?

Example 2.5. Looking at $M^n = \partial B_r^{n+1}(0) \subset \mathbb{R}^{n+1}$, we have

$$\mathrm{Ric}_{ik} = \frac{1}{r^2}g^{jl}(g_{ik}g_{jl} - g_{jk}g_{il}) = \frac{1}{r^2}(ng_{ik} - g_{jk}\delta_i^j) = \frac{n-1}{r^2}g_{ik}.$$

By writing $K = 1/r^2$ for the sectional curvature, we see that $\mathrm{Ric}_{ik} = (n-1)Kg_{ik}$, the point being the Ricci curvature depends on the dimension by $n - 1$.

We want to look at an analog of this in metric spaces. Since we still have a distance function, one could look at triangles and ask how close they are to satisfying the Pythagorean theorem. This is what the Alexandrov school in Russia did around the 1940s. With this, you can derive sectional curvature bounds (morally, bounds for the Riemann tensor) based on triangles.

But if you want bounds for the Ricci tensor, it's not so clear how to do that in a metric space. Especially in a geodesic space we can construct triangles, but it's not apparent how to average – how do we weight the different triangles? So, we need a little more structure.

Why are we interested in the Ricci curvature anyways? To quote Robert: "As my geometry friends tell me, when you need upper curvature bounds you usually need upper bounds on the sectional curvature, but when you need lower curvature bounds it usually suffices to consider Ricci curvature bounds."

In summary, sectional curvature bounds can be formulated in any geodesic space (X, d), but to define Ricci bounds we require a notion of averaging. This will be given by a reference measure $m \in \mathscr{M}_+(X)$ on a Polish (geodesic) space (X, d).

This is an idea which goes back to Lott-Villani ([LV09]) and Sturm ([Stu06a] and [Stu06b]).[6] Use

$$d_2(\mu,\nu)^2 = \inf_{\gamma\in\Gamma} \int_{X\times X} d(x,y)^2 \; d\gamma(x,y)$$

to metrize the probability measures $\mathcal{P}_2(X)$. I.e., lift the geometry from points $x \in X$ to measures $\mu \in \mathcal{P}_2(X)$. Then you have a good notion of geodesic. Where does the reference measure come in? We wanted to define a notion like the L^p norm of the density of a measure, but in a metric space this isn't so clear (what even is the density of a measure?). We could define this for any p, but we'll take the limit as p goes to 1. Doing so gives the Boltzmann (or Shannon) entropy:

$$S(\mu \mid m) = \begin{cases} \int_X d\mu/dm \log(d\mu/dm)\; dm & \mu \ll m \\ \infty & \text{else} \end{cases}$$

on $\mu \in \mathcal{P}_2(X)$. As for choosing m, as soon as you have a metric space there's a notion of Hausdorff measure for each dimension. And there's precisely one dimension which gives a nontrivial measure. So we could use this Hausdorff measure as m.

Here is their idea: Convexity properties of S along d_2-geodesics in $\mathcal{P}_2(X)$ reflect the "Ricci" curvature properties of (X,d,m). The triple (X,d,m) is called a *metric measure space.* The role of the distance is to define geodesics while the role of optimal transport is to lift the notion of geodesics of points to measures, the reference measure defines the entropy and the convexity properties of the entropy along d_2-geodesics reflect the Ricci curvature.

2.2. (K,N)**-convexity.** When you move from a manifold, there are different notions of convexity, so we have to be a little careful. The following definition clarifies the distinction.

Definition 2.6 (Weak vs strong convexity)**.** Let (X,d) be a metric space. We say that $f : X \to \mathbb{R}$ is *weakly geodesically convex* if for all $x_0, x_1 \in X$ there exists a d-geodesic $\sigma : [0,1] \to X$ with $\sigma(0) = x_0, \sigma(1) = x_1$ such that $f(\sigma(s))$ is geodesic. Moreover, $f : X \to \mathbb{R}$ is *strongly geodesically convex* if the above holds for all such geodesics.

Note that in a geodesic space, strong geodesic convexity implies weak. However, in a space with no geodesics, every function is trivially strongly geodesically convex.

Definition 2.7. We say $(X,d,m) \in CD(0,\infty)$ if S is weakly d_2-geodesically convex.

Example 2.8. If $d = d_g$ on a Riemannian manifold $X = M^n$ and $m = \mathrm{vol}_g$, then $(M^n, d_g, \mathrm{vol}_g) \in CD(0,\infty)$ if and only if $\mathrm{Ric} \geq 0$. A detailed proof of this is given in Villani old & new ([Vil09]). The 0 in $CD(0,\infty)$ gives the lower bound of 0, while the ∞ is a dimensional parameter. We'll see soon that $(M^n, d_g, \mathrm{vol}_g) \in CD(K,N)$ if and only if $\mathrm{Ric} \geq Kg$ and $n \leq N$, where $K \in \mathbb{R}$ and $N \in [1,\infty]$. The motivation for looking at Ricci lower bounds and dimensional upper bounds comes from a pre-compactness theorem of Gromov.

Previously we discussed taking limits of Riemannian manifolds in some abstract sense. One can imagine taking a sequence of manifolds converging, say, to a convex polytope. While not a smooth manifold, it is still close to one in a sense. In what space should they live? In a way, this is like passing from the rationals to the reals.

Definition 2.9 ((K,N)-convexity)**.** For $K \in \mathbb{R}$ and $N \in [1,\infty]$ we say $S : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is (K,N)-*convex* when

i) S is semiconvex and
ii) $S''(r) \geq S'(r)^2/N + K$.

When $N = \infty$ we interpret the second condition as $S''(r) \geq K$.

[6]Both papers came out around the same time, the former just took longer to publish.

Remark 2.10. The larger K is and smaller N is, the stronger this requirement is. Also, the parameters K and N are "fake" in some way, by scaling. Consider replacing S by αS. Then, the second condition reads

$$\alpha S''(r) \geq \frac{\alpha^2}{N} S'(r)^2 + K,$$

and so if N is finite we can take $N = 1$. On the other hand, by scaling r by βr then we get

$$\beta^2 S''(\beta r) \geq \frac{\beta^2}{N} S'(\beta r)^2 + K$$

and so we can always take $K \in \{-1, 0, 1\}$.

As a convexity condition, we wonder if there is an interpolation inequality available.

Theorem 2.11 (c.f. [Vil09] Theorem 14.28)**.** *If $S \in C([0,1]) \cap C^2(0,1)$ then the following are equivalent*

i) *$(K,1)$-convexity, i.e. $S'' - (S')^2 \geq K \in \{-1,0,1\}$.*
ii) *Defining $U(r) := e^{-S(r)}$, then $U'' \leq -KU$.*
iii) *For all $x, y \in \mathbb{R}$ and $t \in [0,1]$,*

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

where

$$\sigma_K^{(t)}(r) = \frac{s_K(tr)}{s_K(r)}, \qquad \text{if } Kr^2 \leq \pi^2$$

and

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

Remark 2.12. Of course, we only consider $(K,1)$ convexity due to scaling of N. However, if we wanted general N then we'd set $U_N(r) = e^{-S(r)/N}$, and replace K with K/N in ii) and iii).

Note also that, regardless of the sign of K,

$$s_K'' + K s_K = 0$$

and is the unique solution with initial conditions $s_K(0) = 0$ and $s_K'(0) = 1$.

Proof. The equivalence between i) and ii) is a special case of the Hopf-Cole transform from PDE. If $U = \phi(S)$ then $U' = \phi' S'$ and $U'' = \phi''(S')^2 + \phi' S''$. Now

$$S'' - (S')^2 \geq K \quad \Leftrightarrow \quad \phi' S'' - \phi'(S')^2 \leq K\phi' \quad \Leftrightarrow \quad U'' - (\phi'' + \phi')(S')^2 \leq K\phi'.$$

We will be done if we can find ϕ such that $\phi'' = -\phi'$, $\phi' = -\phi$, and $\phi' \leq 0$. Choosing $\phi(S) = e^{-S}$ suffices.

We now want to show the equivalence of i) or ii) to iii). One direction is a Taylor expansion (iii) implies ii)) while the other comes from a maximum principle. By Taylor expanding with $\Delta r = r_1 - r_0$,

$$\sigma_K^{(1/2)}(\Delta r) = \frac{s_K(\Delta r/2)}{s_K(\Delta r)} = \frac{\Delta r/2 - K/6(\Delta r/2)^3 + O(|\Delta r|^5)}{\Delta r - K(\Delta r)^3/6 + O(|\Delta r|^5)} = \frac{1}{2}\left(1 + \frac{K}{8}(\Delta r)^2 + O(|\Delta r|^4)\right)$$

$$\frac{1}{2}\left(U(r_0) + U(r_1)\right) = U\left(\frac{r_0 + r_1}{2}\right) + \frac{(\Delta r)^2}{8} U''\left(\frac{r_0 + r_1}{2}\right) + o(|\Delta r|^2)$$

Applying iii) with $t = 1/2$ yields

$$\begin{aligned}
0 &\geq \sigma_K^{(1/2)}(|\Delta r|)[U(r_0)+U(r_1)] - U\left(\frac{r_0+r_1}{2}\right)\\
&= \frac{1}{2}[U(r_0)+U(r_1)] + \frac{K}{16}[U(r_0)+U(r_1)](\Delta r)^2 + O(|\Delta r|^4)\\
&\quad - \frac{1}{2}(U(r_0)+U(r_1)) + \frac{1}{8}U''\left(\frac{r_0+r_1}{2}\right)(\Delta r)^2 + o(|\Delta r|^2)\\
&= (\Delta r)^2\left(\frac{K}{16}[U(r_0)+U(r_1)] + \frac{1}{8}U''\left(\frac{r_0+r_1}{2}\right) + o(1)\right).
\end{aligned}$$

To complete the proof, we select $r \in \mathbb{R}$ and r_0, r_1 such that Δr is small and $r = (r_0+r_1)/2$. By taking $r_0, r_1 \to r$, we see that

$$0 \geq (\Delta r)^2\left(\frac{K}{8}[U(r)] + \frac{1}{8}U''(r) + o(1)\right)$$

i.e., $U''(r) \leq -KU(r)$.

The harder direction is to show ii) implies iii). Set $u(t) = U((1-t)r_0 + tr_1)$. So, ii) becomes $u'' + K(\Delta r)^2 u \leq 0$. Let $v(t)$ solve $v'' + K(\Delta r)^2 v = 0$ with $v(0) = u(0)$ and $v(1) = u(1)$. The point is that the right hand side of iii) saturates the differential inequality – that is, it is just $v(t)$ (after replacing U with u). Thus we aim to show $u(t) \geq v(t)$.

If not, there exists $t_* \in (0,1)$ minimizing $u(t) - v(t)$. Then $0 \leq u''(t_*) - v''(t_*)$. But, by our differential inequality for u and equality for v,

$$0 \leq -K(\Delta r)^2(u(t_*) - v(t_*)),$$

which immediately gives a contradiction if $K < 0$. When $K = 0$ we're just saying that U is concave, which is standard. The case $K > 0$ is a little tricky, but similar to what we just did. For us, we'll only be concerned when u is a positive function anyways so let's just prove it in this case. Again we may take $K \leq \pi^2$ (otherwise it is trivial) and let v_ϵ solve $v_\epsilon'' + (K-\epsilon)(\Delta r)^2 v_\epsilon = 0$ with $v_\epsilon(0) = u(0)$ and $v_\epsilon(1) = u(1)$. If we show $u \geq v_\epsilon$ for all $\epsilon > 0$, then $u \geq v$ as desired. Note also for ϵ sufficiently small we have (as the interval we're looking at is a small interval $[0,1]$) that $v_\epsilon > 0$. Consider then $w_\epsilon(t) = u(t)/v_\epsilon(t)$. If this ratio is minimized at $t \in \{0,1\}$ we're done. Otherwise there exists $t_\epsilon \in \arg\min_{t\in(0,1)} w_\epsilon(t)$ where $w_\epsilon''(t_\epsilon) \geq 0$ and $w_\epsilon'(t_\epsilon) = 0$. Expanding this out gives

$$\begin{aligned}
0 &\leq \left(\frac{u}{v_\epsilon}\right)''(t_\epsilon) = \left(\frac{u'v_\epsilon - uv_\epsilon'}{v_\epsilon^2}\right)'(t_\epsilon)\\
&= \left(\frac{(u''v_\epsilon + u'v_\epsilon' - u'v_\epsilon' - uv_\epsilon'')(v_\epsilon)^2 - (u'v_\epsilon - uv_\epsilon')(2v_\epsilon v_\epsilon')}{v_\epsilon^4}\right)(t_\epsilon)\\
&= \left(\frac{u''v_\epsilon^3 - u(v_\epsilon)^2 v_\epsilon''}{v_\epsilon^4}\right)(t_\epsilon) - \frac{2v_\epsilon'(t_\epsilon)}{v_\epsilon(t_\epsilon)}w_\epsilon'(t_\epsilon) = \left(\frac{u''}{v_\epsilon}\right)(t_\epsilon) - \left(\frac{uv_\epsilon''}{v_\epsilon^2}\right)(t_\epsilon)\\
&= \left(\frac{u'' + K(\Delta r)^2 u}{v_\epsilon}\right)(t_\epsilon) - u(t_\epsilon)\left(\frac{v_\epsilon'' + K(\Delta r)^2 v}{v_\epsilon^2}\right)(t_\epsilon)\\
&\leq -\frac{u(t_\epsilon)(\epsilon(\Delta r)^2 v_\epsilon(t_\epsilon))}{v_\epsilon(t_\epsilon)^2} = -\epsilon(\Delta r)^2 w_\epsilon(t_\epsilon)
\end{aligned}$$

where we have applied the differential properties of u'' and v_ϵ. Since $u, v_\epsilon > 0$, so too is w_ϵ – thus no such t_ϵ exists. □

Remark 2.13. The differential inequality $U'' + KU \leq 0$ is saturated by

$$U(r) = \begin{cases} \cos(r\sqrt{K}) & K > 0\\ r & K = 0\\ \cosh(r\sqrt{-K}) & K < 0 \end{cases}.$$

Accordingly,

$$S(x) = \begin{cases} -N \log \cos\left(x\sqrt{\frac{K}{N}}\right) & K > 0,\ |x| \le \frac{\pi}{2}\sqrt{\frac{N}{K}} \\ -N \log x & K = 0,\ x \in (0,\infty) \\ -N \log \cosh\left(x\sqrt{\frac{-K}{N}}\right) & K > 0,\ x \in \mathbb{R} \end{cases}$$

are (K,N)-convex on the specified intervals. These are the prototypical examples.

2.3. $CD^e(K,N)$ **spaces.** The definition of (K,N)-convexity allows us to define $CD^e(K,N)$ spaces.

Definition 2.14. Fix a Polish space (X,d) with locally- and σ-finite Borel measure $m \ge 0$. We say $(X,d,m) \in CD^e(K,N)$ if for all $\mu_0, \mu_1 \in \mathcal{P}_2(X)$ there exists a d_2-geodesic $(\mu_t)_{t\in[0,1]}$ connecting them along which $t \in [0,1] \mapsto S(\mu_t)$ is $(Kd_2(\mu_0,\mu_1)^2, N)$-convex. We say (X,d) is *strongly* $CD^e(K,N)$ if the above happens for all such d_2-geodesics.

Remark 2.15. We'd like to say that a space is $CD^e(K,N)$ if the entropy is (K,N)-convex along geodesics, but it's not that simple. Remember that rescaling r in the definition of (K,N)-convexity multiplies K by the scaling factor squared. If we parametrized the geodesic by arc length then we'd write (K,N)-convexity above, but unless the geodesic has length 1 we cannot do this and keep it parametrized over $[0,1]$.

Remark 2.16. Definition 2.14 is sometimes called *entropic* $CD(K,N)$, and is denoted $CD^e(K,N)$. The original definition involved a more complicated entropy depending on N. Under reasonable conditions the two are equivalent. This entropic definition comes from Erbar-Kuwada-Sturm ([EKS15]).

Remark 2.17. What if we used d_p instead of d_2, for $1 < p < \infty$? This idea was introduced by Kell. Since all the d_p metrize the same topology, it is reasonable to conclude that these definitions would lead to the same idea. Indeed this is the case, as shown by Robert and Afiny Akdemir et. al. in [ACM+21] under a non-branching hypothesis.

2.3.1. *Examples of* $CD^e(K,N)$ *spaces.*

Example 2.18. The simplest examples are given by intervals. Although the sectional curvature is zero you can still get some Ricci curvature by taking a reference measure which is not the Lebesgue measure. Take Euclidean distance on

a) $|x| < \pi/2\sqrt{(N-1)/K}$ with $dm(x) = \cos(x\sqrt{K/(N-1)})^{N-1}dx$ for $K > 0$ and $N < \infty$;
b) $x > 0$ with $dm(x) = x^{N-1}dx$ for $K = 0$ and $N < \infty$;
c) $x \in \mathbb{R}$ with $dm(x) = \cosh(x\sqrt{-K/(N-1)})^{N-1}dx$ for $K < 0$ and $N < \infty$;
d) $x \in \mathbb{R}$ with $dm(x) = e^{-K/2|x|^2}dx$ for $K \in \mathbb{R}$ and $N = \infty$.

(The power $N-1$ should be reminiscent of the change of variables to polar coordinates in $\mathbb{R}^d$, where we get a factor of r^{d-1}). The following example expands upon this.

Example 2.19. Here are some more interesting examples. Consider $(\mathbb{R}^n, |\cdot|, e^{-\phi(x)}dx)$. For which ϕ is the above a $CD^e(K,N)$ space?

Think about $\phi(x) = |x|^2$, so that we get a Gaussian measure. Transporting across the middle of the Gaussian, the reference measure m is much larger in the middle (with respect to the Euclidean volume case). So, $d\mu/dm$ will be much smaller and S will be more convex. Thus, since we know $(\mathbb{R}^n, |\cdot|, dx) \in CD^e(0,\infty)$ we expect, by changing the reference measure to a Gaussian, for this to hold with $K > 0$. This leads us to the following theorem

Theorem 2.20 (Euclidean $CD^e(K,N)$ spaces)**.** *Fix $(X,d) \subset (\mathbb{R}^n, |\cdot|)$ with X a compact subset of $\mathbb{R}^n$ and $\phi \in C^2(\mathbb{R}^n)$. If $dm(x) = e^{-\phi(x)}dx$ then $(X,d,m) \in CD^e(K,N)$ if $N \ge n$ and either*

i) $N = n$, $K \le 0$, and ϕ is constant;
ii) $N > n$ and

$$D^2\phi(x) - \frac{(D\phi \otimes D\phi)(x)}{N-n} - K\,\mathrm{Id} \ge 0$$

for all $x \in X$.

Remark 2.21. The assumption that X is compact is just to simplify the proof, it can probably be done by working with $\mathcal{P}_2(X)$.

The converse is also true, see A.2.

Proof. We start with a heuristic proof, then make it rigorous. Recall

$$S(\mu \mid m) := \int_X \log\left(\frac{d\mu}{dm}\right) d\mu(x)$$

whenever $\mu \ll m$. Then also by the Radon-Nikodym chain rule,

$$\frac{d\mu}{dm} = \frac{d\mu}{dx}\frac{dx}{dm} = e^{\phi}\frac{d\mu}{dx}.$$

Applying this in the above yields

$$S(\mu \mid m) = \int_X \log\left(e^{\phi}\frac{d\mu}{dx}\right) d\mu(x) = \int_X \left[\phi(x) + \log\left(\frac{d\mu}{dx}\right)\right]\ d\mu(x).$$

Now that we've introduced $d\mu/dx$, we can start using our previously developed theory. In particular, recall if $\mu_0, \mu_1 \in \mathcal{P}_2^{ac}(\mathbb{R}^n)$ are joined by a d_2-geodesic $(\mu_t) \subset \mathcal{P}_2^{ac}(\mathbb{R}^n)$ then there exists $G_t : \mathbb{R}^n \to \mathbb{R}^n$ such that $\mu_t = (G_t)_\# \mu_0$ where

$$G_t(x) = (1-t)x + tDU(x) := x + tDu(x)$$

for some convex function U on $\mathbb{R}^n$. Here, we have defined $u(x) = U(x) - 1/2|x|^2$.

We even have a Monge-Ampere equation giving

$$\rho_t(G_t(x)) := \frac{d\mu_t}{dx}(G_t(x)) = \frac{d\mu_0/dx}{\det DG_t(x)}$$

for μ_0-a.e. x. Then,

$$\begin{aligned} e(t) := S(\mu_t \mid m) &= \int_X [\phi(y) + \log(\rho_t(y))]\ d\mu_t(y) = \int_X [\phi(G_t(x)) + \log(\rho_t(G_t(x)))]\, d\mu_0(x) \\ &= \int_X \log(\rho_0(x))\ d\mu_0(x) + \int_X [\phi(G_t(x)) - \log(\det DG_t(x))]\, d\mu_0(x). \end{aligned}$$

The first integral in the last line is simply $S(\mu_0 \mid dx)$, which is finite.[7] To show $(Kd_2(\mu_0,\mu_1)^2, N)$-convexity, we need to compute $e'(t)$ and $e''(t)$. The term $S(\mu_0 \mid dx)$ is constant with respect to t. Moreover, G_t is actually very smooth (in fact linear) in t. So we shouldn't have much trouble differentiating this, but we need to recall how to compute the derivative of logarithms of determinants:

$$\begin{aligned} \frac{d}{dt}\log(\det DG_t(x)) &= \mathrm{Tr}[DG_t^{-1}(x)DG_t'(x)] \\ \frac{d^2}{dt^2}\log(\det DG_t(x)) &= -\,\mathrm{Tr}[(DG_t^{-1}(x)DG_t'(x))^2] \end{aligned}$$

where DG_t^{-1} denotes the inverse of DG_t and DG_t' denotes the derivative of $d/dtG_t = Du(x)$.

Now, a nice trick is when you look along geodesics and you want to differentiate at t corresponding to some midpoint, that's equivalent to differentiating at 0 of some shorter geodesic [because the two geodesics are affine reparameterizations of each other]. So, even though we want to show $e''(t) \geq e'(t)^2/N - Kd_2(\mu_0,\mu_1)^2$ it's enough to check at $t = 0$. In this case, $G_0(x) = x$ and

[7] Actually Robert claims this is $e(0)$, but I don't think that's true.

$DG_0(x) = \mathrm{Id}$. As before, $G_0'(x) = Du(x)$ and so $DG_0'(x) = D^2u(x) \geq -I$ (this latter inequality follows from the definition of u and that U is convex). Hence we have

$$\frac{d}{dt}\log(\det DG_t(x))\Big|_{t=0} = \mathrm{Tr}[DG_0^{-1}(x)DG_0'(x)] = \mathrm{Tr}[D^2u(x)] = \Delta u(x)$$
$$\frac{d^2}{dt^2}\log(\det DG_t(x))\Big|_{t=0} = -\mathrm{Tr}[(DG_0^{-1}(x)DG_0'(x))^2] = -\mathrm{Tr}[(D^2u(x))^2].$$

Computing $e'(0)$ and $e''(0)$ then yields

$$e'(0) = \int_X \left[\langle D\phi(G_t(x)), Du(x)\rangle - \mathrm{Tr}[DG_t^{-1}(x)DG_t'(x)]\right] d\mu_0(x)\Big|_{t=0}$$
$$= \int_X [\langle D\phi(x), Du(x)\rangle - \Delta u]\, d\mu_0(x)$$
$$e''(0) = \int_X \left[Du(x)^T D^2\phi(G_t(x))Du(x) + \mathrm{Tr}[(DG_t^{-1}(x)DG_t'(x))^2]\right] d\mu_0(x)\Big|_{t=0}$$
$$= \int_X \left[Du(x)^T D^2\phi(x)Du(x) + \mathrm{Tr}[(D^2u(x))^2]\right] d\mu_0(x).$$

We'd like to see that $e''(0) \geq 0$. Note that although $D^2(u) \geq -I$, by squaring it you get a symmetric matrix and therefore $\mathrm{Tr}[(D^2u(x)]^2] \geq 0$. If ϕ is convex, then the first term is non-negative too so that $e''(0) \geq 0$. However ϕ may not be convex (in which case we'd have that $(X, d, m) \in CD^e(K, N)$ for some negative K).

i) For ϕ constant, we claim that $e''(0) \geq 1/n(e'(0))^2$. Since ϕ is constant, $e'(0)$ and $e''(0)$ simplify to just be

$$e'(0) = -\int_X \Delta u(x)\ d\mu_0(x)$$
$$e''(0) = \int_X \mathrm{Tr}[(D^2u(x))^2]\ d\mu_0(x).$$

Then, by applying Jensen's inequality to $e'(0)$:

$$(e'(0))^2 \leq \int_X \Delta u(x)^2\ d\mu_0(x) = \int_X \mathrm{Tr}[D^2u(x)]^2\ d\mu_0(x).$$

Note that by Cauchy-Schwarz (for the Hilbert-Schmidt norm)

$$\mathrm{Tr}[AB] \leq \sqrt{\mathrm{Tr}[A^TA]\,\mathrm{Tr}[B^TB]}$$

squaring both sides and applying this with $A = \mathrm{Id}$ and $B = D^2u(x)$ yields

$$\mathrm{Tr}[D^2u(x)]^2 \leq n\,\mathrm{Tr}[(D^2u(x))^2].$$

Applying this in the above then gives

$$(e'(0))^2 \leq n\int_X \mathrm{Tr}[(D^2u(x))^2]\ d\mu_0(x) = ne''(0).$$

Of course, if $K \leq 0$ then we still have $nKd_2(\mu_0, \mu_1)^2 + (e'(0))^2 \leq ne''(0)$ as desired.

ii) We still use the above idea, but it's a little more delicate. By Jensen's inequality, Young's inequality (for some $\epsilon > 0$ to be chosen shortly), and Cauchy-Schwarz for the Hilbert-Schmidt norm,

$$\frac{1}{N}(e'(0))^2 \leq \frac{1}{N}\int_X [\langle D\phi(x), Du(x)\rangle - \Delta u(x)]^2\, d\mu_0(x)$$
$$\leq \frac{1}{N}\int_X \left[(1+\epsilon^{-1})\langle D\phi(x), Du(x)\rangle^2 + n(1+\epsilon)\,\mathrm{Tr}[(D^2u(x))^2]\right] d\mu_0(x).$$

To control this by $e''(0)$, we need the coefficient $n/N(1+\epsilon)$ for the trace term to be 1. Thus,

$$\epsilon := \frac{N-n}{n}, \quad 1+\epsilon^{-1} = \frac{N}{N-n}.$$

With this choice of ϵ,

$$\frac{1}{N}(e'(0))^2 \leq \int_X \left[\frac{Du(x)^T(D\phi\otimes D\phi)(x)Du(x)}{N-n} + \mathrm{Tr}[(D^2u(x))^2]\right] d\mu_0(x)$$
$$\leq \int_X Du(x)^T \left[\frac{(D\phi\otimes D\phi)(x)}{N-n} - D^2\phi(x)\right] Du(x)\ d\mu_0(x) + e''(0)$$

where we have used the definition that $\langle D\phi(x), Du(x)\rangle^2 = Du(x)^T(D\phi\otimes D\phi)(x)Du(x)$. Via the assumption

$$D^2\phi(x) - \frac{(D\phi\otimes D\phi)(x)}{N-n} - K\,\mathrm{Id} \geq 0$$

for all $x \in X$, we get

$$\frac{1}{N}(e'(0))^2 \leq -K\int_X |Du(x)|^2\ d\mu_0(x) + e''(0).$$

Finally, as G_1 is the Brenier map transporting μ_0 to μ_1, we have

$$d_2(\mu_0,\mu_1)^2 = \int_X |G_1(x)-x|^2\ d\mu_0(x) = \int_X |Du(x)|^2\ d\mu_0(x)$$

and thus

$$\frac{1}{N}(e'(0))^2 \leq -Kd_2(\mu_0,\mu_1)^2 + e''(0)$$

as desired.

There is only one technicality to check in this heuristic proof: that we can actually take derivatives under the integral sign when computing $e''(t)$. Recall that there exists a function $g(s,t)$ with $s,t\in[0,1]$ such that

$$\begin{cases} \partial^2 g/\partial s^2 = -\delta(s-t) & (s,t)\in[0,1]^2 \\ g(s,t) = 0 & (s,t)\in\partial[0,1]^2 \end{cases}$$

This is the Green's function for the Laplacian on $[0,1]^2$ (used to solve Poisson's equation on $[0,1]^2$). An explicit formula is

$$g(s,t) = \min\{s,t\} - st.$$

For any $f\in C^2([0,1])$ we have

$$(1-t)f(0) + tf(1) - f(t) = \int_0^1 g(s,t)f''(s)\ ds.$$

This also holds for any f convex or semi-convex (in which case you view $f''(s)$ as a distributional second derivative). Justifying the above is essentially a matter of using integration by parts, where we have to add the term $(1-t)f(0)+tf(1)$ to enforce the left-hand side to vanish along the boundary. We apply this first to the integrand

$$e(x,t) = \phi(G_t(x)) + \log(\rho_0(x)) - \log(\det DG_t(x))$$

yielding

$$(1-t)e(x,0) + te(x,1) - e(x,t) = \int_0^1 g(s,t)e''(x,s)\ ds.$$

Computing

$$e''(x,t) = \left[Du(x)^T D^2\phi(G_t(x))Du(x) + \mathrm{Tr}[(DG_t^{-1}(x)DG_t'(x))^2]\right]$$

is valid for a.e. $x\in \mathrm{spt}\,\mu_0$ since G_t is linear in t. By integrating the above over X with respect to μ_0 and using Fubini, we see that

$$(1-t)e(0) + te(1) - e(t) = \int_0^1 g(s,t)\int_X e''(x,s)\ d\mu_0(x)ds.$$

Now applying the formula above in the opposite direction, we can identify $\int_X e''(x,t)\ d\mu_0(x)$ as the distributional second derivative of $e(t)$. To see that $e(t)$ is semi-convex (in order to apply the above), we split it into two integrals. The first, $\int_X \log(\rho_t(x))\ d\mu_t(x)$, is convex because $(\mathbb{R}^n, |\cdot|, dx) \in CD^e(0,\infty)$. The second, $\int_X \phi(x)\ d\mu_t(x)$ has semi-convexity constant something like $\|\phi\|_{C^2(\mathbb{R}^n)} d_2(\mu_0,\mu_1)^2$.

□

Remark 2.22. We can generalize this to Riemannian manifolds. In this case, $(M^n, d_g, e^{-\phi(x)} d\,\mathrm{vol}_g) \in CD^e(K,N)$ if and only if either ϕ is constant and $N < n$ or

$$\mathrm{Ric}(x)_{ik} + D^2\phi(x)_{ik} - K g_{ij}(x) \geq 0.$$

The left-hand side is sometimes called the N-Ricci tensor, denoted Ric_N.

3. Applications of $CD^e(K,N)$ Spaces.

3.1. **The Generalized Brunn-Minkowski Inequality.**

Let's first recall the Brunn-Minkowski inequality. Given $K, L \subset \mathbb{R}^n$ we have for $t \in [0,1]$

$$|(1-t)K + tL|^{1/n} \geq (1-t)|K|^{1/n} + t|L|^{1/n}$$

where $(1-t)K + tL$ is the Minkowski sum. This geometric inequality has an algebraic analog. If you have symmetric, positive semi-definite $n \times n$ matrices A and B then

$$\det(A+B)^{1/n} \geq \det(A)^{1/n} + \det(B)^{1/n}.$$

The proof is not so bad.

Proof. Assuming A is invertible (if not, perturb it slightly) we get

$$\det(A+B)^{1/n} = \det(A)^{1/n} \det(I + A^{-1}B)^{1/n}.$$

The matrix $A^{-1}B$ is still symmetric and positive definite. If we can show the inequality for I and $A^{-1}B$ then,

$$\det(A+B)^{1/n} \geq \det(A)^{1/n}\left[\det(I)^{1/n} + \det(A)^{-1/n}\det(B)^{1/n}\right] = \det(A)^{1/n} + \det(B)^{1/n}$$

as desired. We just need to show

$$\det(I+B)^{1/n} \geq 1 + \det(B)^{1/n}$$

for symmetric, positive semi-definite matrices B. Let $0 \leq b_1 \leq ... \leq b_n$ denote the eigenvalues of B. Since we can simultaneously diagonalize I and B, we need to show that

$$\prod_{i=1}^n (1+b_i)^{1/n} \geq 1 + \prod_{i=1}^n b_i^{1/n}$$

or equivalently

$$1 \geq \left(\prod_{i=1}^n \frac{1}{1+b_i}\right)^{1/n} + \left(\prod_{i=1}^n \frac{b_i}{1+b_i}\right)^{1/n}.$$

By the inequality of arithmetic and geometric means,

$$\left(\prod_{i=1}^n \lambda_i\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^n \lambda_i$$

whenever $\lambda_1, ..., \lambda_n$ are non-negative. Applying this twice yields

$$\left(\prod_{i=1}^n \frac{1}{1+b_i}\right)^{1/n} + \left(\prod_{i=1}^n \frac{b_i}{1+b_i}\right)^{1/n} \leq \frac{1}{n}\left(\sum_{i=1}^n \frac{1}{1+b_i} + \sum_{i=1}^n \frac{b_i}{1+b_i}\right) = 1$$

as desired. □

We first start with a definition

Definition 3.1 (The set of midpoints)**.** Let (X,d,m) be a metric measure space. Then for all $A_0, A_1 \subset \operatorname{spt} m$ Borel and $t \in [0,1]$ we define the *set of midpoints at time* t to be

$$A_t := \{z \in X \mid \exists\ x \in A_0,\ y \in A_1 \text{ such that } d(x,z)/t = d(z,y)/(1-t) = d(x,y)\}$$

Remark 3.2. The condition

$$\frac{d(x,z)}{t} = \frac{d(z,y)}{1-t} = d(x,y)$$

implies that z lies on a geodesic connecting x to y.

Theorem 3.3 (Generalized Brunn-Minkowski)**.** *Let* $(X,d,m) \in CD^e(K,N)$ *with* $1 \leq N < \infty$. *Then,*

$$m^*(A_t)^{1/N} \geq \sigma^{(1-t)}_{K/N}(\Theta) m(A_0)^{1/N} + \sigma^{(t)}_{K/N}(\Theta) m(A_1)^{1/N}.$$

where

$$\Theta = \begin{cases} \min_{(x,y)\in A_0\times A_1} d(x,y) = \underline{\Theta} & K > 0 \\ 1 & K = 0 \\ \max_{(x,y)\in A_0\times A_1} d(x,y) = \overline{\Theta} & K < 0 \end{cases}$$

Remark 3.4. This says exactly what the usual Brunn-Minkowski inequality says in Euclidean space: the (outer) measure of the set of midpoints is at least as large as the convex combination of the endpoints (if $K = 0$, otherwise it is a weighted combination). Note that we need to take the outer measure since the set A_t need not be measurable even if A_0, A_1 are. However, if A_0, A_1 are both compact then A_t is closed.

Remark 3.5. You'd really expect $d_2(1/m(A_0)\chi_{A_0}, 1/m(A_1)\chi_{A_1})$ to appear instead of Θ. But, σ is monotone in Θ and $\underline{\Theta} \leq d_2(1/m(A_0)\chi_{A_0}, 1/m(A_1)\chi_{A_1}) \leq \overline{\Theta}$. For $K > 0$ (resp. $K < 0$), σ is monotone increasing (decreasing) and you can replace $d_2(\cdot,\cdot)$ by its lower (upper) bound to get a weaker inequality. To see these bounds, let $\mu_{A_0} = 1/m(x)\chi_{A_0}$ (similarly define μ_{A_1}) and recall that

$$d_2(\mu_{A_0}, \mu_{A_1})^2 = \inf_{\gamma\in\Gamma(\mu_{A_0},\mu_{A_1})} \int_{A_0\times A_0} d(x,y)\ d\gamma(x,y).$$

But since $(\pi_1)_\#\gamma = \mu_{A_0}$ and $(\pi_2)_\#\gamma = \mu_{A_1}$, we see that $\operatorname{spt}\gamma = A_0 \times A_1$. Then,

$$d_2(\mu_{A_0}, \mu_{A_1})^2 = \inf_{\gamma\in\Gamma(\mu_{A_0},\mu_{A_1})} \int_{A_0\times A_1} d(x,y)\ d\gamma(x,y) \leq \inf_{\gamma\in\Gamma(\mu_{A_0},\mu_{A_1})} \overline{\Theta}[\gamma(A_0\times A_1)] = \overline{\Theta},$$

and similarly for $\underline{\Theta}$.

There are two advantages to this. First, $\overline{\Theta}$ and $\underline{\Theta}$ are more geometrically nice quantities to work with. Also, they're well defined even if $m(A_0), m(A_1) = 0$.

Proof. First since m is locally-finite it is in particular inner regular. Suppose the inequality holds when A_0 and A_1 are compact. Let $K_0 \subset A_0$ and $K_1 \subset A_1$ be compact. Trivially we have $K_t \subset A_t$, where K_t is the set of midpoints between K_0 and K_1. Then,

$$m^*(A_t)^{1/N} \geq m(K_t)^{1/N} \geq \sigma^{(1-t)}_{K/N}(\Theta) m(K_0)^{1/N} + \sigma^{(t)}_{K/N}(\Theta) m(K_1)^{1/N}.$$

By taking the sup over $K_0 \subset A_0$ and $K_1 \subset A_1$ compact, we get the desired inequality.

So it suffices to assume that A_0, A_1 are compact (and have non-zero measure). Set $\mu_i = 1/m(A_i) m \llcorner A_i$ for $i \in \{0,1\}$. Since $(X,d,m) \in CD^e(K,N)$, there exists $(\mu_t)_{t\in[0,1]}$ a d_2-geodesic along which S is $(Kd_2(\mu_0,\mu_1)^2, N)$-convex. Equivalently, $U_N(\mu) := e^{-S(\mu)/N}$ we have

$$\begin{aligned} U_N(\mu_t) &\geq \sigma^{(1-t)}_{K/N}(d_2(\mu_0,\mu_1)) U_N(\mu_0) + \sigma^{(t)}_{K/N}(d_2(\mu_0,\mu_1)) U_N(\mu_1) \\ &\geq \sigma^{(1-t)}_{K/N}(\Theta) U_N(\mu_0) + \sigma^{(t)}_{K/N}(\Theta) U_N(\mu_1). \end{aligned}$$

When $i \in \{0,1\}$ we have

$$\rho_i := \frac{d\mu_i}{dm} = \begin{cases} 1/m(A_i) & x \in A_i \\ 0 & x \notin A_i \end{cases}$$

so that

$$S(\mu_i) = \int_{A_0} \rho_i \log\rho_i\ dm = \int_{A_i} \frac{1}{m(A_i)} \log\left(\frac{1}{m(A_i)}\right) dm = \log\left(\frac{1}{m(A_i)}\right)$$

and therefore

$$U_N(\mu_i) = e^{-\log(1/m(A_i))/N} = e^{\log(m(A_i)^{1/N})} = m(A_i)^{1/N}.$$

Finally, applying this in the above we have

$$U_N(\mu_t) \geq \sigma^{(1-t)}_{K/N}(\Theta) m(A_0)^{1/N} + \sigma^{(t)}_{K/N}(\Theta) m(A_1)^{1/N}.$$

It remains to show $m(A_t)^{1/N} \geq U_N(\mu_t)$. Recall that μ_t vanishes outside A_t (for the exact reason that for a.e. $z \in \operatorname{spt}\mu_t$, z lies on a geodesic connecting points $x \in \operatorname{spt}\mu_0$ and $y \in \operatorname{spt}\mu_1$ with the ratio $1-t$ and t). Setting $\rho_t = d\mu_t/dm$ we have

$$U_N(\mu_t) = \exp\left[-\frac{1}{N}\int_{A_0} \log\rho_t \; d\mu_t\right].$$

Applying Jensen's inequality with the convex function $f(r) = e^{-r/N}$,

$$U_N(\mu_t) \leq \int_{A_t} \rho_t^{-1/N} \; d\mu_t = m(A_t)\int_{A_t} \frac{\rho_t^{1-1/N}}{m(A_t)} \; dm.$$

Now we apply Jensen's inequality again to $f(r) = -r^{1-1/N}$, using the fact that $m/m(A_t)$ is a probability measure (on A_t), gives

$$U_N(\mu_t) \leq m(A_t)\left(\int_{A_t} \frac{\rho_t}{m(A_t)} \; dm\right)^{1-1/N} = m(A_t)\left(\int_{A_t} \frac{1}{m(A_t)} \; d\mu_t\right)^{1-1/N} = m(A_t)^{1/N}$$

since μ_t is a probability measure supported on A_t. □

Remark 3.6. If instead $N = \infty$ then

$$\log(m(A_t)) \geq (1-t)\log(m(A_0)) + t\log(m(A_1)) + \frac{K}{2}t(1-t)\Theta^2.$$

Proof. Take A_0, A_1 compact and $(\mu_t)_{t\in[0,1]}$ as before. Then $\rho_t = d\mu_t/dm = 0$ outside A_t. By $(Kd_2(\mu_0,\mu_1)^2, N)$-convexity,

$$S(\mu_t) \leq (1-t)S(\mu_0) + tS(\mu_1) - \frac{K}{2}t(1-t)d_2(\mu_0,\mu_1)^2.$$

Recall $S(\mu_i) = \log(1/m(A_i))$ with $i \in \{0,1\}$. Then, depending on the sign of K we can replace $d_2(\mu_0,\mu_1)$ by $\underline{\Theta}$ or $\overline{\Theta}$ as appropriate. It remains to show $S(\mu_t) \geq \log(1/m(A_t))$. We again use Jensen's inequality applied to

$$S(\mu_t) = m(A_t)\int_{A_t} \rho_t \log\rho_t \; \frac{dm}{m(A_t)}$$

with $f(r) = r\log r$. Then,

$$S(\mu_t) \geq m(A_t) f\left(\int_{A_t} \frac{1}{m(A_t)} \; d\mu_t\right) = m(A_t)\left[\frac{1}{m(A_t)}\log\left(\frac{1}{m(A_t)}\right)\right] = \log\left(\frac{1}{m(A_t)}\right)$$

since $\mu_t(A_t) = 1$. □

3.1.1. *Corollaries of the Generalized Brunn-Minkowski Inequality.* Here is a nice corollary of this regarding the atoms of m.

Corollary 3.7 (c.f. [Vil09] Corollary 30.9)**.** *If $(X,d,m) \in CD^e(K,N)$ then m has no atoms (unless* $\operatorname{spt} m = \{x_0\}$*).*

Proof. Recall that $CD^e(K,N) \subset CD^e(K,\infty)$. So, we assume without loss of generality that $N = \infty$. Suppose for a contradiction that $x \in X$ is an atom, i.e. $m(\{x\}) > 0$. Set $A_0 = \{x\}$ and $A_1 = X\setminus\{x\}$. Because A_1 is compact, for every $\epsilon > 0$ and t small enough, $A_t \subset B_\epsilon(x) \setminus \{x\}$. Then, $m(A_t) \to 0$ and $\log(1/m(A_t)) \to \infty$ as $t \to \infty$. On the other hand, by Remark 3.6 we get

$$\log\left(\frac{1}{m(A_t)}\right) \leq (1-t)\log\left(\frac{1}{m(A_0)}\right) + t\log\left(\frac{1}{m(A_1)}\right) - \frac{K}{2}t(1-t)\Theta^2.$$

Taking $t \to 0$ in the above shows that

$$\infty = \lim_{t\to 0} \log\left(\frac{1}{m(A_t)}\right) \le \log\left(\frac{1}{m(A_0)}\right).$$

So, either $m(A_0) = m(\{x\}) = 0$ or ∞. As m is locally finite, it must be that $m(\{x\}) = 0$, a contradiction. □

We claim now that $m(B_r(x))$ is continuous in $r > 0$, i.e. that $m(\partial B_r(x)) = 0$. We'll prove a slightly more general corollary, from which this follows.

Corollary 3.8 (c.f. [Vil09] Corollary 30.10)**.** *If $(X, d, m) \in CD^e(K, N)$ for $N < \infty$ and $Y \subset \operatorname{spt} m$ (bounded?) satisfying $Y = \bigcap_{r>0} Y^r$ where*

$$Y^r := \{y \in X \mid \operatorname{dist}(y, Y) < r\}$$

then for all $x \in Y$

$$m(Y) = \lim_{t\to 1} m(Z_t(\{x\}, Y))$$

Proof. Let $R = \max_{y\in Y} d(x, Y)$. Then,

$$A_t := Z_t(\{x\}, Y) \subset \mathrm{Cl}(Y^{(1-t)R}).$$

By the hypothesis $Y = \bigcap_{r>0} Y^r$, we get

$$\limsup_{t\to 1} m^*(A_t) \le m(Y).$$

Recall that $K > 0$ implies $CD^e(K, N) \subset CD^e(0, N)$, so we may assume $K \le 0$. Generalized Brunn-Minkowski says that

$$m^*(A_t)^{1/N} \ge \sigma^{(t)}_{K/N}(R) m(Y)^{1/N}$$

since $m(\{x\}) = 0$, as m has no atoms. But $\sigma^{(t)}_{K/N}(R) \to 1$ as $t \to 1$, so this shows the reverse inequality. □

3.2. **The Bishop-Gromov Inequality.**

We start with some intuition. For $B_r(x) \subset \mathbb{R}^n$ we can compute the area and volume: $a_0(r) = \mathcal{H}^{n-1}(\partial B_r(x)) = \omega_n r^{n-1}$ and $v_0(r) = \mathcal{H}^n(B_r(x)) = \omega_n/nr^n$. If we look on a sphere ($K = 1$), a ball is a spherical cap. Let $a_1(r)$ and $v_1(r)$ be the corresponding area and volume. Notice that

$$\frac{a_0(r)}{r^{n-1}} = \omega_n, \qquad \frac{v_0(r)}{r^n} = \frac{\omega_n}{n}$$

are constants. On the other hand, on a sphere the area and volume grow less than in Euclidean space. So,

$$\frac{a_1(r)}{\sin(r)^{n-1}}, \qquad \frac{v_1(r)}{\sin(r)^n}$$

are non-increasing. Similarly in hyperbolic space,

$$\frac{a_{-1}(r)}{\sinh(r)^{n-1}}, \qquad \frac{v_{-1}(r)}{\sinh(r)^n}$$

are non-increasing. The Bishop-Gromov inequality generalizes this to spaces with non-constant curvature.

Theorem 3.9 (Bishop-Gromov)**.** *If $(X, d, m) \in CD^e(K, N)$ with $N < \infty$ then for all $x_0 \in \operatorname{spt} m$ the map $r \in (0, \pi\sqrt{N/K_+}] \mapsto ra(r)/s_{K/N}(r)^N$ is non-increasing, as is $r \mapsto v_r / \int_0^r s_{K/N}(t)^N/t \, dt$ (over the same interval). Here,*

$$a(r) = \limsup_{\delta\to 0} \frac{1}{\delta} m(\mathrm{Cl}(B_{r+\delta}(x_0)) \setminus B_r(x_0)), \qquad v(r) = m(\mathrm{Cl}(B_r(x_0)))$$

Remark 3.10. We expect an integral of $s_{K/N}$ to appear for v as volume is essentially an integral of area. The fact that we use $ra(r)/s_{K/N}(r)^N$ means this is not a sharp statement. The sharp statement would be that $r \mapsto a(r)/\sin(r\sqrt{K/(N-1)})^{N-1}$ and $r \mapsto v(r)/\int_0^r \sin(t\sqrt{K/(N-1)})^{N-1}\,dt$ for $r \in (0, \pi/\sqrt{(N-1)/K})$. This is more naturally expected from our previous analysis of the Ricci curvature on a sphere; it also holds for any $CD(K,N)$ space. In fact, for non-branching spaces $CD(K,N)$ and $CD^e(K,N)$ are equivalent. The latter is more useful to prove the non-sharp Bishop-Gromov inequality, so this is what we will do.

Before beginning, we need the following lemma.

Lemma 3.11 (c.f. [Vil09] Lemma 18.9)**.** *Suppose $g : (0,b) \to (0,\infty)$ (with $b \leq \infty$) is continuous and $G(r) := \int_0^r g(s)\,ds < \infty$. Let $F : [0,b) \to [0,\infty)$ be non-decreasing with $F(0) = 0$ and $f(r) = d^+F/dr$ its upper derivative. Then, if f/g is non-increasing so too is F/G.*

Proof. Let r, R be such that $0 < r < R$. By monotonicity of f/g, we have $f(R) \leq G(R)f(r)/g(r)$. Thus, f is locally bounded and in particular F is locally Lipschitz on $[0,b)$. Hence, F is the anti-derivative of f,

$$F(r) = \int_0^r f(s)\,ds.$$

Let $h = f/g$. Then $\inf_{s\in(0,r)} h(s) \geq a \geq \sup_{s\in(r,R)} h(s)$. Applying this yields

$$\begin{aligned}\int_0^r f(s)\,ds \int_r^R g(s)\,ds &= \int_0^r g(s)h(s)\,ds \int_r^R g(s)\,ds \geq \int_0^r ag(s)\,ds \int_r^R g(s)\,ds \\ &= \int_0^r g(s)\,ds \int_r^R ag(s)\,ds \geq \int_0^r g(s)\,ds \int_r^R h(s)g(s)\,ds \\ &= \int_0^r g(s)\,ds \int_r^R f(s)\,ds.\end{aligned}$$

So, we've just shown

$$\frac{F(r)}{G(r)} \geq \frac{F(R) - F(r)}{G(R) - G(r)}.$$

Multiplying through yields, after cancellation

$$F(r)G(R) \geq F(R)G(r)$$

as was desired. □

Proof of Theorem 3.9. The second monotonicity statement follows immediately from the first using Lemma 3.11. Just take $F(r) = v(r)$ and $g(r) = s_{K/N}(r)^N/r$.

To prove the first part, let $\epsilon, \delta > 0$. Set $X = \mathrm{Cl}(B_\epsilon(x_0))$ and $Y = \mathrm{Cl}(B_{R(1+\delta)}(x_0)) \setminus B_R(x_0)$. Given $0 < r < R$ we apply Brunn-Minkowski to estimate $m(A_t)$ with $t = r/R$ and $A_t = Z_t(X,Y)$. Thinking about this, if $(x,y) \in X \times Y$ the triangle inequality says

$$R - \epsilon \leq d(x,y) \leq R(1+\delta) + \epsilon.$$

Now, A_t is some interpolation between the ball X and the annulus Y (so it maybe looks like an annulus too). But every point in A_t divides a geodesic starting in X and ending in Y in ratio t to $1-t$; that is, if $z \in A_t$ then there exist $x \in X$ and $y \in Y$ such that $d(x,z) = td(x,y)$ and $d(y,z) = (1-t)d(x,y)$. Even though A_t is not quite an annulus, we can at least say that

$$A_t \subset \mathrm{Cl}(B_{r(1+\delta)+o(1)}(x_0)) \setminus B_{r-o(1)}(x_0)$$

where $o(1) \to 0$ as $\epsilon \to 0$ (the signs on the error terms are not important). The triangle inequality, combined with the above bounds, proves this claim. Then,

$$v(r(1+\delta) + o(1)) - v(r - o(1))^{1/N} \geq m(A_t)^{1/N}.$$

By Brunn-Minkowski, we can bound the right-hand side from below to give:

$$[v(r(1+\delta)+o(1)) - v(r-o(1))]^{1/N} \geq \sigma_{K/N}^{(1-r/R)}(R \mp (\delta R+\epsilon))m(B_\epsilon(x_0))^{1/N} + \sigma_{K/N}^{(r/R)}(R \mp (\delta R+\epsilon))m(Y)^{1/N}$$

where we choose $-$ above if $K>0$ and $+$ above if $K<0$ (these are just weakened bounds on $d(x,y)$ we have before). Taking $\epsilon \to 0$ we obtain

$$[v(r(1+\delta)) - v(r)]^{1/N} \geq \sigma_{K/N}^{(r/R)}(R(1\mp\delta))m(Y)^{1/N}.$$

Now because Y is an annulus,

$$m(Y) = v(R(1+\delta)) - v(R).$$

Using this and taking $\delta \to 0$ gives

$$a(r) = \lim_{\delta\to 0}\frac{v(r(1+\delta)) - v(r)}{r\delta} \geq \frac{R}{r}\sigma_{K/N}^{(r/R)}(R)^N \lim_{\delta\to 0}\frac{v(R(1+\delta)) - v(R)}{R\delta} = \frac{R}{r}\sigma_{K/N}^{(r/R)}(R)^N a(R).$$

Recalling that

$$\sigma_{K/N}^{(t)}(\theta) = \frac{s_{K/N}(t\theta)}{s_{K/N}(\theta)},$$

we get

$$a(r) \geq \frac{R}{r}\left(\frac{s_{K/N}(r)}{s_{K/N}(R)}\right)^N a(R)$$

as desired. □

Remark 3.12. When $K>0$, why do we restrict ourselves to the interval $(0,\pi\sqrt{N/K}]$? Notice that, since m is locally finite,

$$\frac{v(r(1+\delta)) - v(r)}{r\delta} < \infty.$$

In particular, by the bound

$$\frac{v(r(1+\delta)) - v(r)}{r\delta} \geq \frac{R}{r}\sigma_{K/N}^{(r/R)}(R(1-\delta))^N \frac{v(R(1+\delta)) - v(R)}{R\delta}$$

we see that $\sigma_{K/N}^{(r/R)}(R(1\mp\delta))$ is finite. Define $R_* = \pi\sqrt{N/K}$ and select $R = R_*$ and $0<r<R$. When $\delta \to 0^+$, $\sigma_{K/N}^{(r/R)}(R(1-\delta))^N$ goes to ∞ (as the denominator is $s_{K/N}(\theta) = \sin(\theta\sqrt{K/N})$). This implies that $v(R(1+\delta)) - v(R) = 0$ for $\delta > 0$, that is v is constant beyond this critical value of R.

3.2.1. *Corollaries of the Bishop-Gromov Inequality.*

Corollary 3.13 (Non-sharp Bonnet-Myers)**.** *If* $K>0$ *and* $N<\infty$*, then* $\mathrm{Diam}(\operatorname{spt} m) \leq 2\pi\sqrt{N/K}$

Remark 3.14. The sharp version, as expected via dimensional analysis, should be $\mathrm{Diam}(\operatorname{spt} m) \leq \pi\sqrt{(N-1)/K}$. Equality holds on the round sphere $\partial B^{N+1}_{1/\sqrt{K}}(0)$.

We briefly review the concept of Hausdorff dimension in order to introduce the next corollary. Recall if (X,d) is a metric space, given $n \geq 0$ and $\delta > 0$ we can define

$$\mathcal{H}^n_\delta(X) = \inf\left\{\sum_{i=1}^\infty (\mathrm{Diam}(S_i))^n \;\middle|\; X \subset \bigcup_{i=1}^\infty S_i,\ \mathrm{Diam}(S_i) \leq \delta\right\}.$$

As δ increases, there are more covers to look at and so $\mathcal{H}^n_\delta$ is monotone in delta. We can therefore define the limit:

$$\mathcal{H}^n(X) = \lim_{\delta\to 0^+}\mathcal{H}^n_\delta(X),$$

which defines a Borel measure (each $\mathcal{H}_\delta$ is an outer measure). One can show that there exists a D such that whenever $n > D$, $\mathcal{H}^n = \infty$ and when $n < D$, $\mathcal{H}^n = 0$. The number n is called the *Hausdorff dimension* of (X,d) and is denoted $n = \dim_H(X)$.

Corollary 3.15 (c.f. [Vil09] Corollary 30.12)**.** *If* $(X,d,m) \in CD^e(K,N)$ *and* $X = \operatorname{spt} m$ *then* $\dim_H(X) \leq N$.

Proof. If $B_r(x_0) \subset B_R(z) \subset B_{2R}(x_0)$ then Bishop-Gromov tells us

$$\frac{m(B_r(x_0))}{\int_0^r s_{K/N}(t)^N/t\ dt} \geq \frac{m(B_{2R}(x_0))}{\int_0^{2R} s_{K/N}(t)^N/t\ dt} \geq \frac{m(B_R(z))}{\int_0^{2R} s_{K/N}(t)^N/t\ dt}.$$

On the other hand, we know that $s_{K/N}(t) = t(1 + O_{K/N}(t^2))$ as $t \to 0$. So the above yields

$$m(B_r(x_0)) \geq C(z, R, K, N)\ r^N(1 + O_{K,N}(r^2)).$$

By local finiteness of m, we know $m(B_R(z)) < \infty$ for all $z \in X$, $R < \infty$. Then any disjoint collection of balls $\{B_r(x_i)\}_{i=1}^J$ such that $B_r(x_i) \subset B_R(z)$ satisfies

$$m(B_R(z)) \geq \sum_{i=1}^J m(B_r(x_i)) \geq JCr^N(1 + O_{K/N}(r^2)).$$

Thus we get a bound on the number of disjoint balls.

Choose now a maximal disjoint collection of r-balls in $B_R(z)$. Maximality implies that

$$B_R(z) \subset \bigcup_{i=1}^J B_{2r}(x_i).$$

Taking $D > N$ and $\delta > 2r$,

$$\mathcal{H}^D_\delta(B_R(z)) \leq \sum_{i=1}^J (2r)^D = \frac{m(B_R(z))2^D r^{D-N}}{C(1 + O_{K/N}(r^2))}.$$

If $D > N$, then $\mathcal{H}_\delta(B_R(z)) \to 0$ as $r, \delta \to 0$. This exactly means that $\dim_H(X) \leq N$. □

There is one more nice corollary of 3.9.

Corollary 3.16. *If $(X, d, m) \in CD^e(K, N)$ and $N < \infty$ then*
i) m is locally doubling and
ii) (X, d) has the Heine-Borel property (i.e. boundedly compact).

To show this we need one quick lemma.

Definition 3.17. A metric space (X, d) is *totally bounded* if for every $r > 0$ there exists $N(r) < \infty$ such that there exists an r-net $\{x_1, ..., x_n\} \subset X$ (meaning that $X \subset \bigcup_{i=1}^n B_r(x_i)$) with $n \leq N(r)$.

Lemma 3.18 (c.f. Burago-Burago-Ivanov ([BBI01]) Theorem 1.6.5)**.** *A metric space (X, d) is compact if and only if (X, d) is complete and totally bounded.*

Proof. The forward direction is straightforward. The converse uses a diagonal argument and the pigeonhole principle (to trap a sequence in a given ball and force subsequential convergence). □

Proof of Corollary 3.16. The proofs of these have the same flavor as the proof of Corollary 3.15. We omit the proof of i) and focus on ii).

Given $B_R(z) \subset X$ we wish to show for all closed $C \subset B_R(z)$ that C is in fact compact. Since $CD^e(K, N)$ spaces are by definition complete, it suffices to show that $B_R(z)$ is totally bounded. Given $0 < r < R$ let $\{B_r(x_i)\}_{i=1}^J$ be a maximal disjoint collection of balls such that $B_r(x_i) \subset B_R(z)$ for $i = 1, ..., n$. Aiming to bound J, we first note that

$$J \min_{i \in \{1,...,J\}} m(B_r(x_i)) \leq \sum_{i=1}^J m(B_r(x_i)) \leq m(B_R(z))$$

and, together with Bishop-Gromov

$$J \leq \frac{m(B_R(z))}{m(B_r(x_i))} \leq \frac{\int_0^{2R} s_{K/N}(t)^N/t\ dt}{\int_0^r s_{K/N}(t)^N/t\ dt} := N_R(2r)$$

as $B_r(x_i) \subset B_R(z)$ for any $i \in \{1, ..., J\}$. We claim $N_R(2r)$ is the desired bound. Indeed, by doubling the radii of all the balls $B_r(x_i)$, we know (due to maximality) that $B_R(z) \subset \bigcup_{i=1}^J B_{2r}(x_i)$. □

4. Convergence of Metric Spaces.

We claimed before that Riemannian manifolds with Ricci curvature bounded below by $K(N-1)$ and dimensional upper bounds by N are to $CD^e(K,N)$ spaces as $\mathbb{Q}$ is to $\mathbb{R}$.[8] We know now that the $CD^e(K,N)$ spaces (when $N<\infty$) are fairly nice, so we can actually work towards formalizing this. An essential first step is to discuss the convergence of metric spaces.

4.1. **The Hausdorff distance.** First we must discuss a notion of distance between subsets of the same metric space.

Definition 4.1 (Hausdorff distance)**.** On subsets $A,B\subset X$ there is the *Hausdorff metric*:

$$d_H(A,B)=\inf\left\{r>0 \,\middle|\, A\subset B^r,\ B\subset A^r\right\}$$

where $A^r=\bigcup_{x\in A}B_r(x)$.

Remark 4.2. Actually $d_H:2^X\to[0,\infty]$ is a semi-metric, i.e. for $A,B,C\in 2^X$

i) $d_H(A,B)=d_H(B,A)$;
ii) $d_H(A,B)\le d_H(A,Z)+d_H(Z,B)$;
iii) $0\le d_H(A,B)$ and $d_H(A,B)=0$ if $A=B$ (but the converse need not be true!)

It ends up being that $d_H(A,\mathrm{Cl}(A))=0$, so that d_H is a semi-metric.

Whenever you have a semi-metric you can convert it into a metric by quotienting. See [BBI01] Proposition 1.1.5. If d is a semi-metric on X then the quotient space $\tilde{X}=X/\sim$ where $A\sim B$ if and only if $d(A,B)=0$ becomes a metric space under d.

Corollary 4.3. *d_H is a metric on closed sets $\mathscr{C}(X)$ of X.*

Even more, $(\mathscr{C}(X),d_H)$ inherits certain properties from (X,d).

Proposition 4.4 (c.f. [BBI01] Proposition 7.3.7)**.** *If (X,d) is complete, then so too is $(\mathscr{C}(X),d_H)$ is too.*

Proof. Given a Cauchy sequence $\{C_n\}_{n=1}^\infty\subset\mathscr{C}(X)$, we claim that $C_n\to C$ in d_H where

$$C:=\{x\in X\mid \forall r>0,\ B_r(x)\text{ intersects infinitely many }C_n\}.$$

Since $d_H(C,\mathrm{Cl}(C))=0$, we can assume $C=\mathrm{Cl}(C)$. Given any $\epsilon>0$ there exists $N(\epsilon)$ such that $n,m\ge N(\epsilon)$ implies $d_H(C_n,C_m)<\epsilon$. Now we show for any $n\ge N(\epsilon)$ that $d_H(C_n,C)<2\epsilon$. To do this we need to show that for any $x\in C$ and $y\in C_n$, both $d(x,C_n)<2\epsilon$ and $d(y,C)<2\epsilon$.

For the first, by definition there exists $m\ge N(\epsilon)$ arbitrarily large with $B_\epsilon(x)\cap C_m\ne\emptyset$. I.e., there exists $x_m\in C_m$ with $d(x,x_m)<\epsilon$. On the other hand, since $d_H(C_n,C_m)<\epsilon$ there exists some $x_n\in C_n$ such that $d(x_n,x_m)<\epsilon$. Combining the two yields $d(x,x_n)<2\epsilon$ and thus also $d(x,C_n)<2\epsilon$.

For the second, we set $n_1=n$ and for all $k\in\mathbb{N}$ we choose $n_k>n_{k-1}$ such that

$$d_H(C_p,C_q)<\frac{\epsilon}{2^k}$$

if $p,q\ge n_k$. Letting $x_1=y\in C_{n_1}$ we recursively choose $x_k\in C_{n_k}$ such that $d(x_k,x_{k-1})<\epsilon/2^{k-1}$. We can do this as $d_H(C_{n_k},C_{n_{k-1}})<\epsilon/2^{k-1}$. Then $\{x_k\}_{k=1}^\infty$ is Cauchy because

$$d(x_p,x_q)<\sum_{i=k+1}^{\infty}\frac{\epsilon}{2^i}<\frac{\epsilon}{2^k}$$

[8] Note: We are not claiming here that the completion of such Riemannian manifolds in some suitable sense is $CD^e(K,N)$, but that the completion lies in this space.

as soon as $p, q \geq n_{k+1}$. But (X,d) is complete so that $x_k \to x \in X$. Since the x_k are in some C_{n_k}, and $x_k \to x$, it follows that $x \in C$ (otherwise we could bound $d(x_k, x)$ from below by a non-negative number for all but finitely many k). Taking $p \to \infty$ and $q \to 1$ we see that

$$d(x,y) = d(x,x_1) < \sum_{i=1}^{\infty} \frac{\epsilon}{2^i} = \epsilon$$

and hence $d(y, C) < \epsilon$. □

Proposition 4.5 (c.f. [BBI01] Theorem 7.3.8)**.** *If (X,d) is compact, then so too is $(\mathscr{C}(X), d_H)$.*

Proof. Since (X,d) is compact, it is complete, and therefore $(\mathscr{C}(X), d_H)$ is complete by Proposition 4.4. So it suffices to check $(\mathscr{C}(X), d_H)$ is totally bounded. By compactness (X,d) is totally-bounded. So for every $\epsilon > 0$ there exists a finite ϵ-net $\{x_1, ..., x_N\} \in X$. Denoting $X_\epsilon = \{x_1, ..., x_N\}$, we check that 2^{X_ϵ} is an ϵ-net in $(\mathscr{C}(X), d_H)$. That is, for each $C \subset X$ closed there exists $C_\epsilon \in 2^{X_\epsilon}$ such that $d_H(C, C_\epsilon) < \epsilon$. Probably choose

$$C_\epsilon := \{x \in X_\epsilon \mid d(x, C) < \epsilon\}.$$

Clearly $C_\epsilon \subset X_\epsilon \subset C^\epsilon$. On the other hand, we also need $C \subset (C_\epsilon)^\epsilon$ – meaning for every $x \in C$ we need to find $x_\epsilon \in C_\epsilon$ such that $d(x_\epsilon, x) < \epsilon$. Since $\{x_1, ..., x_N\}$ is an ϵ-net, some x_i is such that $d(x_i, x) < \epsilon$. By definition, this means too that $x_i \in C_\epsilon$, so we choose this as x_ϵ. □

4.2. **The Gromov-Hausdorff distance.** This is great, but it has the downside of comparing subesets of a particular space. We're interested in comparing different metric spaces to each other. One of Gromov's ideas was to compare (X, d_1) to (Y, d_Y) by looking at all isometric embeddings i_1 and i_2 of (X, d_X) and (Y, d_Y) into an arbitrary third space (X, d). This defines the so-called Gromov-Hausdorff distance.

Definition 4.6 (Gromov-Hausdorff distance)**.** A map $i : (X, d_X) \to (Y, d_Y)$ is an *isometry* if for all $x, x' \in X$ we have $d_Y(i(x), i(x')) = d_X(x, x')$ and i is a bijection. We say i is an *isometric embedding* if there is a bijection between X and $i(X)$. Then, the *Gromov-Hausdorff distance* is defined by

$$d_{GH}((X,d_X),(Y,d_Y)) = \inf \left\{ d_H(i_1(X), i_2(Y)) \;\middle|\; \begin{array}{l} (X,d) \text{ is a metric space and} \\ i_1 : (X, d_X) \to (Z, d_Z) \\ \text{and } i_2 : (Y, d_Y) \to (Z, d_Z) \text{ are} \\ \text{isometric embeddings for } k = 1,2 \end{array} \right\}.$$

Remark 4.7. As with the Hausdorff distance, d_{GH} is a semi-metric on compact metric spaces. This is because $d_{GH}((X, d_X), (Y, d_Y)) = 0$ if X and Y are isometric (the converse holds when X, Y are compact). Using the same quotient construction as before gives a metric on isometry classes of compact metric spaces.

We aim to show the following meta "theorem":

Theorem 4.8. *For every $K \in \mathbb{R}$ and $N < \infty$, $CD^e(K,N)$ is pre-compact for d_{GH}.*

Why the quotation marks? When $K > 0$, the spaces genuinely are compact. Since $(X,d) \in CD^e(K,N)$ has the Heine-Borel property, it suffices to check that X is closed and bounded. Closedness follows from completeness (in the definition, as X is a Polish space). Boundedness follows from the non-sharp Bonnet-Myers bound. So, we are in good shape when $K > 0$.

However when $K \leq 0$ we need a notion of measuring distance between non-compact spaces. This is more subtle. Consider a cylinder which is shaped like a wine bottle – thick at one end and narrow at the other, which extends to infinity in both directions. Taking a sequence of translations towards the narrow end yields a narrow cylinder, but taking a sequence translating towards the thick end yields a thick cylinder. Of course, taking a constant sequence gives itself, so in every instance we're getting a different limit. We need a better notion of convergence (called pointed Gromov-Hausdorff).

Can we hope to strengthen this to a compactness result? In order to do so, we also need to keep track of the measures. Thus we must also modify the notion of convergence to take into account the measures. In this sense, we get compactness.

We saw in Propositions 4.4 and 4.5 that completeness and compactness are inherited for d_H. We're not going to use completeness and compactness in quite the same way at the level of Gromov-Hausdorff, but we are going to use the following: when we want to show that a sequence of metric (measure) space Gromov-Hausdorff converges, we'll use this same kind of reduction to checking collections of finite points as in Proposition 4.5 (which is a generalization of the Blaschke selection theorem). As a corollary we also get

Corollary 4.9. *Convex subsets of $(K, |\cdot|)$ with $K \subset \mathbb{R}^n$ form a d_H-closed subset of $\mathscr{C}(K)$.*

It's possible to weaken this by relaxing the boundedness assumption – we want a result that does not a priori assume boundedness. This is similar to the meta "theorem" 4.8; we need compactness, which is not guaranteed in the $K \leq 0$ case. But, by changing the convergence a little bit, we can get compactness.

4.3. Alternative representations of the Gromov-Hausdorff distance.

4.3.1. *Metric extensions.* One of the downsides of Definition 4.6 for the Gromov-Hausdorff metric is that it's too big to work with – there are too many possible spaces to embed into. On the other hand, we can think about embedding X and Y into the union $X \cup Y$ and equipping this with a metric extension $d_{X\cup Y}$, that is $d_{X\cup Y}$ is a metric such that $d_{X\cup Y}(x,x') = d_X(x,x')$ whenever $x, x' \in X$ and $d_{X\cup Y}(y,y') = d_Y(y,y')$ whenever $y, y' \in Y$. The inclusions $i_1 : X \to X \cup Y$ and $i_2 : Y \to X \cup Y$ are isometries so that

$$d_{GH}((X,d_X),(Y,d_Y)) \leq \inf\left\{ d_H(X,Y) \;\middle|\; d_{X\cup Y} \text{ is a metric extension of } d_X \text{ and } d_Y \text{ to } X \cup Y \right\}.$$

In the above, we take the Hausdorff distance in the space $(X \cup Y, d_{X\cup Y})$ and we naturally regard $i_k(X_k) = X_k$. For the converse, let (Z, d_Z) be a metric space and $i_1 : X \to Z$ and $i_2 : Y \to Z$ isometric embeddings. Let $\delta > 0$ and define $d_{X\cup Y}$ a metric extension on $X \cup Y$ by

$$d_{X\cup Y}(x,y) = d_{X\cup Y}(y,x) = d_Z(i_1(x), i_2(y)) + \delta$$

for $x \in X$ and $y \in Y$, and $d_{X\cup Y}(x,x') = d_X(x,x')$ for $x, x' \in X$ (similarly for $y, y' \in Y$). We need this constant δ as $i_1(X) \cap i_2(Y)$ may not be empty. By definition $d_{X\cup Y}$ is symmetric and non-negative. If $d_{X\cup Y} = 0$ then it must be either $d_{X\cup Y}(x,x') = 0$ or $d_{X\cup Y}(y,y') = 0$, in which case we apply the fact that $d_{X\cup Y}$ is an extension to conclude $x = x'$ or $y = y'$. All that is left to check is that $d_{X\cup Y}$ satisfies the triangle inequality. Given triplets in X or Y this is trivial (by the extension property). Instead we have two cases:

i) Given $x, x' \in X$ and $y \in Y$ we show

$$d_{X\cup Y}(x,x') \leq d_{X\cup Y}(x,y) + d_{X\cup Y}(y,x').$$

Since the i_k are isometries and d_Z is a metric,

$$\begin{aligned} d_{X\cup Y}(x,x') &= d_X(x,x') = d_Z(i_1(x), i_1(x')) \leq d_Z(i_1(x), i_2(y)) + d_Z(i_2(y), i_1(x')) \\ &\leq [d_Z(i_1(x), i_2(y)) + \delta] + [d_Z(i_1(x'), i_2(y)) + \delta] = d_{X\cup Y}(x,y) + d_{X\cup Y}(y,x'). \end{aligned}$$

ii) Given $x, x' \in X$ and $y \in Y$ we show

$$d_{X\cup Y}(x,y) \leq d_{X\cup Y}(x,x') + d_{X\cup Y}(x',y).$$

The proof is more or less the same as above:

$$\begin{aligned} d_{X\cup Y}(x,y) &= d_Z(i_1(x), i_2(y)) + \delta \leq d_Z(i_1(x), i_1(x')) + d_Z(i_1(x'), i_2(y)) + \delta \\ &= d_X(x,x') + d_{X\cup Y}(x',y) = d_{X\cup Y}(x,x') + d_{X\cup Y}(x',y). \end{aligned}$$

The remaining cases follow by reversing roles of x and y. Note that, by definition, $d_H^{(d_{X\cup Y})}(X,Y) \leq d_H^{(d)}(i_1(X), i_2(Y))$ as $d_{X\cup Y} \leq d$. Since we can construct a metric extension $d_{X\cup Y}$ satisfying this for any pair of embeddings i_1, i_2 into a third space (X,d), it follows that

$$d_{GH}((X,d_X),(Y,d_Y)) = \inf\left\{ d_H(X,Y) \;\middle|\; d_{X\cup Y} \text{ is a metric extension of } d_X \text{ and } d_Y \text{ to } X\cup Y \right\}.$$

The following proposition illustrates the utility of this reformulation.

Proposition 4.10 (c.f. [BBI01] Proposition 7.3.16)**.** *The Gromov-Hausdorff distance satisfies the triangle inequality.*

Proof. Given metric spaces (X,d_X), (Y,d_Y), and (Z,d_Z) if $d_{X\cup Y}$ is a metric extension of d_X and d_Y to $X\cup Y$ and $d_{Y\cup Z}$ similarly for $Y\cup Z$. How do we define $d_{X\cup Z}$? We know that $d_{X\cup Z}$ should agree with d_X if both arguments are in X. Similarly, $d_{X\cup Z}$ should agree with d_Z when both are in Z. So we only need to define $d_{X\cup Z}(x,z)$ when $x\in X$ and $z\in Z$. The correct way to do this is:

$$d_{X\cup Z}(x,z) = d_{X\cup Z}(z,x) = \inf_{y\in Y}\{d_{X\cup Y}(x,y) + d_{Y\cup Z}(y,z)\}$$

We need to check a couple things. The first is that $d_{X\cup Z}$ satisfies the triangle inequality, which involves two cases:

i) For all $x,x'\in X$ and $z\in Z$, $d_{X\cup Z}(x,x') \leq d_{X\cup Z}(x,z) + d_{X\cup Z}(z,x')$ and
ii) for all $x,x'\in X$ and $z\in Z$, $d_{X\cup Z}(x,z) \leq d_{X\cup Z}(x,x') + d_{X\cup Z}(x',z)$.

Since the definition of $d_{X\cup Z}$ is symmetric in x and z, we only need to check these two cases (and not the two others involving $x\in X$ and $z,z'\in Z$). Let's show these now

i) First $d_{X\cup Y}$ and $d_{X\cup Z}$ are extensions and $x,x'\in X$, we know that $d_{X\cup Z}(x,x') = d_{X\cup Y}(x,x') = d_X(x,x')$. Via the triangle inequality, for any $y\in Y$ it follows $d_{X\cup Y}(x,x') \leq d_{X\cup Y}(x,y) + d_{X\cup Y}(y,x')$. Since $d_{Y\cup Z}(y,z)\geq 0$ we have
$$\begin{aligned} d_{X\cup Z}(x,x') &= d_{X\cup Y}(x,x') \leq d_{X\cup Y}(x,y) + d_{X\cup Y}(x',y) \\ &\leq [d_{X\cup Y}(x,y) + d_{Y\cup Z}(y,z)] + [d_{X\cup Y}(x',y) + d_{Y\cup Z}(y,z)]\,. \end{aligned}$$
Taking the infimum over $y\in Y$ on both sides gives the desired result.
ii) Select $y\in Y$. As $d_{X\cup Y}$ is a metric we have $d_{X\cup Y}(x,y) \leq d_{X\cup Y}(x,x') + d_{X\cup Y}(x',y)$. But as before $d_{X\cup Y}(x,x') = d_{X\cup Z}(x,x')$. Applying this and adding $d_{Y\cup Z}(y,z)$ to both sides yields
$$d_{X\cup Y}(x,y) + d_{Y\cup Z}(y,z) \leq d_{X\cup Z}(x,x') + d_{X\cup Y}(x',y) + d_{Y\cup Z}(y,z).$$
Taking the infimum over $y\in Y$ now gives the result.

Next we need to show that

$$d_H^{(d_{X\cup Z})}(X,Z) \leq d_H^{(d_{X\cup Y})}(X,Y) + d_H^{(d_{Y\cup Z})}(Y,Z)$$

where $d_H^{(d)}$ indicates that the Hausdorff distance is taken with respect to the metric d. Let $r_{ij} = d_H^{(d_{ij})}(X_i,X_j)$. We first show that $X\subset Z^{r_{12}+r_{23}}$, i.e. for any $x\in X$ there exists a $z\in Z$ such that $d_{X\cup Z}(x,z) < r_{12}+r_{23}$. But since $r_{12} = d_H^{(d_{X\cup Y})}(X,Y)$, this means that there exists a $y\in Y$ such that $d_{X\cup Y}(x,y) < r_{12}$. Similarly since $r_{23} = d_H^{(d_{Y\cup Z})}(Y,Z)$, for the previously chosen $y\in Y$ there exists a $z\in Z$ such that $d_{Y\cup Z}(y,z) < r_{23}$. The triangle inequality for $d_{X\cup Z}$ shows that $d_{X\cup Z}(x,z) < r_{12}+r_{23}$ for some $z\in Z$ as desired. By symmetry, $Z\subset X^{r_{12}+r_{23}}$. □

4.3.2. *Correspondences.* There is yet another way to think about the Gromov-Hausdorff distance through correspondences.

Definition 4.11. A subset $R\subset X\times Y$ is called a *correspondence* if the projections of R onto X and Y are surjective. If (X,d_X) and (Y,d_Y) are metric spaces and $R\subset X\times Y$ a correspondence, the *distortion* is

$$\operatorname{dis}(R) := \sup\left\{ \left|d_X(\tilde{x},\tilde{x}') - d_Y(\tilde{y},\tilde{y}')\right| \;\middle|\; (\tilde{x},\tilde{y}),(\tilde{x}',\tilde{y}')\in R \right\}.$$

This leads us to an intrinsic characterization of the Gromov-Hausdorff distance. The previous formulations involved embedding (isometrically) both (X, d_X) and (Y, d_Y) into another metric space (whether it be some arbitrary one (Z, d_Z) or $(X \cup Y, d_{X\cup Y})$ for a metric extension $d_{X\cup Y}$).

Theorem 4.12 (Intrinsic characterization of d_{GH}, c.f. [BBI01] Theorem 7.3.25)**.** *For any metric space (X, d_X) and (Y, d_Y),*

$$d_{GH}((X,d_X),(Y,d_Y)) = \frac{1}{2} \inf \left\{ \operatorname{dis}(R) \;\middle|\; R \subset X \times Y \text{ is a correspondence} \right\}.$$

Proof. Denote the infimum above by I. First suppose that $d_{GH}(X,Y) < r$. Using the metric extension characterization, we have $d_H(X,Y) < r$ where d_H is computed using $d_{X\cup Y}$, a metric extension on $X \cup Y$. Now define

$$R := \left\{ (x,y) \in X \times Y \;\middle|\; d_{X\cup Y}(x,y) < r \right\}.$$

We claim that R is a correspondence and $\operatorname{dis}(R) \le 2r$. For any $x \in X$ there exists $y \in Y$ such that $d_{X\cup Y}(x,y) < r$. Since $(x,y) \in R$, the projection of R onto X is surjective. Similarly, the projection of R onto Y is surjective. To check the distortion, we use the triangle inequality. Let $(x,y), (\tilde{x},\tilde{y}) \in R$. As $d_{X\cup Y}$ is a metric extension, we have

$$\begin{aligned}
|d_X(x,\tilde{x}) - d_Y(y,\tilde{y})| &= |d_{X\cup Y}(x,\tilde{x}) - d_{X\cup Y}(y,\tilde{y})| \\
&= |[d_{X\cup Y}(x,\tilde{x}) - d_{X\cup Y}(\tilde{x},y)] + [d_{X\cup Y}(\tilde{x},y) - d_{X\cup Y}(y,\tilde{y})]| \\
&\le |d_{X\cup Y}(x,\tilde{x}) - d_{X\cup Y}(\tilde{x},y)| + |d_{X\cup Y}(\tilde{x},y) - d_{X\cup Y}(y,\tilde{y})| \\
&\le d_{X\cup Y}(x,y) + d_{X\cup Y}(\tilde{x},\tilde{y}) < 2r.
\end{aligned}$$

Taking the sup over pairs of points then gives $\operatorname{dis}(R) \le 2r$.

Notice if we can show $d_{GH}(X,Y) \le 1/2 \operatorname{dis}(R)$ for any correspondence R then we are done. To see this, suppose the inequality were strict. Then there exists r such that $d_{GH}(X,Y) < r < 1/2 \operatorname{dis}(R)$. But, we showed that $d_{GH}(X,Y) < r$ implies $\operatorname{dis}(R) < 2r$, a contradiction.

Given a correspondence $R \subset X \times Y$ with $\operatorname{dis}(R) = 2r$ we need to find a metric extension $d_{X\cup Y}$ of d_X and d_Y to $X \cup Y$ such that $d_{GH}(X,Y) \le r$. We claim that

$$d_{X\cup Y}(x,y) = d_{X\cup Y}(y,x) = \inf_{(\tilde{x},\tilde{y})\in R} \{d_X(x,\tilde{x}) + r + d_Y(\tilde{y},y)\}$$

for all $(x,y) \in X \times Y$ works. Let $x \in X$. Then by definition of R there exists $\tilde{y} \in Y$ such that $(x,\tilde{y}) \in R$. Testing this in the above definition, we see that $d_{X\cup Y}(x,\tilde{y}) < r$, in particular $X \subset Y^r$ (in the metric $d_{X\cup Y}$). Similarly $Y \subset X^r$ so that $d_H(X,Y) \le r$. If $d_{X\cup Y}$ is a metric, then by definition it's an extension and so $d_{GH}(X,Y) \le d_H(X,Y) \le r$.

To check that $d_{X\cup Y}$ is a metric, we just need to show it satisfies the triangle inequality. As before, we check

i) For all $x, x' \in X$ and $y \in Y$,

$$d_{X\cup Y}(x,x') \le d_{X\cup Y}(x,y) + d_{X\cup Y}(x',y).$$

To show this, let $(\tilde{x},\tilde{y}), (\tilde{x}',\tilde{y}') \in R$. By two instances of the triangle inequality and the fact that $\operatorname{dis}(R) = 2r$,

$$\begin{aligned}
d_2(\tilde{y},\tilde{y}') &\le d_Y(y,\tilde{y}) + d_Y(y,\tilde{y}') \\
d_{X\cup Y}(x,x') = d_X(x,x') &\le d_X(x,\tilde{x}) + d_X(x',\tilde{x}') + d_X(\tilde{x},\tilde{x}') \\
d_X(\tilde{x},\tilde{x}') - d_Y(\tilde{y},\tilde{y}') &< 2r
\end{aligned}$$

Combining these three together yields

$$d_{X\cup Y}(x,x') < [d_X(x,\tilde{x}) + r + d_Y(y,\tilde{y})] + [d_X(x',\tilde{x}') + r + d_Y(y,\tilde{y}')]$$

Taking the infimum over $(\tilde{x},\tilde{y}),(\tilde{x}',\tilde{y}') \in R$ on the right-hand side gives us the desired inequality. Why do we need to fuss over having two different points in R for the infimum? Consider $f,g:[0,1]\to\mathbb{R}$ by $f(x)=x$ and $g(x)=1-x$. Then,

$$\inf_{x\in[0,1]}\{f(x)+g(x)\}=1$$

but $\inf_{x\in[0,1]} f(x)=0=\inf_{x\in[0,1]} g(x)$.

ii) For all $x,x'\in X$ and $y\in Y$,

$$d_{X\cup Y}(x,y)\le d_{X\cup Y}(x,x')+d_{X\cup Y}(x',y).$$

Once more, let $(\tilde{x},\tilde{y})\in R$. By applying the triangle inequality and adding $r+d_Y(\tilde{y},y)$ to both sides we get,

$$d_X(x,\tilde{x})+r+d_Y(\tilde{y},y)\le d_X(x,x')+[d_X(x',\tilde{x})+r+d_Y(\tilde{y},y)].$$

Taking the inf over $(\tilde{x},\tilde{y})\in R$ concludes.

□

4.4. **Gromov-Hausdorff and isometry classes.** What's useful about correspondences (other than it gives us an intrinsic characterization of Gromov-Hausdorff)? In Remark 4.7 we saw that d_{GH} is only a semi-metric. The following theorem explains why.

Theorem 4.13 (c.f. [BBI01] Theorem 7.3.30)**.** *For every pair of compact metric spaces (X,d_X) and (Y,d_Y), we have $d_{GH}((X,d_X),(Y,d_Y))\ge 0$ with equality if and only if (X,d_X) is isometric to (Y,d_Y).*

To prove this we need a few more definitions.

Definition 4.14. A (possibly discontinuous) map $f:X\to Y$ between metric spaces is called an *ϵ-isometry* if

i) $\mathrm{dis}(\mathrm{Graph}(f))<\epsilon$ and
ii) $f(X)$ is an ϵ-net in Y, i.e. $Y\subset\bigcup_{x\in X}B_\epsilon(f(x))$.

Corollary 4.15 (c.f. [BBI01] Theorem 7.3.28)**.** *Let (X,d_X) and (Y,d_Y) be metric spaces.*

i) If $d_{GH}(X,Y)<\epsilon$ then there exists a 2ϵ-isometry from X to Y and
ii) if there exists an ϵ-isometry from X to Y then $d_{GH}(X,Y)\le 3\epsilon/2$.

Proof. We'll use correspondences to show this.

i) If $d_{GH}(X,Y)<\epsilon$ then there exists a correspondence $R\subset X\times Y$ such that $\mathrm{dis}(R)<2\epsilon$. We'd like to turn this correspondence to a mapping. For every $x\in X$ there exists (possibly multiple) $\tilde{y}\in Y$ such that $(x,\tilde{y})\in R$. Set $f(x)=\tilde{y}$ for one of these possible candidates (this uses the axiom of choice). We claim this map gives a 2ϵ-isometry. Now $\mathrm{Graph}(f)$ is still a correspondence, and moreover $\mathrm{Graph}(f)\subset R$. Since the distortion is a supremum, we have $\mathrm{dis}(\mathrm{Graph}(f))\le\mathrm{dis}(R)<2\epsilon$. It remains to show $f(X)$ is a 2ϵ-net. Let $y\in Y$. Then there exists $\tilde{x}\in X$ such that $(\tilde{x},y)\in R$. Consider $\tilde{y}=f(\tilde{x})$, which by definition is such that $(\tilde{x},\tilde{y})\in R$. Testing both these points in the definition of the distortion gives

$$d_Y(y,f(\tilde{x}))=|d_X(\tilde{x},\tilde{x})-d_Y(y,\tilde{y})|\le\mathrm{dis}(R)<2\epsilon$$

so that $y\in B_{2\epsilon}(f(\tilde{x}))$.

ii) Given a ϵ-isometry $f:X\to Y$, we need to define a relation with $\mathrm{dis}(R)\le 3\epsilon$. For instance,

$$R:=\{(x,y)\in X\times Y\mid d_Y(y,f(x))<\epsilon\}$$

defines a correspondence with $\mathrm{dis}(R)\le 3\epsilon$. Let $(\tilde{x},\tilde{y})$ and $(\tilde{x}',\tilde{y}')$ be in R. Since

$$d_Y(\tilde{y},\tilde{y}')\le d_Y(\tilde{y},f(\tilde{x}))+d_Y(f(\tilde{x}),f(\tilde{x}'))+d_Y(f(\tilde{x}'),\tilde{y}')<2\epsilon+d_Y(f(\tilde{x}),f(\tilde{x}'))$$

we can estimate the distortion as:

$$|d_Y(\tilde{y},\tilde{y}')-d_Y(\tilde{x},\tilde{x}')|\le 2\epsilon+|d_Y(f(\tilde{x}),f(\tilde{x}'))-d_X(\tilde{x},\tilde{x}')|.$$

Taking the sup over $(\tilde{x},\tilde{y}),(\tilde{x}',\tilde{y}')\in R$ the yields

$$\mathrm{dis}(R)\le 2\epsilon+\mathrm{dis}(\mathrm{Graph}(f))\le 3\epsilon$$

as desired.

□

We're now ready to prove Theorem 4.13

Proof of Theorem 4.13. Since X is compact, there exists a dense $\{x_n\}_{n=1}^\infty \subset X$. As $d_{GH}(X,Y)=0$, by Corollary 4.15 for every $k \in \mathbb{N}$ there exists a $1/k$-isometry $f_k : X \to Y$ (apply part i) with $\epsilon = 1/(2k)$).

Let $y_n^k = f_k(x_n)$. By compactness of Y, for every $n \in \mathbb{N}$ there exists a subsequence (in k) with $y_n^k \to y_n$. By a diagonal argument there exists a subsequence $y_n^{k(j)}$ such that

$$\lim_{j\to\infty} y_n^{k(j)} \to y_n$$

for all $n \in \mathbb{N}$. We usually won't bother with relabelling in the future. Since $\mathrm{dis}(\mathrm{Graph}(f_k)) \le 1/k$,

$$|d_Y(y_n^k, y_m^k) - d_X(x_n, x_m)| \le \frac{1}{k}$$

and by taking $k \to \infty$ we see

$$|d_Y(y_n, y_m) - d_X(x_n, x_m)| = 0$$

for all $n, m \in \mathbb{N}$. Setting $f(x_n) = y_n$, we see that f is an isometry onto its image. But the domain of f is dense in X, and f is 1-Lipschitz because it is an isometry. So we can extend f to all of X isometrically. Now, is the image $f(X)$ all of Y? For this part we use the fact that each $f_k(X)$ is a $1/k$-net in Y to show that $\{y_n\}_{n=1}^\infty$ is dense in Y.

To start, let $\epsilon > 0$ and choose $y \in Y$. Then there exists $k \in \mathbb{N}$ such that $1/k < \epsilon/3$. As f_k is a $1/k$-net, there exists some $x \in X$ such that $d_Y(y, f_k(x)) < 1/k$. Since $\{x_n\}_{n=1}^\infty$ is dense in X, there exists x_n such that $d_X(x, x_n) < 1/k$. Because f_k is a $1/k$-isometry, $\mathrm{dis}(\mathrm{Graph}(f_k)) < 1/k$; in particular $d_Y(f_k(x), f_k(x_n)) < \mathrm{dis}(\mathrm{Graph}(f_k)) + d_X(x_1, x_n) < 2/k$. Combining these facts together yields

$$d_Y(y, y_n) \le d_Y(y, f_k(x)) + d_Y(f_k(x), f_k(x_n)) + d_Y(f_k(x_n), y_n) < \frac{1}{k} + \frac{2}{k} + d_Y(y_n^k, y_n) < \epsilon + d_Y(y_n^k, y_n).$$

Since y_n^k converges to y_n, taking the limit as $k \to \infty$ yields $d_Y(y, y_n) \le \epsilon$. This holds for all $\epsilon > 0$, $\{y_n\}_{n=1}^\infty$ is dense in Y. Finally, f is 1-Lipschitz and therefore continuous. As X and Y are compact, f maps closed sets to closed sets. In particular,

$$f(X) = f(\mathrm{Cl}(\{x_n\}_{n=1}^\infty)) = \mathrm{Cl}(f(\{x_n\}_{n=1}^\infty)) = \mathrm{Cl}(\{y_n\}_{n=1}^\infty) = Y,$$

i.e. f is surjective. □

Corollary 4.16. *d_{GH} gives a metric on isometry classes of compact metric spaces.*

4.5. **Reductions to finite sets.** When showing properties of the Hausdorff distance we sometimes worked with a finite set. The following gives the appropriate generalization for the Gromov-Hausdorff distance.

Definition 4.17. We say metric space (X, d_X) and (Y, d_Y) are (δ,ϵ)-*approximations* to each other if there exist $X' \subset X$ and $Y' \subset Y$ such that $N = |X'| = |Y'| < \infty$ and there exists a correspondence $R = \{(x_i, y_i)\}_{i=1}^N \subset X' \times Y'$ with distortion $\mathrm{dis}(R) \le \delta$ and X_i' is an ϵ-net in X_i for $i = 1, 2$. If $\delta = \epsilon$ we simply write ϵ-approximation.

Here is an example of the utility of this definition. Showing that (δ, ϵ)-approximations are close in the Gromov-Hausdorff distance.

Theorem 4.18 (c.f. [BBI01] Theorem 7.4.11)**.** *Given compact metric space (X, d_X) and (Y, d_Y),*

a) If Y is a (δ,ϵ)-approximation of X then $d_{GH}(X,Y) \le 2\epsilon + \delta$.
b) If $d_{GH}(X,Y) < \epsilon$ then Y is a 5ϵ-approximation of X.

Proof. The proof is essentially just applying the definitions.

a) If Y is a (δ,ϵ)-approximation of X, then there exists $R = \{(x_i, y_i)\}_{i=1}^N$ with $\text{dis}(R) \leq \delta$ and projections $X_i' \subset X_i$ giving ϵ-nets. Since R is a correspondence on $X' \times Y'$,
$$d_{GH}(X', Y') \leq \frac{1}{2} < \delta.$$
Now since X_i' is already isometrically embedded in X_i, computing the Gromov-Hausdorff distance amounts to computing the Hausdorff distance. And, as X_i' is an ϵ-net in X_i, we know that
$$d_{GH}(X_i', X_i) = d_H(X_i', X_i) \leq \epsilon.$$
Next, using the triangle inequality
$$d_{GH}(X, Y) \leq d_{GH}(X', X_i) + d_{GH}(Y', Y) + d_{GH}(X', Y') < 2\epsilon + \delta.$$

b) If $d_{GH}(X,Y) < \epsilon$ then there exists a 2ϵ-isometry $f : X \to Y$. By compactness there exists a finite ϵ-net $X' = \{x_i\}_{i=1}^N \subset X$. Define $Y' = f(X')$ and note that $f(X)$ is a 2ϵ-net in Y. Then, if $y \in Y$ we can find $x \in X$ such that $d_Y(y, f(x)) < 2\epsilon$ and we can find $x_i \in X'$ such that $d_X(x, x_i) < \epsilon$. Since $\text{dis}(\text{Graph}(f)) < 2\epsilon$, we also get $d_Y(f(x), f(x_i)) < \text{dis}(\text{Graph}(f)) + d_X(x, x_i) < 3\epsilon$. Combining these gives
$$d_Y(y, f(x_i)) \leq d_Y(y, f(x)) + d_Y(f(x), f(x_i)) < 5\epsilon,$$
showing that Y' is a 5ϵ-net in Y (as $f(x_i) \in Y'$).

We must also check that there exists a correspondence R with $\text{dis}(R) \leq 5\epsilon$. Selecting $R = \{(x_i, f(x_i))\}_{i=1}^N \subset \text{Graph}(f)$, we know that $\text{dis}(R) \leq \text{dis}(\text{Graph}(f)) < 2\epsilon < 5\epsilon$.

□

With this in hand, we can recast Gromov-Hausdorff convergence as the existence of arbitrarily good (δ,ϵ)-approximations.

Proposition 4.19 (Finite d_{GH} convergence criteria, c.f. [BBI01] Proposition 7.4.12)**.** *If (X_n, d_n) is a sequence of compact metric spaces then $d_{GH}(X_n, X) \to 0$ if and only if for every $\epsilon > 0$ there exist finite ϵ-nets $S_n \subset X_n$ for all $n \in \mathbb{N}$ such that $d_{GH}(S_n, S) \to 0$. Moreover, for n sufficiently large $|S_n| = |S|$.*

In other words, to check Gromov-Hausdorff convergence it's enough to check on finite subsets (provided they can be taken sufficiently fine).

Proof. We show the reverse direction (which is more important for us). Assume for all $\epsilon > 0$ that there exist ϵ-nets $S_n \subset X_n$ as above. Since $d_{GH}(S_n, S) \to 0$, for some N large enough and $n \geq N$ we can find a correspondence R_n such that $\text{dis}(R_n) < \epsilon$. Because S_n and S are by assumption ϵ-nets in X_n and X respectively and eventually have the same cardinality, we see that (X_n, d_n) is an ϵ-approximation to (X, d) for all n sufficiently large. Now, by applying Theorem 4.18 we have $d_{GH}(X_n, X) \leq 3\epsilon$ for all $n \geq N$, i.e. X_n Gromov-Hausdorff converges to X. □

4.6. **Stability of uniformly totally bounded classes of compact metric spaces.** Before continuing we need one quick definition.

Definition 4.20. A class $\mathcal{X}$ of metric spaces is *uniformly totally bounded* if for every $\epsilon > 0$ there exists $N(\epsilon)$ such that every $(X, d) \in \mathcal{X}$ is covered by at most $N(\epsilon)$ many ϵ-balls and $\max_{(X,d)\in\mathcal{X}} \text{Diam}(X) = D < \infty$.

Remark 4.21. The diameter bound can be removed in length spaces, as the diameter will be bounded by something like $2\epsilon N(\epsilon)$.

We're finally ready to prove a compactness theorem for d_{GH} (which serves as our starting point for a compactness theorem of $CD^e(K,N)$ spaces).

Theorem 4.22 (Gromov pre-compactness theorem, c.f. [BBI01] Theorem 7.4.15)**.** *Any uniformly totally bounded class $\mathcal{X}$ of compact metric spaces is pre-compact. I.e., any sequence of spaces $(X_n, d_n) \in \mathcal{X}$ has a subsequence admitting a d_{GH}-limit.*

Proof. Let D and $N(\epsilon)$ be the uniform bounds on the diameter and number of ϵ-nets in $\mathcal{X}$. Set $N_1 = N(1)$ and for $k > 1$ define inductively $N_k = N_{k-1} + N(1/k)$. Given any sequence $\{(X_n, d_n)\}_{n=1}^{\infty} \subset \mathcal{X}$, for every $n \in \mathbb{N}$ there exists a dense set $S_n = \{x_n^i\}_{i=1}^{\infty} \subset X_n$ such that the first N_k points in S_n form a $1/k$ net in X_n. So for any n, we look at the points necessary to make a 1-net first (there are $N(1)$ many of these) and call them $x_n^1, ..., x_n^{N_1}$. Then we add in the points necessary to make a 1/2-net (there are $N(2)$ many of these) and call them $x_n^{N_1+1}, ..., x_n^{N_2}$. In this way, the first N_k many points form a $1/k$-net in X_n (we have many leftover points, but these are irrelevant).

Now for all $i, j \in \mathbb{N}$,

$$0 \le \max_{n \in \mathbb{N}} d_n(x_n^i, x_n^j) \le D,$$

and so along some subsequence $\lim_{n\to\infty} d_n(x_n^i, x_n^j)$ exists. This may depend on n, but using a diagonal argument we can get all these subsequences to converge simultaneously. Set $d(i, j) = \lim_{n\to\infty} d_n(x_n^i, x_n^j)$. The Gromov-Hausdroff limit (along this subsequence) is constructed as follows. First, note that

$$d(i, j) \le d(i, k) + d(k, j),$$

inherited from the fact that each d_n obeys the triangle inequality. So, $d(i, j)$ is a $[0, D]$-valued semi-metric on $\mathbb{N}$. Letting $\sim$ be the equivalence relation identifying natural numbers i, j such that $d(i, j) = 0$, we see that $X = \mathbb{N}/\sim$ is a metric space for d. This may not be complete, but we can always take its completion $\mathrm{Cl}(X)$ with respect to d. We claim this is a Gromov-Hausdorff limit.

It remains to show that $\mathrm{Cl}(X)$ is compact and $d_{GH}(X_n, \mathrm{Cl}(X)) \to 0$ (along the subsequence – we won't mention this point anymore). To show compactness, we only need to show that $(\mathrm{Cl}(X), d)$ is totally bounded. We claim for all $k \in \mathbb{N}$ that $S^k := \{[i] \in \mathbb{N}/\sim \mid i \le N_k\}$ is a $1/k$-net in $\mathrm{Cl}(X)$. By construction, $S_n^k := \{x_n^j \mid j \le N_k\}$ is a $1/k$-net in X_n. Let $i > N_k$ (if $i \le N_k$, then $[i] \in S^k$ and there is nothing to check), we wish to find $[j(i)] \in S^k$ such that $d([i], j(i)] \le 1/k$. For any n we can find $x_n^j \in S_n^k$ with $d_n(x_n^i, x_n^j) \le 1/k$. Although j depends on i and n, for every i there are infinitely many such j and $j(i, n) \le N_k$. By the pigeonhole principle, infinitely many must lie in the same equivalence class which we call $[j(i)]$. For fixed i, each such $j \in [j(i)]$ corresponds to some n, so we obtain a subsequence of n, and along this subsequence we have $d_n(x_n^i, x_n^j) \to d([i], [j(i)])$ (essentially just by definition). So, $d([i], [j(i)]) \le 1/k$, where $[j(i)] \in S^k$.

To show $d_{GH}(X_n, \mathrm{Cl}(X)) \to 0$, we use the finite convergence criterion and show $d_{GH}(S^k, S_n^k) \to 0$ as $n \to \infty$. To do this, define the following correspondence R^k:

$$R^k := \left\{([i], x_n^j) \in S^k \times S_n^k \mid j \in [i]\right\}.$$

Given $\epsilon > 0$ and $i, j \le N_k$, there exists $N \in \mathbb{N}$ such that if $n \ge N$ then

$$|d_n(x_n^i, x_n^j) - d([i], [j])| < 2\epsilon.$$

The N above may be taken uniform in i, j (since there are only finitely many possible combinations of indices). Thus, taking the sup over $([i], x_n^i) \in R^k$, we get $d_{GH}(S_n^k, S^k) \le 1/2\, \mathrm{dis}(R) < \epsilon$. ☐

4.7. **Pointed Gromov-Hausdorff convergence and stability.** We'd like a version of this for non-compact spaces (for $K < 0$, a $CD^e(K, N)$ space may not be compact). To do this, we need a new notion of Gromov-Hausdorff convergence. The idea is to fix a sequence of points and say what those points correspond with in each space along the sequence. Henceforth, we write X for a metric space (X, d).

Definition 4.23. Let (X_n, p_n) be a sequence of metric spaces with distinguished point p_n. We say (X_n, p_n) converges in the *pointed Gromov-Hausdorff* sense to (X, p) if for every $0 < r, R$ there exists $N(r, R)$ such that if $n \ge N(r, R)$ there exists a (possibly discontinuous) map $f_n : B_R(p_n) \to X$ such that

i) $f_n(p_n) \to p$,
ii) $\mathrm{dis}(\mathrm{Graph}(f_n)) \le r$,
iii) $B_{R-r}(p) \subset f_n(B_R(p_n))^r$.

We denote this with $(X_n, p_n) \xrightarrow{GH} (X, p)$.

The second condition is like being an ϵ-isometry with size r, but instead of globally it is only localized to the points in the sequence, while the third says the r-neighborhood of $f_n(B_R(p_n))$ in X contains $B_{R-r}(p)$.

If the limiting space X is a length space and $(X_n, p_n) \xrightarrow{GH} (X, p)$ then $B_R^{d_n}(p_n) \xrightarrow{GH} B_R^d(p)$ for every $R > 0$. This is somehow what you want. If for example the X_n are σ-compact, then applying this for arbitrarily large balls seems like a good strategy. But we don't necessarily have length spaces, hence the more abstract Definition 4.23.

Remark 4.24. We have the same technicality as with the Gromov-Hausdorff distance, namely that if $(X_n, p_n) \xrightarrow{GH} (X, p)$ then $(X_n, p_n) \xrightarrow{GH} (\mathrm{Cl}(X), p)$. So, we might as well assume that our space are complete as this convergence cannot distinguish them.

We need a couple more quick results before continuing.

Theorem 4.25 (c.f. [BBI01] Theorem 8.1.7)**.** *Let (X, p) and (X', p') both be complete pointed Gromov-Hausdorff limits of $\{(X_n, p_n)\}_{n=1}^{\infty}$. If (X, p) is boundedly compact then X and X' are pointedly isometric, i.e. there exists an isometry $f : X \to X'$ with $f(p) = f(p')$.*

We proved earlier that on compact sets the Gromov-Hausdorff convergence defines a metric on isometry classes. What this theorem is telling us is that remains true in the non-compact case provided we restrict ourselves to boundedly compact.

Proof. Similar to showing d_{GH} metrizes the isometry classes of compact metric spaces. Except, you need to use a diagonal argument twice. First as $r \to 0$, then as $R \to \infty$. □

With this, we can extend our pre-compactness result to non-compact spaces.

Theorem 4.26 (Pointed pre-compactness, c.f. [BBI01] Theorem 8.1.10)**.** *Let $\mathcal{X}$ be a class of pointed, boundedly compact metric spaces such that for every $r, R > 0$ there exists $N(r, R)$ such that $(X, p) \in \mathcal{X}$ the ball $B_R(p)$ in X admits an r-net of no more than $N(r, R)$ points. Then $\mathcal{X}$ is pre-compact in the sense that any sequence in $\mathcal{X}$ admits a (boundedly compact) pointed Gromov-Hausdorff limit.*

The proof is virtually the same, again with the caveat of needing two diagonal arguments.

5. Stability of $CD^e(K, N)$ Spaces.

To get us started with limits of $CD^e(K, N)$ spaces, here is some philosophy: convexity of $u : \mathbb{R}^n \to \mathbb{R}$ can be checked in two ways

a) If $u \in C^2(\mathbb{R}^n)$ then u is convex if and only if $D^2 u(x) \geq 0$ for all $x \in \mathbb{R}^n$.
b) If $u \in C^0(\mathbb{R}^n)$ then u is convex if and only if $u((x_0 + x_1)/2) \leq u(x_0)/2 + u(x_1)/2$ for all $x_0, x_1 \in \mathbb{R}^n$.

The latter is a more general definition but it is harder to check (it is non-local, as you need to look at all pairs of points). When you're computing things, you use the former. But the latter is more useful when deducing corollaries of convexity. So when we're working with limits of $CD^e(K, N)$ spaces, we'll use the latter (limits hold in the much weaker C^0 sense rather than the C^2 sense).

5.1. **Pre-compactness of $CD^e(K, N)$.** Now how is all of Section 4 useful for us? We have the following theorem (introduced as the "meta" theorem 4.8 previously).

Corollary 5.1. *For every $(K, N) \in \mathbb{R} \times [1, \infty)$ the class $CD^e(K, N)$ is pointedly Gromov-Hausdorff pre-compact.*

Remark 5.2. In fact, $CD^e(K, N)$ is pointedly Gromov-Hausdorff compact. We just haven't said yet what the measure on the limiting space should be in order to limit the curvature dimension bound of the approximating sequence.

Proof. Let $(X,d,m) \in CD^e(K,N)$ and fix $p \in X$. Given $0 < r < R < \infty$, Bishop-Gromov tells us for some $M(r,R,K,N)$ then

$$m(B_r(p)) \geq M(r,R,K,N)m(B_R(p))$$

provided $B_r(p) \subset B_R(p) \subset X$. In particular, there exists a r-net in $B_R(p)$ with cardinality at most $1/M(r,R,K,N)$ – use the usual maximal collection and doubling argument.[9] We checked in Corollary 3.16 that $CD^e(K,N)$ spaces are boundedly compact, so by Theorem 4.26 some subsequence pointedly Gromov-Hausdorff converges. □

5.2. **Approximate inverses and limits of geodesics.** We now work towards showing that a (pointed) Gromov-Hausdorff limit of $CD^e(K,N)$ spaces is in fact a $CD^e(K,N)$ space. In order to test that a space is in $CD^e(K,N)$, we check the convexity of the entropy along geodesics. It'll be useful for us to know when you have a sequence of metric spaces converging to a limiting space, then geodesics in the sequence of spaces converge to a geodesic in the limit space. We define first

$$\text{Geo}(X) = \text{Geo}(X,d) = \{\sigma : [0,1] \to X \mid d(\sigma(s),\sigma(t)) = |t-s|d(\sigma(0),\sigma(1))\},$$

i.e. the set of geodesics in (X,d). We also introduce the concept of approximate inverses

Definition 5.3. An r-isometry $f : X \to Y$ has an *approximate inverse* $\tilde{f} : Y \to X$ defined as: for all $y \in Y$ there exists $x \in X$ such that $d_Y(y, f(x)) < r$ (of course, since $f(X)$ is an r-net there could be multiple). Set $\tilde{f}(y) = x$.

There are some easy consequences we can prove. For example, $\tilde{f} : Y \to X$ is a $3r$-isometry, $d_X(\tilde{f}(f(x)), x) \leq 2r$, and $d_Y(y, f(\tilde{f}(y))) \leq r$.

To see this, let $y, y' \in Y$ and choose $x, x' \in X$ such that $d_Y(y, f(x)) < r$ (and similarly for y' and $f(x')$). Since f is an r-isometry, we have that

$$|d_Y(f(x), f(x')) - d_X(x,x')| < r.$$

Then,

$$\begin{aligned} d_X(x,x') &\leq r + d_Y(f(x), f(x')) \\ &\leq r + d_Y(f(x), y) + d_Y(y,y') + d_Y(y', f(x')) < 3r + d_Y(y,y'). \end{aligned}$$

In particular, the above holds for $x = \tilde{f}(y)$ and $x' = \tilde{f}(y')$. Thus,

$$d_X(\tilde{f}(y), \tilde{f}(y')) - d_Y(y,y') < 3r.$$

Similar work shows the reverse inequality so that $\tilde{f}$ is a $3r$-isometry. Next, for $x \in X$, setting $y = f(x)$ we know the approximate inverse satisfies $d_Y(y, f(\tilde{f}(y))) < r$. Since f is an r-isometry, we have that

$$d_X(x, \tilde{f}(f(x))) = d_X(x, \tilde{f}(y)) < r + d_Y(f(x), f(\tilde{f}(y))) = r + d_Y(y, f(\tilde{f}(y))) < 2r.$$

The last inequality is the easiest. For $y \in Y$, by definition $x = \tilde{f}(y)$ is such that $d_Y(y, f(x)) < r$. But this just means $d_Y(y, f(\tilde{f}(y))) < r$. Using the idea of approximate inverses, we can prove an Arzela-Ascoli type result for ϵ-isometries.

Lemma 5.4 (Non-smooth Arzela-Ascoli, c.f. [LV09], Appendix A. [GP91])**.** *Let (X, d_X) and (Y, d_Y) be compact spaces. If $f^i : X_i \to X$ and $g^i : Y_i \to Y$ are ϵ_i-isometries with $\epsilon_i \to 0$ and $\tilde{f}^i : X \to X_i$, $\tilde{g}^i : Y \to Y_i$ approximate inverses as above, and $h^i : X_i \to Y_i$ are asymptotically equicontinuous*[10]*, then a subsequence of the (possibly discontinuous) maps $g^i \circ h^i \circ \tilde{f}^i : X \to Y$ converges uniformly to a continuous limit.*

[9] I'm not sure I fully believe this.

[10] Meaning, for every $\epsilon > 0$ there is a $\delta > 0$ and $N < \infty$ such that for all $i \geq N$ we have $d_{X_i}(x_i, x_i') < \delta$ implies $d_Y(h^i(x_i), h^i(x_i')) < \epsilon$

The essence is as follows: because the f^i and g^i are ϵ_i-isometries with $\epsilon_i \to 0$, they do not distort continuity too much. Moreover, the h^i are continuous at scale ϵ for i large enough and so continuous (in fact equicontinuous) as $i \to \infty$.

Really we'll be applying a specific case of this with $X = [0,1]$, $f^i(t) = t = \tilde{f}^i(t)$, and $h^i \in \mathrm{Geo}(Y_i, d_{Y_i})$ to prove that limits of geodesics are geodesics.

Corollary 5.5 (c.f. [LV09] Corollary 4.3)**.** *If $g^i : Y_i \to Y$ are ϵ_i-isometries of compact metric spaces with $\epsilon_i \to 0$ and $\sigma^i \in \mathrm{Geo}(Y_i)$ then $g_i \circ \sigma^i$ converges uniformly (along a subsequence) to $\sigma \in \mathrm{Geo}(Y)$. Moreover, $d_{Y_i}(\sigma^i(0), \sigma^i(1)) \to d_Y(\sigma(0), \sigma(1))$.*

Proof. Applying Lemma 5.4 shows that $\sup_{t\in[0,1]} d_Y(g_i(\sigma^i(t)), \sigma(t)) \to 0$ along a subsequence, and it provides $\sigma \in C([0,1], Y)$. We want to show that σ is in fact a geodesic. For every $s, t \in [0,1]$ we have

$$d_Y(\sigma(s), \sigma(t)) \leq d_Y(\sigma(s), g^i(\sigma^i(s))) + d_Y(g_i(\sigma^i(s)), g_i(\sigma^i(t))) + d_Y(g_i(\sigma^i(t)), \sigma(t)).$$

For i large, uniform convergence of $g_i \circ \sigma^i$ to σ (restated as the limit of the sup distances before) goes to zero. Because the g_i are ϵ_i isometries we have

$$\begin{aligned} d_Y(g_i(\sigma^i(s)), g_i(\sigma^i(t))) &\leq \epsilon_i + d_{Y_i}(\sigma^i(s), \sigma^i(t)) = \epsilon_i + |s-t| d_{Y_i}(\sigma^i(0), \sigma^i(1)) \\ &\leq 2\epsilon_i + |s-t| d_Y(g^i(\sigma^i(0)), g^i(\sigma^i(1))) \\ &\leq 2\epsilon_i + |s-t| d_Y(g^i(\sigma^i(0)), \sigma(0)) + |s-t| d_Y(\sigma(0), \sigma(1)) \\ &\qquad + |s-t| d_Y(\sigma(1), g^i(\sigma^i(1))). \end{aligned}$$

Once more by uniform convergence, the first and third distance in the final inequality go to zero. So, taking the limit as $i \to \infty$ in all we get

$$d_Y(\sigma(s), \sigma(t)) \leq |s-t| d_Y(\sigma(0), \sigma(1)),$$

which is all we need by the triangle inequality: For $0 \leq s \leq t \leq 1$,

$$\begin{aligned} d_Y(\sigma(0), \sigma(1)) &\leq d_Y(\sigma(0), \sigma(s)) + d_Y(\sigma(s), \sigma(t)) + d_Y(\sigma(t), \sigma(1)) \\ &\leq s d_Y(\sigma(0), \sigma(1)) + (t-s) d_Y(\sigma(0), \sigma(1)) + (1-t) d_Y(\sigma(0), \sigma(1)) \\ &\leq d_Y(\sigma(0), \sigma(1)) \end{aligned}$$

and so equality holds throughout. To show the final convergence, we use the fact that g^i is an ϵ_i isometry:

$$\begin{aligned} d_Y(\sigma(0), \sigma(1)) - d_{Y_i}(\sigma^i(0), \sigma^i(1)) &\leq d_Y(\sigma(0), g^i(\sigma^i(0))) + d_Y(g^i(\sigma^i(0)), g^i(\sigma^i(1))) \\ &\qquad + d_Y(g^i(\sigma^i(1)), \sigma^i(1)) - d_{Y_i}(\sigma^i(0), \sigma^i(1)) \\ &\leq o(1) + \epsilon_i \end{aligned}$$

which goes to zero as $i \to \infty$. The other direction is similar. □

5.3. **Lifting results to probability measures.** The geodesics we're interested in are geodesics of probability measures. Another crucial proposition is the following, which is inspired by Lott and Villani but they botched it, so we have to correct it.

Proposition 5.6 ([LV09] Theorem 4.1 corrected)**.** *If $f : X \to Y$ is an r-isometry (of compact spaces)[11] then $f_\# : (\mathcal{P}(X), d_2^Y) \to (\mathcal{P}(Y), d_2^Y)$ is a $7r$-isometry.*

[11] This may not matter much.

Proof. Given $\mu_0,\mu_1\in\mathcal{P}(X)$ let $\gamma\in\Gamma(\mu_0,\mu_1)$ be d_X^2-optimal. Set $\gamma' := (f\times f)_\#\gamma\in\Gamma(f_\#\mu_0,f_\#\mu_1)$. Then,

$$\begin{aligned}
[d_2^Y(f_\#\mu_0,f_\#\mu_1)]^2 &\le \int_{Y\times Y} d_Y(y,y')^2\ d\gamma'(y,y')\\
&= \int_{X\times X} d_Y(f(x),f(x'))^2\ d\gamma(x,x') \le \int_{X\times X}(d_X(x,x')+r)^2\ d\gamma(x,x')\\
&\le r^2+[d_2^X(\mu_0,\mu_1)]^2+2r\int_{X\times X} d_X(x,x')\ d\gamma(x,x')\\
&\le r^2+[d_2^X(\mu_0,\mu_1)]^2+2r\left(\int_{X\times X} d_X(x,x')^2\ d\gamma(x,x')\right)^{1/2}\\
&= (r+d_2^X(\mu_0,\mu_1))^2
\end{aligned}$$

where we used Jensen's inequality. Hence, $d_2^Y(f_\#\mu_0,f_\#\mu_1)\le r+d_2^X(\mu_0,\mu_1)$. For the other inequality, recall that if $\tilde f: Y\to X$ is an approximate inverse then it is a $3r$-isometry. Applying the same argument with $f_\#\mu_0, f_\#\mu_1$ and $\tilde f_\#$ in place of μ_0,μ_1 and $f_\#$, we see that

$$d_2^X(\tilde f_\# f_\#\mu_0,\tilde f_\# f_\#\mu_1)\le 3r+d_2^Y(f_\#\mu_0,f_\#\mu_1).$$

But since $\tilde f$ is a $3r$-isometry and f is an r-isometry, the composition $\tilde f\circ f$ is a $2r$-isometry.[12] Hence by the above work again, for $i=1,2$

$$d_2^X(\mu_i,\tilde f_\# f_\#\mu_i)\le 2r+d_2^X(\mu_i,\mu_i)=2r.$$

Combining everything with the triangle inequality gives $d_2^X(\mu_0,\mu_1)\le 7r+d_2^Y(f_\#\mu_0,f_\#\mu_1)$.

We must show now that $f_\#(\mathcal{P}(X))$ is a $7r$-net in $\mathcal{P}(Y)$. For all $\nu\in\mathcal{P}(Y)$, consider $f_\#\tilde f_\#\nu$. Since $d_Y(f(\tilde f(y)),y)\le 2r$ for all $y\in Y$ (in fact, less than r) we get

$$d_2^Y(\nu,(f\circ\tilde f)_\#\nu)^2\le\int_Y d_Y(y,(f\circ\tilde f)(y))^2\ d\nu(y)\le 4r^2\nu(Y)=4r^2,$$

that is $d_2^Y(\nu,(f\circ\tilde f)_\#\nu)\le 2r$. Letting $\mu=\tilde f_\#\mu$ we are done, and actually we have shown that $f_\#(\mathcal{P}(X))$ is a $2r$-net in $\mathcal{P}(Y)$. □

Corollary 5.7. *If $d_{GH}(X_i,X)\to 0$ along a sequence of compact spaces with compact limit then $d_{GH}(\mathcal{P}(X),\mathcal{P}(Y))\to 0$ as well.*

5.4. **Behavior of the entropy along limits.** We now know if we have d_2-geodesics in the approximating spaces $\mathcal{P}(X_i)$ then we can extract (from our Arzela-Ascoli type result) a geodesic in the limiting space $\mathcal{P}(X)$ whose length is the limit of the lengths of the approximating geodesics (along the subsequence). The other thing we need to know is how the entropy will behave along such limits. The proper proof is in [LV09].[13]

Proposition 5.8 (c.f. [Vil09] Proposition 29.19)**.** *Let (X,d) be a compact metric space and $m,\mu\in\mathscr{M}_+(X)$ Radon measures. Setting $U(\rho)=\rho\log\rho$, the Legendre transform is $U^*(s)=e^{s-1}$. This yields*

$$\begin{aligned}
S(\mu\mid m) &:= \int_X \frac{d\mu}{dm}\log\frac{d\mu}{dm}\ dm\\
&= \sup_{\phi\in C(X)}\left[\int_X \phi\ d\mu-\int_X U^*(\phi(x))\ dm(x)\right]\\
&= \sup_{\phi\in L^\infty(X)}\left[\int_X \phi\ d\mu-\int_X U^*(\phi(x))\ dm(x)\right]
\end{aligned}$$

for $\mu\in\mathrm{Dom}(S(\cdot\mid m))$.

[12] Robert claims this, but I don't see how. When I try to prove it I get a $4r$-isometry, which is good enough I think.

[13] Robert remarks here that [Vil09] is somehow "terribly convoluted and terribly complicated."

Remark 5.9. If (X,d) is Polish, so is $(C([0,1],X),\|\cdot\|_\infty)$. Since $\mathrm{Geo}(X)\subset C[0,1]$, Villani likes to represent a d_2-geodesic $(\mu_t)_{t\in[0,1]}\subset\mathcal{P}(X)$ using a measure $\pi\in\mathcal{P}(\mathrm{Geo}(X))$ such that $\mu_t=(e_t)_\#\pi$ where $e_t:\sigma\in\mathrm{Geo}(X)\to\sigma(t)\in X$. We typically represent μ_t using $\gamma\in\Gamma(\mu_0,\mu_1)$ optimal. We can construct π using a measurable selection $\Sigma:(x,y)\in X\times X\to\sigma_{x,y}\in\mathrm{Geo}(X)$ such that $\sigma_{x,y}(0)=x$ and $\sigma_{x,y}(1)=y$. This is a non-constructive proof as it goes through the axiom of choice, but [LV09] actually give a constructive proof.

The use of these so-called dynamic transference plans is one of the reasons why Robert thinks [Vil09] is complicated. For example, it's less natural to think of measures on $\mathcal{P}(\mathrm{Geo}(X))$ than it is to think of geodesics on $\mathcal{P}(X)$. Another is the following: before [EKS15] developed entropic curvature dimension, people used Boltzmann entropy only to talk about $CD(K,\infty)$ and used entropies[14] of the form

$$E_N(\mu\mid m)=\int_X\rho^{1-1/N}\ dm$$

with $\rho=d\mu/dm$ for $CD(0,N)$. Because of this, $CD(K,N)$ is hard to write down in terms of either entropy unless you have the notion of K,N-convexity. So there were these very complicated formulas and [LV09] used these very complicated definitions. The payoff is that they obtained sharp constants (e.g. in the Bishop-Gromov). It's a big pain to show that these $CD(K,N)$ spaces are essentially equivalent to $CD^e(K,N)$ spaces, see e.g. the pre-print [CM] by Cavalletti-Milman which "may someday appear" (it has been a preprint for almost five years).

Proof. We give the idea, since the proof takes several pages. We have that

$$U\left(\frac{d\mu}{dm}(x)\right)=\sup_{\phi\in C(X)}\left(\phi(x)\frac{d\mu}{dm}(x)-U^*(\phi(x))\right),$$

as $U(\rho)$ is continuous (defining $U(0)=0$) and convex; i.e. $U^{**}=U$. Integrating this with respect to m yields

$$S(\mu\mid m)=\int_X U\left(\frac{d\mu}{dm}\right)dm=\int_X\sup_{\phi\in C(X)}\left(\phi(x)\frac{d\mu}{dm}(x)-U^*(\phi(x))\right)dm(x).$$

We'd like to interchange the integral with the supremum on the right-hand side. If we can do this, then

$$\begin{aligned}S(\mu\mid m)&=\sup_{\phi\in C(X)}\int_X\left[\phi(x)\frac{d\mu}{dm}(x)-U^*(\phi(x))\right]dm(x)\\&=\sup_{\phi\in C(X)}\left[\int_X\phi(x)\frac{d\mu}{dm}(x)\ dm(x)-\int_X U^*(\phi(x))\ dm(x)\right]\\&=\sup_{\phi\in C(X)}\left[\int_X\phi(x)\ d\mu(x)-\int_X U^*(\phi(x))\ dm(x)\right]\end{aligned}$$

as desired. All the hard work is in showing we can commute the supremum and integral. ☐

What is this good for? We can use it to show the lower semi-continuity and contractivity of the entropy.

Corollary 5.10 (c.f. [Vil09] Theorem 29.20 ii)). *For (X,d_X) a compact metric space and $\mu,m\in\mathscr{M}_+(X)$ Radon, $S(\mu\mid m)$ is jointly weakly lower semi-continuous in (μ,m) and moreover*

$$S(f_\#\mu\mid f_\#m)\le S(\mu\mid m)$$

for any measurable $f:X\to Y$ with (Y,d_Y) compact.

Proof. Writing

$$S(\mu\mid m)=\sup_{\phi\in C(X)}\left[\int_X\phi(x)\ d\mu(x)-\int_X U^*(\phi(x))\ dm(x)\right].$$

[14]Sometimes this is called the Renyi entropy.

We could've taken the sup over $\phi \in L^\infty(X)$, but using the above we can exploit the fact that ϕ is continuous. If $\mu_k \overset{\star}{\rightharpoonup} \mu$, then by definition of weak-$*$ convergence we have

$$\int_X \phi(x)\ d\mu_k(x) \to \int_X \phi\ d\mu(x)$$

as X is compact. This shows that in fact the mapping $\mu \in C(X) \mapsto \int_X \phi\ d\mu$ is weak-$*$ continuous. Similarly, $U^*(s) = e^{s-1}$ is continuous so that $U^* \circ \phi \in C(X)$. Thus we have a supremum of weakly (μ, m) continuous functions, which is lower semi-continuous.

To show contractivity, we use the same representation:

$$\begin{aligned} S(f_\# \mu \mid f_\# m) &= \sup_{\varphi \in L^\infty(Y)} \left[\int_Y \varphi(y)\ d(f_\#\mu)(y) - \int_Y U^*(\varphi(y))\ d(f_\# m)(y) \right] \\ &= \sup_{\varphi \in L^\infty(Y)} \left[\int_X \varphi(f(x))\ d\mu(x) - \int_X U^*(\varphi(f(x)))\ dm(x) \right] \\ &= \sup_{\phi \in L^\infty(X), \phi = \varphi \circ f} \left[\int_X \phi(x)\ d\mu(x) - \int_X U^*(\phi(x))\ dm(x) \right] \\ &\leq \sup_{\phi \in L^\infty(X)} \left[\int_X \phi(x)\ d\mu(x) - \int_X U^*(\phi(x))\ dm(x) \right] = S(\mu \mid m) \end{aligned}$$

□

5.5. **Measured Gromov-Hausdorff and stability for compact $CD^e(K,N)$ spaces.** We're almost ready to show our compactness theorem. As we saw before, we only had pre-compactness since we did not impose a measure on the limit space. Thus there was no way to test if it was in $CD^e(K,N)$. We develop now another form of Gromov-Hausdorff convergence to take care of this.

Definition 5.11. For a sequence of compact spaces $\{(X_k, d_k)\}_{k=1}^\infty$ and compact limit (X,d) we say that (X_k, d_k, m_k) converges to (X,d,m) in the *measured Gromov-Hausdorff sense*, denoted $(X_k, d_k, m_k) \xrightarrow{mGH} (X,d,m)$ if there exist ϵ_k-isometries $f^k : X_k \to X$ with $\epsilon_k \to 0$ and $f^k_\# m_k \overset{\star}{\rightharpoonup} m$, i.e.

$$\lim_{k\to\infty} \int_{X_k} \phi \circ f^k\ dm_k = \int_X \phi\ dm$$

for all $\phi \in C(X)$.

We can finally prove one of our main theorems:

Theorem 5.12 (Stability for compact $CD^e(K,N)$ spaces)**.** *If $(X_j, d_j, m_j) \in CD^e(K,N)$, with $K \in \mathbb{R}$ and $N \in [1,\infty]$, are all compact and $(X_j, d_j, m_j) \xrightarrow{mGH} (X,d,m)$ (also compact) then $(X,d,m) \in CD^e(K,N)$ if $\operatorname{spt} m = X$*

Proof. Fix $\mu_0, \mu_1 \in \mathcal{P}(X)$. Assume $\rho_i := d\mu_i/dm \in C(X)$ for $i = 0,1$. Because we have measured Gromov-Hausdorff convergence, we can find approximate ϵ_j-isometries $f^j : X_j \to X$ with $\epsilon_j \to 0$ such that $f^j_\# m_j \overset{\star}{\rightharpoonup} m$. We use the f_j to define approximations of the endpoints μ_i in X_j. Set $\rho_i^j = \rho_i \circ f^j / Z_i^j$ with

$$Z_i^j = \int_{X_j} \rho_i \circ f^j\ dm_j = \int_X \rho_i d(f^j_\# m_j)$$

a normalization factor. Because $f^j_\# m_j \overset{\star}{\rightharpoonup} m$, we have that $Z_i^j \to 1$ for $i = 0,1$. Hence for j large enough, $Z_i^j > 0$ and the ρ_i^j induce absolutely continuous probability measures on X_j. We define μ_i^j by $d\mu_i^j = \rho_i^j dm_j$. Now since $(X_j, d_j, m_j) \in CD^e(K,N)$ there exist approximating geodesics $(\mu_t^j)_{t\in[0,1]} \in \mathrm{Geo}(\mathcal{P}(X_j))$. Hence by Corollary 5.5, the curves $f^j_\# \mu_t^j$ converge uniformly (along some subsequence) to $(\mu_t)_{t\in[0,1]} \in \mathrm{Geo}(\mathcal{P}(X))$. By the $CD^e(K,N)$ condition, we know that $e_j(t) := S(\mu_t^j \mid m_j)$ is $(K d_2^{X_j}(\mu_0^j, \mu_1^j)^2, N)$-convex.

i) $N=\infty$: We want to show that

$$e(t) \le (1-t)e(0) + te(1) - \frac{Kt(1-t)}{2} d_2^X(\mu_0,\mu_1)^2.$$

Since $e_j(t)$ is $K d_2^{X_j}(\mu_0^j,\mu_1^j)^2$-convex, we have the same inequality in the approximating spaces:

$$e_j(t) \le (1-t)e_j(0) + te_j(1) - \frac{Kt(1-t)}{2} d_2^{X_j}(\mu_0^j,\mu_1^j)^2.$$

We hope to take a limit in this. In order to do so, there are three things we must verify.

a) For $i=0,1$ we have $\lim_{j\to\infty} e_j(i) = e(i)$;
b) $e(t) \le \lim_{j\to\infty} e_j(t)$;
c) $\lim_{j\to\infty} d_2^{X_j}(\mu_0^j,\mu_1^j) = d_2^X(\mu_0,\mu_1)$.

The last claim follows immediately from Corollary 5.5, so we just need to check the first two.

a) By definition,

$$\begin{aligned} e_j(i) = S(\mu_i^j \mid m_j) &= \int_{X_j} \frac{d\mu_i^j}{dm_j} \log\left(\frac{d\mu_i^j}{dm_j}\right) dm_j = \int_{X_j} \rho_i^j \log(\rho_i^j)\ dm_j \\ &= \frac{1}{Z_i^j}\int_{X_j} (\rho_i\circ f^j)\log(\rho_i\circ f^j)\ dm_j - \frac{\log(Z_i^j)}{Z_i^j}\int_{X_j} \rho_i\circ f^j\ dm_j \\ &= \frac{1}{Z_i^j}\int_X \rho_i\log(\rho_i)\ d(f_\#^j m_j) - \frac{\log(Z_i^j)}{Z_i^j}\int_X \rho_i\ d(f_\#^j m_j). \end{aligned}$$

Now we know that $Z_i^j \to 1$ and $f_\#^j m_j \overset{\star}{\rightharpoonup} m$. By assumption $\rho_i \in C(X)$ so that $\rho_i\log(\rho_i)\in C(X)$. Then by definition of weak-$*$ convergence we get that $e_j(i)\to e(i)$.

b) Contractivity tells us that

$$S(f_\#^j\mu_t^j \mid f_\#^j m_j) \le S(\mu_t^j \mid m_j) = e_j(t).$$

Combined with the fact that $S(\mu \mid m)$ is jointly weakly lower semi-continuous, since $f_\#^j\mu_t^j \overset{\star}{\rightharpoonup} \mu_t$ (d_2^X metrizes weak-$*$ convergence) and $f_\#^j m_j \overset{\star}{\rightharpoonup} m$ we get

$$S(\mu_t \mid m) \le \liminf_{j\to\infty} S(f_\#^j\mu_t^j \mid f_\#^j m_j) \le \lim_{j\to\infty} e_j(t)$$

as desired.

It remains to show we can drop the continuity assumption on the ρ_i. To do this, we show any $\mu_0,\mu_1\in\mathcal{P}(X)$ of finite entropy admit approximations $d_2^X(\mu_i^k,\mu_i)\to 0$ such that $S(\mu_i \mid m) \ge \limsup_{k\to\infty} S(\mu_i^k \mid m)$ where the μ_0^k have continuous densities.[15] By the previous step, we'll have shown that

$$S(\mu_t^k \mid m) \le (1-t)S(\mu_0^k \mid m) + tS(\mu_1^k \mid m) - \frac{Kt(1-t)}{2} d_2^X(\mu_0^k,\mu_1^k)^2.$$

Since the μ_t^k are all geodesics, the condition $d_2^X(\mu_i^k,\mu_i)\to 0$ at the endpoints is enough to show that it holds along the entire geodesic. That is, $\mu_t^k \overset{\star}{\rightharpoonup} \mu_t$ for all $t\in[0,1]$. By the triangle inequality we have

$$\begin{aligned} d_2^X(\mu_0^k,\mu_1^k) &\le d_2^X(\mu_0^k,\mu_0) + d_2^X(\mu_1^k,\mu_1) + d_2^X(\mu_0,\mu_1) \\ &\le 2d_2^X(\mu_0^k,\mu_0) + 2d_2^X(\mu_1^k,\mu_1) + d_2^X(\mu_0^k,\mu_1^k) \end{aligned}$$

so that taking the limit as $k\to\infty$ on both sides yields equality throughout. That is, $\lim_{k\to\infty} d_2^X(\mu_0^k,\mu_1^k) = d_2^X(\mu_0,\mu_1)$. Then by lower semi-continuity of the entropy and our

[15] This kind of approximation should remind the reader of Γ-convergence. More about this can be found in A.4.

limsup assumption, we have

$$\begin{aligned} S(\mu_t \mid m) &\leq \liminf_{k\to\infty} S(\mu_t^k \mid m) \\ &\leq (1-t)\limsup_{k\to\infty} S(\mu_0^k \mid m) + t\limsup_{k\to\infty} S(\mu_1^k \mid m) \\ &\quad - \frac{Kt(1-t)}{2}\lim_{k\to\infty} d_2^X(\mu_0^k,\mu_1^k)^2 \\ &\leq (1-t)S(\mu_0 \mid m) + tS(\mu_1 \mid m) - \frac{Kt(1-t)}{2} d_2^X(\mu_0,\mu_1)^2. \end{aligned}$$

We defer to proof of finding the desired approximations μ_i^k until after the proof.

ii) $N \in [1,\infty)$: Recall $u_j(t) = U_N(\mu_t^j) = \exp\Big(-1/NS(\mu_t^j \mid m_j)\Big)$. Then we have the following concavity inequality

$$u_j(t) \geq \sigma_{K/N}^{(1-t)}(d_2^{X_j}(\mu_0^j,\mu_1^j))u_j(0) + \sigma_{K/N}^{(t)}(d_2^{X_j}(\mu_0^j,\mu_1^j))u_j(1).$$

So as before we just need to take a limit. Item a) in our previous claim carries through since the exponential is continuous, so $u_j(i) \to u(i)$ for $i = 0,1$. Item c) also still holds, and since the $\sigma_{K/N}^{(t)}$ are continuous we have that $\sigma_{K/N}^{(t)}(d_2^{X_j}(\mu_0^j,\mu_1^j)) \to \sigma_{K/N}^{(t)}(d_2^X(\mu_0,\mu_1))$. Due to the minus sign in the exponential, any inequality for the entropy gets reversed at the level of U_N. So item b) tells us that $u(t) \geq \lim_{j\to\infty} u_j(t)$. So, taking limits we get exactly what we want:

$$u(t) \geq \sigma_{K/N}^{(1-t)}(d_2^X(\mu_0,\mu_1))u(0) + \sigma_{K/N}^{(t)}(d_2^X(\mu_0,\mu_1))u(1).$$

Removing the continuity assumption follows as before.

□

Remark 5.13. The weak $CD^e(K,N)$ condition tells us there exists a midpoint μ_t^k satisfying the convexity type inequality, but we have much less control on how it approximates the μ_t coming from our Arzela-Ascoli theorem. We're relieved by the fact that we only need lower semi-continuity at the midpoints (because we have less control).

Similarly, we assumed continuous densities. But really it's the same thing. We construct continuous approximations where convergence is nice at the endpoints. Then, after taking a limit we get a geodesic which we have little control over in the midpoints, but we only need lower semi-continuity.

5.6. **Regularizing kernels and approximating measures with continuous densities.** So how do we find the desired approximations μ_i^k? If we were in Euclidean space, the idea is to mollify it. Because $\rho\log\rho$ is convex, by Jensen's inequality when you mollify you lower the entropy. We need some kind of mollification that works on a compact metric space.

Definition 5.14 (c.f. [Vil09] Definition 29.34)**.** Let (X,d,m) be boundedly compact with m Radon. For $Y \subset X$ compact a (Y,m)-*regularizing kernel* is a family $\{K_r\}_{r>0} \subset C(X\times X)$ such that

i) For $x,y \in X$ we have $K_r(x,y) = K_r(y,x) \geq 0$ (symmetry);
ii) If for $x,y \in X$ we have $d(x,y) > r$ then $K_r(x,y) = 0$ (local);
iii) $\int_X K_r(x,y)\ dm(y) = 1$ for every $x \in Y$ (normalization).

The associated linear operator on $\mu \in \mathscr{M}(X)$ (and on $f \in L^1(X,m)$) are

$$(K_r\mu)(x) := \int_X K_r(x,y)\ d\mu(y)$$
$$(K_rf)(x) := \int_X K_r(x,y)f(y)\ dm(y).$$

Lemma 5.15. *Let (X,d,m) be boundedly compact and $Y \subset X$ compact. Then,*

(1) There exists a (Y,m)-regularizing kernel.
(2) The operator $K_r : \mathscr{M}(X) \to C(X)$ satisfies
  a) If $f \in C(X)$ then $\lim_{r\to 0} \|f - K_rf\|_{L^\infty(Y,m)} = 0$;

b) If $\mu \in \mathscr{M}(Y)$ *then* $\|(K_r\mu)_m\|_{TV(Y)} \le \|\mu\|_{TV(Y)}$ *with equality if* $\mu \ge 0$*; Also,* $d_2^X(\mu, (K_r\mu)m) \to 0$. *Do we need* μ *finite?*

c) If $f \in L^1(Y,m)$ *then* $\|f - K_r f\|_{L^1(Y,m)} \to 0$.

These properties motivate the definition of a regularizing kernel.

Proof. The proof is straightforward.

(1) Given $r > 0$ cover Y by finitely many balls $\{B_{r/2}(x_i)\}_{i=1}^N$ and let $\{\phi_i\}_{i=1}^N \subset C(X,[0,1])$ be a subordinate partition of unity: i.e. $\sum_{i=1}^N \phi_i(x) = 1$ on Y and $\phi_i(x) = 0$ unless $x \in B_{r/2}(x_i)$. Now we define K_r by
$$K_r(x,y) = \sum_{i=1}^N \frac{\phi_i(x)\phi_i(y)}{\int_X \phi_i \, dm}.$$
Properties i) and iii) are immediate. To show ii), note that if $d(x,y) > r$ it cannot be that both x and y lie in $B_{r/2}(x_i)$ for some $i \in 1,...,N$. Thus $\phi_i(x)\phi_i(y) = 0$ for all y.

(2) We don't need an explicit choice of regularizing kernel, just the properties in the definition.

a) Given $\epsilon > 0$ and $f \in C(X)$, there exists an $r > 0$ small enough such that $d(x,y) < r$ implies $|f(x) - f(y)| < \epsilon$ (i.e. continuous functions are always uniformly continuous on compact sets). Then,
$$\begin{aligned}|f(x) - (K_r f)(x)| &= \left| \int_X f(x)K_r(x,y)\, dm(y) - \int_X K_r(x,y)f(y)\, dm(y)\right| \\ &\le \int_X |f(x) - f(y)|K_r(x,y)\, dm(y)\end{aligned}$$
Now since $K_r(x,y) = 0$ if $d(x,y) > r$, we need only consider when $d(x,y) < r$. But in this regime $|f(x) - f(y)| < \epsilon$ so that
$$|f(x) - (K_r f)(x)| < \epsilon \int_X K_r(x,y)\, dm(y) = \epsilon.$$
We emphasize here that uniform convergence is *only* on Y, as we need $x \in Y$ in order for $\int_X K_r(x,y)\, dm(y) = 1$.

b) We want to show $K_r\mu \overset{\star}{\rightharpoonup} \mu$ as $r \to 0$. A simple way of doing this is to look at the action on continuous functions. Recall this is given by $\mu(f) = \int_X f\, d\mu$. Define $\mu_r = (K_r\mu)m$. We show that $\mu_r \overset{\star}{\rightharpoonup} \mu$ by looking at the action on continuous functions. For $f \in C(X)$ we have
$$\begin{aligned}[(K_r\mu)m](f) &= \int_X f(x) d[(K_r\mu)m](x) = \int_X f(x) K_r\mu(x)\, dm(x) \\ &= \int_X f(x) \left[\int_Y K_r(x,y)\, d\mu(y)\right] dm(x) \\ &= \int_Y \left[\int_X f(x)K_r(x,y)\, dm(x)\right] d\mu(y) = \int_Y K_r f(y)\, d\mu(y).\end{aligned}$$
where we use the fact that
$$\begin{aligned}(K_r f)(z) &= \int_X K_r(z,y) f(y)\, dm(y) \\ &= \int_X K_r(y,z)f(y)\, dm(y) = \int_X K_r(x,z) f(x)\, dm(x)\end{aligned}$$
owing to the symmetry of $K_r(\cdot,\cdot)$. Now, since $K_r f$ converges uniformly on Y, for $\epsilon > 0$ there exists $r > 0$ such that $\|K_r f - f\|_{L^\infty(Y,m)}$. Then,
$$\begin{aligned}|\mu(f) - [(K_r\mu)m](f)| &= \left|\int_Y [f(y) - K_r f(y)]\, d\mu(y)\right| \\ &\le \int_Y \|K_r f - f\|_{L^\infty(Y,m)}\, d\mu(y) < \epsilon\mu(Y).\end{aligned}$$

Now if $\mu \geq 0$ we want to show $\|(K_r\mu)m\|_{TV} = \|\mu\|_{TV}$. Then by Fubini,

$$(K_r\mu)m(Y) = \int_Y K_r\mu(x)\ dm(x) = \int_Y 1\ d\mu(y) = \mu(Y).$$

If μ is signed, then the total variation can only decrease due to cancellations.

c) Let $f \in L^1(Y,m)$. We want to show that $\|K_r f - f\|_{L^1(Y,m)} \to 0$ as $r \to 0$. By a) and the fact that Y is compact, we get $L^1(Y,m)$ convergence for continuous functions on Y. But $C(Y)$ is dense in $L^1(Y,m)$ so that we get $L^1(Y,m)$ convergence on all of $L^1(Y,m)$.

□

We're now ready to find our approximating sequence.

Proposition 5.16 (c.f. [Vil09] Theorem 29.20 iii))**.** *Fix (X,d) compact with $m \in \mathscr{M}_+(X)$ a Radon measure with $\operatorname{spt} m = X$. Any $\mu \in \mathcal{P}(X)$ can be weak-$*$ approximated by $\mu_k \in \mathcal{P}^{ac}(X)$ having $d\mu_k/dm \in C(X)$ and $S(\mu \mid m) \geq \limsup_{k\to\infty} S(\mu_k \mid m)$.*

Proof. There exists an (X,m)-regularizing kernel K_r which we use to mollify μ. Set

$$\rho_r(x) := \int_X K_r(x,y)\ d\mu(y)$$

and define μ_r by $d\mu_r(x) = \rho_r(x)dm(x)$. By Lemma 5.15, part 2b) tells us that $\mu_r \overset{\star}{\rightharpoonup} \mu$. If $S(\mu \mid m) = \infty$ then we're done, so we might as well assume it is finite. Then, $\mu \in \mathcal{P}^{ac}(X)$. Let $\rho(x) = d\mu/dm(x)$, which need not be continuous. We can rewrite ρ_r as

$$\rho_r(x) = \int_X K_r(x,y)\rho(y)\ dm(y).$$

Letting $U(\rho) = \rho\log\rho$, Jensen's inequality tells us that

$$U(\rho_r(x)) = U\left(\int_X K_r(x,y)\rho(y)\ dm(y)\right) \leq \int_X K_r(x,y)U(\rho(y))\ dm(y)$$

since $\int_X K_r(x,y)\ dm(y) = 1$ for $x \in X$. Integrating the above with respect to m and applying Fubini gives

$$\begin{aligned} S(\mu_r \mid m) &\leq \int_X \left[\int_X K_r(x,y)U(\rho(y))\ dm(y)\right] dm(x) \\ &= \int_X U(\rho(y)) \left[\int_X K_r(x,y)\ dm(x)\right] dm(y) = S(\mu \mid m), \end{aligned}$$

and the inequality is preserved after taking the limsup. □

5.7. **Stability for boundedly compact $CD^e(K,N)$ spaces.** With this proposition we have fully proven the compactness result for *compact* $CD^e(K,N)$ spaces. What about the non-compact case? As with the Gromov-Hausdorff pre-compactness theorem, the best we can hope for is to relax ourselves to the boundedly compact case.

Theorem 5.17 (Stability for boundedly compact $CD^e(K,N)$ spaces)**.** *If $(X_j,d_j,m_j,p_j) \in CD^e(K,N)$ is a boundedly compact sequence with pointed measured Gromov-Hausdorff limit (X,d,m,p) then $(X,d,m) \in CD^e(K,N)$ if $X = \operatorname{spt} m$.*

Remark 5.18. Villani defines $(X,d,m) \in CD(K,N)$ if for every $\mu_0,\mu_1 \in \mathcal{P}(X)$ of compact support there exists a d_2-geodesic $(\mu_t)_{t\in[0,1]}$ along which the entropy is $(Kd_2(\mu_0,\mu_1)^2,N)$-convex. In the pointed Gromov-Hausdorff setting we have a designated point $p \in X$, and because $\operatorname{spt}\mu_0, \operatorname{spt}\mu_1$ are compact there exists $R > 0$ such that $B_R(p)$ contains both supports. Although $\operatorname{spt}\mu_t$ might escape this ball for $t \in (0,1)$, it certainly will always be contained in $B_{2R+1}(p)$ (as traveling to and from p will be shorter than traveling outside $B_{2R+1}(p)$). So to extend these entropic convexity properties to geodesics for which the endpoints are compactly supported we can use the same proof as before.

Robert also remarks here that Villani seems able to drop the condition $X = \operatorname{spt} m$ by appealing to the Tietze extension theorem. His rough explanation is as follows: We had this mollification procedure where we took $\mu \in \mathcal{P}(X)$ and mollified it to get a sequence $\mu_r \in \mathcal{P}^{ac}(X)$ with continuous densities such that $\mu_r \xrightarrow{\star} \mu$. If we do this mollification within $\operatorname{spt}\mu$ then we get $\operatorname{spt}\mu_r \subset \operatorname{spt}\mu$. But we don't want the densities to drop discontinuously to zero as we cross the boundary of the support. So maybe what we could do is use the Tietze extension theorem so that the densities drop quickly to zero. You won't have a probability measure anymore so you have to divide by a constant.

Remark 5.19. There's a more recent approach which allows us to extend this result to Polish spaces which are not necessarily boundedly compact. This is really only helpful in the $N = \infty$ case, as for $N < \infty$ every $CD^e(K,N)$ space is boundedly compact. This comes from a paper by Gigli-Mondino-Savare ([GMS15]). They invent a slightly weaker notion of convergence which turns out to be equivalent in the $CD(K,N)$ setting with $N < \infty$. This convergence, called pointed measured Gromov convergence, is defined in the following way: we say that $(X_1, d_1, m_1, p_1) \simeq (X_2, d_2, m_2, p_2)$ if and only if there exists a Polish space (X,d) and isometric embeddings $i_k : (X_k, d_k) \to (X,d)$ for $k = 1,2$, $i_1(p_1) = i_2(p_2)$, and $(i_1)_\# m_1 = (i_2)_\# m_2$. This implies that $i_1(\operatorname{spt} m_1) = i_2(\operatorname{spt} m_2)$, and you essentially don't care about the rest of the space.[16] Then (X_j, d_j, m_j, p_j) converges in the pointed measured Gromov sense to (X,d,m,p) if all of them embed into the same metric space, $i_j(p_j) = i(p)$, and $(i_j)_\# m_j \to i_\# m$ against continuous test functions of bounded support. This implies stability of $CD^e(K,\infty)$ without the assumption of boundedly compact. This feels like a more natural assumption, and it's a little simpler (since we don't ask for Gromov-Hausdorff convergence of the spaces).

6. Additional Properties of $CD^e(K,N)$ spaces.

6.1. **Building new $CD^e(K,N)$ spaces from old ones.** There are two ways that we can generate $CD^e(K,N)$ spaces. Recall in Theorem 2.20 saw that $(\mathbb{R}^n, |\cdot|, m) \in CD^e(K,N)$ for $dm(x) = \phi(x)dx$ and ϕ non-constant if $N > n$ and

$$D^2\phi(x) - \frac{(D\phi \otimes D\phi)(x)}{N-n} - K\,\mathrm{Id} \geq 0$$

for all $x \in \mathbb{R}^n$. For example, when we have a Gaussian $\phi(x) = c|x|^2/2$ then $D^2\phi(x) = c\,\mathrm{Id}$ and $(D\phi \otimes D\phi)(x) = c^2(x \otimes x)$. Thus we require for any $y \in \mathbb{R}^n$ that

$$c|y|^2 - \frac{c^2\langle y, x\rangle^2}{N-n} \geq K|y|^2.$$

If $N = \infty$, this amounts to having $c \geq K$. However for $n < N < \infty$, the term $\langle y, x\rangle$ always dominates for $|x|$ large enough, and so it is never in $CD^e(K,N)$ for finite N.

Another thing we can do is change the norm. Consider for example $(\mathbb{R}^n, |\cdot|_p, dx)$. If the unit ball under a given norm is convex (as is the case in a Banach space) then straight lines are always geodesics. Moreover for $1 < p < \infty$ these are the only geodesics. For $p = \infty$, there are many geodesics. Connecting any two points with a straight line, we can always replace this by paths parallel to the axes. In particular, the space is branching. Since the geodesics for $|\cdot|_p$ (again $1 < p < \infty$) are the same as $|\cdot|_2$, we have that $(\mathbb{R}^n, |\cdot|_p, dx) \in CD^e(K,N)$ if and only if $K \leq 0$ and $N \geq n$. On the other hand, $(\mathbb{R}^n, |\cdot|_\infty, dx) \in CD^e(K,N)$ for $K \leq 0$ and $N \geq n$, but it is not a strong $CD^e(K,N)$ space. The manifolds $(\mathbb{R}^n, |\cdot|_p)$ give an example of Finsler manifolds, as the limit as $p \to 1$ or $p \to \infty$ gives a unit ball which is not an ellipse (instead, a diamond or square respectively).

See also A.3 for a homework problem which lets us obtain new $CD^e(K,N)$ spaces by changing the reference measure in a fairly general way.

6.2. **Local to global property of $CD^e(K,N)$ spaces.** We discussed previously that $CD^e(K,N)$ is equivalent to the original $CD(K,N)$ condition under a non-branching hypothesis.[17] Both kinds of spaces enjoy a local to global property, but it is much easier to prove with the entropic definition.

[16] Recalling our previous compactness theorems, this should remind us of the assumption that $X = \operatorname{spt} m$.

[17] This is all we know as of writing these notes, it could be true in greater generality.

Theorem 6.1 (Local to global property)**.** *Let (X,d,m) be a Polish and geodesic space. If $(B_r(x),d,m) \in CD^e(K,N)$ for all $x \in \operatorname{spt} m$ and $r > 0$ sufficiently small, then $(X,d,m) \in CD^e(K,N)$.*

Proof. The proof will be mostly impressionistic. Given $\mu_0,\mu_1 \in \mathcal{P}(X)$ with compact support there exists a d_2-geodesic $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}_2(X)$. Since the endpoints are compactly supported, the velocity cannot be too large (it is bounded by something like the greatest distance between two points in the supports). By covering both supports with balls of radius $r/2$ (where $(B_r(x),d,m) \in CD^e(K,N)$), we are essentially reduced to looking at what happens on small balls.

With this reduction in place, the support doesn't move too far. It moves in a ball of radius r at each step (I imagine here we mean to discretize the motion by partitioning $[0,1]$ and look at the supports at each finite time). By decomposing $\mu_t = \sum_i \mu_t^i$, where all the μ_t^i are mutually singular, along these balls we get $S(\mu_t \mid m) = \sum_i S(\mu_t^i \mid m)$. Now we want to use the fact that the entropic curvature dimension holds along these small balls to conclude it holds along the entire geodesic. Recall if $e_i(t) = S(\mu_t^i \,|m)$ we want $e_i''(t) - e_i'(t)^2/N \geq K d_2(\mu_0,\mu_1)^2$. The left-hand side seems local in t, which is good, but the right-hand side seems global. However, this is merely apparent, and is just an artifact due to insisting on parametrizing geodesics over $[0,1]$. E.g., by making the change of variables $s = t d_2(\mu_0,\mu_1)^2$ we get

$$e_i''(s) - \frac{1}{N} e_i'(s)^2 \geq K$$

which is now local in s (parametrized by arc-length). By covering the whole geodesic by balls (of radius r), since we have the convexity condition on each ball we have it on the whole geodesic. $\square$

6.3. **Tangent cones and $RCD(K,N)$ spaces.** We saw that $CD^e(K,N)$ is Gromov-Hausdorff compact and measured Gromov-Hausdorff compact. In particular if you take $(X,d,m) \in CD^e(K,N)$ and $p \in X$, then $B_r(p) \in CD(K,N)$.[18] Considering $(B_r(p), d/r, m)$, we see that geodesics remain unchanged (except their arc length might dilate). There is no sense in modifying the measure because

$$S(\mu \mid cm) = \int_X \log\left(\frac{1}{c}\frac{d\mu}{dm}\right) d\mu = S(\mu \mid m) - \log(c)$$

for $c > 0$. The concavity properties of the functional are unaffected by adding a constant. Then we get something like $(B_r(p), d/r, m) \in CD^e(r^2K, N)$. Taking a subsequence as $r \to 0$ we get a limit in $CD^e(0,N)$. If the limit exists (without passing to a subsequence) it is called the Gromov-Hausdorff tangent cone.

If what you started with was a Riemannian manifold, i.e. $(X,d,m) = (M^n, d_g, \mathrm{vol}_g)$ then the limiting space above is the tangent space, which is Euclidean. This leads us to the following definition:

Definition 6.2 ($RCD(K,N)$ space)**.** The *Riemannian curvature dimension space* is defined as

$$\begin{aligned} RCD(K,N) = \{X \in CD(K,N) \mid &\text{ for } m\text{-a.e. } p \in X, \text{ the Gromov-Hausdorff} \\ &\text{tangent cone at } p \text{ exists and is Euclidean.}\} \end{aligned}$$

This was not the original definition and was defined by Ambrosio-Gigli-Savare in [AGS14]. They showed that

i) $RCD(K,N)$ is closed under measured Gromov-Hausdorff limits and;
ii) $X \in RCD(K,N)$ implies X is strongly $CD(K,N)$.

Later, Rajala-Sturm ([RS14]) showed X in strong $CD(K,N)$ implies X is essentially non-branching.

Definition 6.3. (X,d,m) is *essentially non-branching* (ENBR) if for $\mu_0,\mu_1 \in \mathcal{P}_2(X)$ and d_2-geodesic $(\mu_t)_{t\in[0,1]}$ coming from $\pi \in \mathcal{P}(\mathrm{Geo}(X,d))$ there exists a set $\Omega \subset \mathrm{Geo}(X,d)$ with $\pi(\Omega) = 1$ and no two geodesics $\gamma, \tilde\gamma \in \Omega$ branch.

[18] There is a small caveat here, where you need all geodesics connecting two points in $B_r(p)$ to lie in $B_r(p)$.

Although branching is not a stable property (recall the spaces $(\mathbb{R}^n, |\cdot|_p)$), and only weak $CD^e(K,N)$ is a stable property, it turns out that $RCD(K,N)$ is a stable property and it implies essentially non-branching. So you might have a few geodesics which branch in the space, but you never need to use them to generate optimal geodesics.

Appendix A. Selected Homework Problems.

Listed are various relevant homework problems assigned during the course, with (mostly correct) solutions.

A.1. **Problem 1 [Riemannian geodesics and distance].**

a) Let $g_{ij}(x)$ be a smooth map from $\mathbb{R}^n$ into symmetric positive definite $n \times n$ matrices, and fix $x_0, x_1 \in \mathbb{R}^n$. If $x : [0,1] \to \mathbb{R}^n$ minimizes the energy functional
$$E[x] := \frac{1}{2}\int_0^1 g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)\; dt$$
among all smooth curves starting at $x(0) = x_0$ and ending at $x(1) = x_1$, use calculus to show that $x(t)$ satisfies the ordinary differential equation
$$\ddot{x}^i(t) + \Gamma^i_{jk}(x(t))\dot{x}^j(t)\dot{x}^k(t) = 0$$
where
$$\Gamma^i_{jk} = \frac{1}{2}g^{im}\left(\frac{\partial g_{mk}}{\partial x^j} + \frac{\partial g_{mj}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^m}\right)$$
are the so-called Christoffel symbols. (Repeated indices are summed from 1 to n, and g^{im} denotes the matrix inverse of g_{mi})

b) In that case, show (using Jensen's inequality, for example) that $x(t)$ also minimizes the length functional
$$L[x] := \int_0^1 (g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t))^{1/2}\; dt$$
in the same class of curves. Its value at the miminum defines $d(x_0, x_1) := L[x]$.

c) Using a), show by induction that if $x : [0,1] \to \mathbb{R}^n$ minimizes $E[x]$ in the larger class of continuous, piecewise smooth curves, then $x \in C^k$ for all $k \in \mathbb{N}$.

d) Show $d(x_0, x_1) \leq d(x_0, y) + d(y, x_1)$ for all $y \in \mathbb{R}^n$ and $d(x(s), x(t)) = |s-t| d(x_0, x_1)$ for all $s, t \in [0,1]$.

Solution:

a) We compute the Euler-Lagrange equation of $E[x]$. To start, let us clarify something. Since g_{ij} is supposed to denote the components of a (Riemannian) metric tensor, I assume that we should have $g : \mathbb{R}^n \to \mathrm{Sym}_{n\times n}(\mathbb{R})$ so that each $g_{ij} : \mathbb{R}^n \to \mathbb{R}$ is a component of this matrix function. That is, $g_{ij}(x) = (g(x))_{ij}$. The Euler-Lagrange equations read
$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}^m} = \frac{\partial \mathcal{L}}{\partial x^m}$$
Where the Lagrangian is $\mathcal{L}(t, x, \dot{x}) = 1/2 g_{im}(x(t))\dot{x}^i(t)\dot{x}^m(t)$. Note: the indices used in defining the Lagrangian are dummy indices and can be changed at will.

Let us begin with the right-hand side. Treating x^m like a variable, we get
$$\frac{\partial \mathcal{L}}{\partial x^m} = \frac{1}{2}\frac{\partial g_{jk}}{\partial x^m}\dot{x}^j\dot{x}^k.$$
Note that we changed the indices in $\mathcal{L}$ from i and m to j and k at no consequence. We will use this trick one more time later on. The left-hand side is slightly more complicated. Treating $\dot{x}^m$ as a variable,
$$\frac{\partial \mathcal{L}}{\partial \dot{x}^m} = g_{im}(x(t))\dot{x}^i(t)$$
We do not differentiate the first two terms because, even though $\dot{x}^j$ depends on x and invariably affects $x(t)$ and $\dot{x}^i(t)$, in these types of computations we treat everything as

independent variables. Furthermore, we get a factor of 2 due to symmetry. Next, the time derivative is

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}^m} = \frac{d}{dt}(g_{im}(x(t))\dot{x}^i(t) + g_{im}(x(t))\ddot{x}^i(t) = \frac{\partial g_{im}}{\partial x^k}\dot{x}^i\dot{x}^k + g_{im}\ddot{x}^i.$$

Now, since the first term is summed over i and k, we can change these indices to whatever letter we like (so long as they aren't the same letter). We then split the sum into two terms as follows

$$\frac{\partial g_{im}}{\partial x^k}\dot{x}^i\dot{x}^k = \frac{1}{2}\frac{\partial g_{im}}{\partial x^k}\dot{x}^i\dot{x}^k + \frac{1}{2}\frac{\partial g_{im}}{\partial x^k}\dot{x}^i\dot{x}^k = \frac{1}{2}\frac{\partial g_{mk}}{\partial x^j}\dot{x}^j\dot{x}^k + \frac{1}{2}\frac{\partial g_{mj}}{\partial x^k}\dot{x}^j\dot{x}^k$$

(we have also used the symmetry of the g_{ij} to interchange lower indices). Putting everything into the Euler-Lagrange equations gives

$$\begin{aligned} \frac{1}{2}\frac{\partial g_{mk}}{\partial x^j}\dot{x}^j\dot{x}^k + \frac{1}{2}\frac{\partial g_{mj}}{\partial x^k}\dot{x}^j\dot{x}^k + g_{im}\ddot{x}^i &= \frac{1}{2}\frac{\partial g_{jk}}{\partial x^m}\dot{x}^j\dot{x}^k \\ g_{im}\ddot{x}^i + \frac{1}{2}\left(\frac{\partial g_{mk}}{\partial x^j} + \frac{\partial g_{mj}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^m}\right)\dot{x}^j\dot{x}^k &= 0 \\ \ddot{x}^i + \Gamma^i_{jk}\dot{x}^j\dot{x}^k &= 0 \end{aligned}$$

where in the last step we contracted with the inverse g^{im}.

b) We break this problem into several steps. The first provides a bound of the length in terms of the energy.

Step 1: Let $C(x_0, x_1)$ denote the class of smooth curves $x : [0,1] \to \mathbb{R}^n$ such that $x(0) = x_0$ and $x(1) = x_1$. Let $x \in C(x_0, x_1)$. Then, $1/2L[x]^2 \le E[x]$ with equality if and only if $\mathcal{L}(t, x, \dot{x})$ is constant.

Proof. We use Hölder's inequality, instead of Jensen's inequality as suggested. We have that

$$L[x] = \int_0^1 (2\mathcal{L}(t,x,\dot{x}))^{1/2}\, dt \le \left(\int_0^1 dt\right)^{1/2}\left(\int_0^1 2\mathcal{L}(t,x,\dot{x})\, dt\right)^{1/2} = \sqrt{2}E[x]^{1/2}.$$

It follows that $1/2L[x]^2 \le E[x]$, with equality if and only if $\mathcal{L}(t,x,\dot{x})$ is constant (i.e., the speed is constant). □

The second step gives us some information about minimizers of $E[x]$.

Step 2: Let $x \in C(x_0, x_1)$ be a minimizer of $E[x]$. Then x has constant speed, and in particular $1/2L[x]^2 = E[x]$.

Proof. Let $x \in C(x_0, x_1)$. Then,

$$\begin{aligned} \frac{d}{dt}\mathcal{L}(t,x,\dot{x}) &= \frac{1}{2}\frac{d}{dt}\left(g_{im}(x(t))\dot{x}^i(t)\dot{x}^m(t)\right) \\ &= \frac{1}{2}\left(\left[\frac{d}{dt}g_{jk}(x(t))\right]\dot{x}^j\dot{x}^k + 2g_{im}(x(t))\ddot{x}^i\dot{x}^m\right) \\ &= \frac{1}{2}\left(\frac{\partial g_{jk}}{\partial x^m}\dot{x}^j\dot{x}^k\dot{x}^m + 2g_{im}(x(t))\ddot{x}^i\dot{x}^m\right) \\ &= \frac{1}{2}\left(\frac{\partial g_{jk}}{\partial x^m}\dot{x}^j\dot{x}^k\dot{x}^m + \frac{\partial g_{jk}}{\partial x^m}\dot{x}^j\dot{x}^k\dot{x}^m - \frac{\partial g_{jk}}{\partial x^m}\dot{x}^j\dot{x}^k\dot{x}^m + 2g_{im}(x(t))\ddot{x}^i\dot{x}^m\right) \\ &= \frac{1}{2}\left(\frac{\partial g_{mk}}{\partial x^j}\dot{x}^j\dot{x}^k\dot{x}^m + \frac{\partial g_{mj}}{\partial x^k}\dot{x}^j\dot{x}^k\dot{x}^m - \frac{\partial g_{jk}}{\partial x^m}\dot{x}^j\dot{x}^k\dot{x}^m + 2g_{im}(x(t))\ddot{x}^i\dot{x}^m\right) \\ &= \left[\Gamma^i_{jk}\dot{x}^j\dot{x}^k + \ddot{x}^i\right]\dot{x}^m g_{im}(x(t)) \end{aligned}$$

where we used the same index-relabeling trick as before. So, minimizers of E have constant speed. □

To conclude we need access to a constant speed reparameterization. First, we show that the length functional $L[x]$ is independent of reparameterization.

Step 3: Let $x \in C(x_0, x_1)$ and $\alpha : [0,1] \to [0,1]$ a monotone, smooth map such that $\alpha(0) = 0$ and $\alpha(1) = 1$. Then $y(t) = x(\alpha(t)) \in C(x_0, x_1)$ and $L[y] = L[x]$

Proof. With notation as above,

$$\begin{aligned} L[y] &= \int_0^1 (g_{ij}(y(t))\dot{y}^i(t)\dot{y}^j(t))^{1/2}\, dt \\ &= \int_0^1 (g_{ij}(x(\alpha(t)))\alpha'(t)^2\dot{x}^i(\alpha(t))\dot{x}^j(\alpha(t)))^{1/2}\, dt \\ &= \int_0^1 \alpha'(t)(g_{ij}(x(\alpha(t)))\dot{x}^i(\alpha(t))\dot{x}^j(\alpha(t)))^{1/2}\, dt \\ &= \int_0^1 (g_{ij}(x(t)))\dot{x}^i(t)\dot{x}^j(t))^{1/2}\, dt = L[x] \end{aligned}$$

where we implicitly used the fact that $|\alpha'| = \alpha'$. □

There is a small technicality we must overcome. If $x \in C(x_0, x_1)$, it could be that $x(t)$ is constant for some sub-interval. But, since $x((0,1))$ is a smooth curve, there always exists a parameterization with $\dot{x}'(t) \neq 0$ everywhere. Roughly speaking, shrink these sub-intervals to a point. We now move to the final step.

Step 4: For each $x \in C(x_0, x_1)$ there exists a reparameterization x with constant speed.

Proof. First notice that $L[x] > 0$ and by the above remark we may assume $\dot{x}'(t) \neq 0$ everywhere. Now define

$$S(t) = \frac{1}{L[x]} \int_0^t [g_{ij}(x(\tau))\dot{x}^i(\tau)\dot{x}^j(\tau)]^{1/2}\, d\tau$$

so that

$$S'(t) = \frac{1}{L[x]}[g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)]^{1/2} > 0.$$

Non-negativity is guaranteed because

$$[g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)]^{1/2} = \langle \dot{x}(t), \dot{x}(t)\rangle_{x(t)}^{1/2} > 0$$

where $< \cdot, \cdot >_p$ denotes the Riemannian metric at a point p. That the Riemannian metric is positive definite implies the above, together with our assumption that $\dot{x}(t) \neq 0$.

We've then defined a strictly increasing function $S : [0,1] \to [0,1]$. Thus, there is a smooth inverse $T : [0,1] \to [0,1]$. Now consider the curve $\mathsf{x}(s) = x(T(s))$. By the chain rule, $\dot{\mathsf{x}}(s) = \dot{x}(T(s))T'(s)$. To evaluate $T'(s)$, we can use the fact it is an inverse to $S(t)$. That is,

$$T(S(t)) = t \;\Rightarrow\; T'(S(t))S'(t) = 1 \;\Rightarrow\; T'(S(t)) = \frac{L[x]}{[g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)]^{1/2}}.$$

Evaluating at $t = T(s)$ and rearranging yields

$$T'(s)[g_{ij}(x(T(s)))\dot{x}^i(T(s))\dot{x}^j(T(s))]^{1/2} = L[x].$$

Fix an $s \in [0,1]$. Then using the above, the speed of $\mathsf{x}(s)$ is computed as

$$\begin{aligned} [g_{ij}(\mathsf{x}(s))\dot{\mathsf{x}}^i(s)\dot{\mathsf{x}}^j(s)]^{1/2} &= [g_{ij}(x(T(s)))T'(s)^2\dot{x}^i(T(s))\dot{x}^j(T(s))]^{1/2} \\ &= T'(s)[g_{ij}(x(T(s)))\dot{x}^i(T(s))\dot{x}^j(T(s))]^{1/2} = L[x]. \end{aligned}$$

So, the speed is constant. □

We are now ready to answer the problem. The proof concludes as follows. Let $x \in C(x_0,x_1)$ be a minimizer of $E[x]$ and $y \in C(x_0,x_1)$ another competitor. Our goal is to show that $L[x] \leq L[y]$. First, since x has constant speed we have by $1/2L[x]^2 = E[x]$. Reparameterizing y to have constant speed, we obtain a curve y such that $L[\mathsf{y}] = L[y]$ and $1/2L[\mathsf{y}]^2 = E[\mathsf{y}]$. Finally, since y is a valid competitor and x is a minimizer of E, $E[x] \leq E[\mathsf{y}]$. Using all these,

$$\frac{1}{2}L[x]^2 = E[x] \leq E[\mathsf{y}] = \frac{1}{2}L[\mathsf{y}]^2 = \frac{1}{2}L[y]^2.$$

Remark: I think we can prove step 4 without the assumption that $\dot{x}(t) \neq 0$ everywhere. To do this, one defines $S(t)$ in the same way but instead defines $T(s)$ by

$$T(s) = \min\{t \in [0,1] \mid S(t) = s\}.$$

This is a left continuous, monotone increasing, *piecewise* smooth function. So, some care needs to be had, but I think everything works out fine. In particular, we probably need part c) in order to extend our class to piecewise smooth curves. This is necessary for the comparison step $E[x] \leq E[\mathsf{y}]$.

Remark: In fact, we can show that $x \in C(x_0,x_1)$ minimizes E if and only if x minimizes L and x has constant speed. The forward direction was just shown. For the latter, assume x minimizes L and x has constant speed. For a competitor $y \in C(x_0,x_1)$, we have $L[x] \leq L[y]$. Finally by step 1,

$$E[x] = \frac{1}{2}L[x]^2 \leq \frac{1}{2}L[y]^2 \leq E[y].$$

c) Let us first show the base case, i.e. for $x \in C_p(x_0,x_1)$ (continuous, piecewise smooth curves) we show that $x \in C^2$. We start by proving some regularity on $\dot{x}$.

Step 1: $\dot{x}^i \in L^2([0,1])$ for each $i = 1,...,n$.

Proof. To see this, since x is piecewise smooth there exists $K \in \mathbb{N}$ such that $\dot{x}^i$ is discontinuous at finitely many points $t_1,...,t_K$. Set $t_0 = 0$ and $t_{K+1} = 1$, we have that for each $k = 0,...,K$, x is smooth on $[t_k,t_{k+1}]$. Denote with a subscript k the restriction of a function to $[t_k,t_{k+1}]$. With this notation, we see that $\dot{x}^i_k$ is continuous for each k. Being continuous on a compact set, $\dot{x}^i_k$ attains a finite maximum and hence

$$\int_{t_k}^{t_{k+1}} \dot{x}^i_k(t)^2 \, dt \leq \|\dot{x}^i_k\|^2_{L^\infty([t_k,t_{k+1}])}|t_{k+1}-t_k| < \infty.$$

Finally,

$$\int_0^1 \dot{x}^i_k(t)^2 \, dt = \sum_{k=0}^{K} \int_{t_k}^{t_{k+1}} \dot{x}^i_k(t)^2 < \infty$$

owing to our finite partition. □

The geodesic equation provides a relationship between $\ddot{x}$ and essentially $\dot{x}^2$. If we can prove the geodesic equation holds at least almost everywhere, we can use the regularity of $\dot{x}$ to prove some regularity of $\ddot{x}$.

Step 2: $\ddot{x}$ satisfies the geodesic equation $\mathscr{L}^1$-a.e. in $[0,1]$ and $\ddot{x}^i \in L^1([0,1])$ for each $i = 1,...,n$.

Proof. Using the same notation as above, since x^i_k is smooth on (t_k,t_{k+1}) and x is a minimizer of $E[x]$ in $C(x_0,x_1)$, then in particular x_k is a minimizer of $E[x]$ in $C(x(t_k),x(t_{k+1}))$. Therefore, reparameterizing as necessary, by part a), we know that x_k satisfies the geodesic equation. Importantly since we have a finite partition, and x agrees with x_k on each $[t_k,t_{k+1}]$, we see that x satisfies the geodesic equation except at finitely many points.

Now, we know that $\dot{x}^i \in L^2([0,1])$ for each $i = 1,...,n$. Because each g_{ij} is smooth, the Christoffel symbols (being a smooth combination of partial derivatives and inverses of the g_{ij}) are also smooth. Since x is continuous and each Γ^i_{jk} is smooth, the composition

$\Gamma^i_{jk}(x(t))$ is continuous. Thus $\Gamma^i_{jk}(x(t))$ attains a maximum on the compact set $[0,1]$. Setting $G = \max\{\|\Gamma^i_{jk} \circ x\|_{L^\infty([0,1])}\}$ we see that $G < \infty$. Finally, $\ddot{x}^i$ agrees almost everywhere with $-\Gamma^i_{jk}(x(t))\dot{x}^j(t)\dot{x}^k(t)$, so that

$$\begin{aligned}
\int_0^1 |\ddot{x}^i(t)| \, dt &\leq \int_0^1 |\Gamma^i_{jk}(x(t))||\dot{x}^j(t)||\dot{x}^k(t)| \, dt \\
&\leq G \sum_{j,k=1}^n \int_0^1 |\dot{x}^j(t)||\dot{x}^k(t)| \, dt \\
&\leq G \sum_{j,k=1}^n \left(\int_0^1 \dot{x}^j(t)^2 \, dt \right)^{1/2} \left(\int_0^1 \dot{x}^k(t)^2 \, dt \right)^{1/2} < \infty
\end{aligned}$$

where in the last step we used Hölder's inequality. Thus $\ddot{x}^i \in L^2([0,1])$. □

Finally, we can use the fundamental theorem of calculus to transform this into a statement about continuity.

Step 3: Each $\dot{x}^i$ is absolutely continuous, and therefore $\ddot{x}^i \in C^1([0,1])$.

Proof. Since x is piecewise smooth, $\dot{x}^i$ has a derivative $\ddot{x}^i$ which exists $\mathscr{L}^1$ almost everywhere. By step 2, we know that $\ddot{x}^i$ is also integrable. Thus, by the fundamental theorem of calculus $\dot{x}^i$ is absolutely continuous everywhere. Now $\ddot{x}^i$ is continuous almost everywhere and $\ddot{x}^i = -\Gamma^i_{jk}(x(t))\dot{x}^j(t)\dot{x}^k(t)$ almost everywhere. But the right hand side is a smooth combination of continuous functions, and so $\ddot{x}^i$ is continuous. □

To complete the induction, assume x is C^n. Differentiating the geodesic equation $n-1$ many times expresses the components of the $n+1$ derivative in terms of the components of the $n, n-1, ..., 1$ derivatives and x itself. So, x is C^{n+1}.

d) To show $d(x_0,x_1) \leq d(x_0,y) + d(y,x_1)$ for all $y \in \mathbb{R}^n$, let $x(t), y_0(t), y_1(t) : [0,1] \to \mathbb{R}^n$ be minimizers of $E[x]$ such that $x(0) = y_0(0) = x_0$, $x(1) = y_1(1) = x_1$, and $y_0(1) = y_1(0) = y$. Define $Y(t) = (y_1 * y_0)(t)$ to be the concatenation of y_1 and y_0 so that $Y(t) = y_0(2t)$ for $t \in [0,1/2]$ and $Y(t) = y_1(2t-1)$ for $t \in [1/2,1]$. Then,

$$\begin{aligned}
L[Y] = &\int_0^{1/2} 2(g_{ij}(y_0(2t))\dot{y}_0^i(2t)\dot{y}_0^j(2t))^{1/2} \, dt \\
&+ \int_{1/2}^1 2(g_{ij}(y_1(2t-1))\dot{y}_1^i(2t-1)\dot{y}_1^j(2t-1))^{1/2} \, dt \\
= &\int_0^1 (g_{ij}(y_0(t))\dot{y}_0^i(t)\dot{y}_0^j(t))^{1/2} \, dt + \int_0^1 (g_{ij}(y_1(t))\dot{y}_1^i(t)\dot{y}_1^j(t))^{1/2} \, dt \\
= &\, d(x_0,y) + d(y,x_1)
\end{aligned}$$

where the factors of 2 come from applying the chain rule. Finally, Y is a competitor in the class of piecewise smooth curves so that

$$d(x_0,x_1) = L[x] \leq L[Y] = d(x_0,y) + d(y,x_1).$$

For the second part, minimizers of $E[x]$ always have constant speed so that $\mathcal{L}(t,x,\dot{x})$ is constant. Fix $t < s \in [0,1]$. Let

$$y(\tau) = x\left((s-t)\tau + t\right)$$

so that $y \in C(x(t), x(s))$. Since x is a minimizer in $C(x_0, x_1)$, it follows that y is a minimizer in $C(x(t), x(s))$. Thus,

$$\begin{aligned} d(x(t),x(s)) &= \int_0^1 (g_{ij}(y(\tau))\dot{y}^i(\tau)\dot{y}^j(\tau))^{1/2}\ d\tau \\ &= \int_0^1 (s-t)(g_{ij}(x((s-t)\tau+t))\dot{x}^i((s-t)\tau+t)\dot{x}^j((s-t)\tau+t))^{1/2}\ d\tau \\ &= \int_t^s (g_{ij}(x(\tau))\dot{x}^i(\tau)\dot{x}^j(\tau))^{1/2}\ d\tau = \int_t^s \sqrt{2}\mathcal{L}(\tau,x,\dot{x})^{1/2}\ d\tau \\ &= \sqrt{2}\mathcal{L}(\tau,x,\dot{x})^{1/2}|s-t| = |s-t|\int_0^1 \sqrt{2}\mathcal{L}(t,x,\dot{x})^{1/2}\ d\tau \\ &= |s-t|d(x(0),x(1)) = |s-t|d(x_0,x_1). \end{aligned}$$

Remark A.1. We have to assume x is a minimizer of E and not just a minimizer of L. For example, consider $\mathbb{R}^2$ with the Euclidean metric, $x_0 = (0,0)$ and $x_1 = (1,1)$. Then $x(t) = (t^2, t^2)$ minimizes the length functional, but

$$\begin{aligned} d(x(t),x(s)) &= \int_t^s (g_{ij}(x(\tau))\dot{x}^i(\tau)\dot{x}^j(\tau))^{1/2}\ d\tau = \int_t^s \sqrt{\dot{x}^1(\tau)^2+\dot{x}^2(\tau)^2}\ d\tau \\ &= \int_t^s 2\sqrt{2}\tau\ d\tau = \sqrt{2}|s-t|^2 = |s-t|^2 d(x_0,x_1). \end{aligned}$$

So, if we assume that x is a minimizer of L, then we need to also assume it has constant speed. But, this is equivalent to assuming x minimizes E, as shown in the remark in part b).

A.2. **Problem 2 [Converse of Theorem 2.20].** Fix $N > 0$, $K \in \mathbb{R}$, $\phi \in C^2(\mathbb{R}^n)$ and let $d(x,y) = |x-y|$ and vol denote the Euclidean distance and volume on $X := \mathbb{R}^n$. Setting $dm(x) = e^{-\phi(x)}d\,\mathrm{vol}(x)$, we've seen that $(X,d,m) \in CD^e(K,N)$ if $N = n$ and ϕ is constant and $K \leq 0$, or if $N > n$ and

$$D^2\phi - \frac{1}{N-n}D\phi \otimes D\phi - KI \geq 0 \tag{1}$$

for all $x \in \mathbb{R}^n$. Show the converse by constructing d_2-geodesics along which $(Kd_2(\mu_0,\mu_1)^2, N)$-convexity of the entropy fails if

a) $N > n$ but (1) is violated for some $x \in \mathbb{R}^n$;
b) $N = n$ but ϕ is constant or $K > 0$;
c) $0 < N < n$.

Solution: Let $\mu_0, \mu_1 \in \mathcal{P}_2^{ac}(\mathbb{R}^n)$ and recall that

$$e'(0) = \int_{\mathbb{R}^n} [\langle Du(x), D\phi(x)\rangle - \Delta u(x)]\, d\mu_0(x)$$

$$e''(0) = \int_{\mathbb{R}^n} \left[\langle Du(x), D^2\phi(x)Du(x)\rangle + \mathrm{Tr}\left([D^2u(x)]^2\right)\right] d\mu_0(x)$$

where $u(x) = U(x) - 1/2|x|^2$ and $G(x) = DU(x)$ is the Brenier map from μ_0 to μ_1. Note that we can easily construct geodesics simply by specifying u and μ_0. If u is λ-convex with $\lambda \geq -1$, then $U(x)$ is convex and we may define $G(x) = DU(x) = Du(x) + x$. Setting $\mu_1 = G_\# \mu_0$, then by uniqueness G is the Brenier map from μ_0 to μ_1. Let $G_t(x) = (1-t)x + tG(x) = x + tDu(x)$; then the geodesic connecting μ_0 to μ_1 is $\mu_t = (G_t)_\# \mu_0$.

a) In Theorem 2.20

$$D^2\phi(x) - \frac{1}{N-n}(D\phi \otimes D\phi)(x) - K\,\mathrm{Id} \geq 0$$

for all $x \in \mathbb{R}^n$, then we showed (K,N)-convexity as follows:

$$\begin{aligned}
\frac{1}{N}[e'(0)]^2 &\leq \frac{1}{N}\int_{\mathbb{R}^n}\left[\langle Du(x), D\phi(x)\rangle - \Delta u(x)\right]^2 d\mu_0(x)\\
&\leq \frac{1}{N}\int_{\mathbb{R}^n}\left[(1+\epsilon^{-1})\langle Du(x), D\phi(x)\rangle^2 + (1+\epsilon)\left(\mathrm{Tr}[D^2u(x)]\right)^2\right] d\mu_0(x)\\
&\leq \int_{\mathbb{R}^n}\left[\frac{1}{N}(1+\epsilon^{-1})\langle Du(x), D\phi(x)\rangle^2 + \frac{n}{N}(1+\epsilon)\,\mathrm{Tr}\left([D^2u(x)]^2\right)\right] d\mu_0(x)\\
&= \int_{\mathbb{R}^n}\left[\frac{1}{N-n}\langle Du(x), D\phi(x)\rangle^2 + \mathrm{Tr}\left([D^2u(x)]^2\right)\right] d\mu_0(x)\\
&= e''(0) - \int_{\mathbb{R}^n}\left[\langle Du(x), D^2\phi(x)Du(x)\rangle - \frac{1}{N-n}\langle Du(x), D\phi(x)\rangle^2\right] d\mu_0(x)\\
&\leq e''(0) - K\int_{\mathbb{R}^n}|Du(x)|^2\, d\mu_0(x) = e''(0) - K\int_{\mathbb{R}^n}|G(x)-x|^2\, d\mu_0(x)\\
&= e''(0) - Kd_2(\mu_0,\mu_1)^2
\end{aligned}$$

where we have applied Jensen's inequality, Young's inequality, and the following inequality

$$(\mathrm{Tr}[A])^2 \leq n\,\mathrm{Tr}[A^2]$$

for a positive definite, symmetric matrix A, and $\epsilon = N/n - 1$.

The goal then is to judiciously choose u and μ_0 so that the first three inequalities are equalities (or, near equality), but the final inequality fails.

Let $x_0, v \in \mathbb{R}^n$ be such that

$$\langle v, D^2\phi(x_0)v\rangle - \frac{1}{N-n}\langle v, D\phi(x_0)\rangle < K|v|^2.$$

First note that Jensen's inequality is saturated when μ_0 is close to a Dirac mass – choose $d\mu_0(x) = \chi_{B_r(x_0)}(x)/|B_r(x_0)|dx$ with $r > 0$ very small. We will construct u so that Young's inequality and the trace inequality above are saturated, and that $Du(x_0) = v$. First, equality in

$$(a-b)^2 \leq (1+\epsilon^{-1})a + (1+\epsilon)b$$

holds when $b = -a/\epsilon$. Since μ_0 is essentially a Dirac mass at x_0, we only need to saturate Young's inequality at x_0. Here, $a = \langle Du(x_0), D\phi(x_0)\rangle$ while $b = \Delta u(x_0)$.
The inequality $\mathrm{Tr}[A]^2 \leq n\,\mathrm{Tr}[A^2]$ follows from Cauchy-Schwarz with the trace norm. So, equality holds when $A = \eta\,\mathrm{Id}$ for some $\eta \in \mathbb{R}_+$. As with Young's inequality, we only need to saturate this near the point x_0. Taking $A = D^2u(x_0)$ we have that $D^2u(x_0) = \eta\,\mathrm{Id}$ for some $\eta \in \mathbb{R}_+$ (note: this is consistent with u needing to be λ-convex for $\lambda \geq -1$, since this implies u is η-convex). To summarize,

$$Du(x_0) = v, \qquad D^2u(x_0) = \eta\,\mathrm{Id}, \qquad \Delta u(x_0) = \frac{n}{n-N}\langle Du(x_0), D\phi(x_0)\rangle$$

with $\eta \geq 0$ to be selected. Combining all three we have

$$n\eta = \frac{n}{n-N}\langle Du(x_0), D\phi(x_0)\rangle = \frac{n}{n-N}\langle v, D\phi(x_0)\rangle,$$

and so

$$\eta = \frac{1}{n-N}\langle v, D\phi(x_0)\rangle.$$

(I guess we should also select v so that the above is non-negative). The following quadratic choice of u works:

$$u(x) = \langle v, x\rangle + \frac{1}{2(n-N)}\langle v, D\phi(x_0)\rangle|x-x_0|^2.$$

b) We must do something slightly different, as the previous chain of inequalities to show $1/N[e'(0)]^2 \leq e''(0) - Kd_2(\mu_0,\mu_1)^2$ will not work. First if ϕ is constant then

$$e'(0) = -\int_{\mathbb{R}^n} \Delta u(x)\ d\mu_0(x), \qquad e''(0) = \int_{\mathbb{R}^n} \mathrm{Tr}\left([D^2u(x)]^2\right)\ d\mu_0(x).$$

Consequently, we need only Jensen's inequality and the trace inequality:

$$\frac{1}{n}[e'(0)]^2 \leq \frac{1}{n}\int_{\mathbb{R}^n} [\Delta u(x)]^2\ d\mu_0(x) \leq \int_{\mathbb{R}^n} \mathrm{Tr}([D^2u(x)]^2)\ d\mu_0(x) = e''(0)$$

Equality holds when μ_0 is close to a Dirac mass at some point $x_0 \in \mathbb{R}^n$, and $u(x) = \eta/2|x|^2$ for $\eta \geq 0$. Defining $U(x) = u(x) + 1/2|x|^2 = (\eta+1)/2|x|^2$ and $\mu_1 = (DU)_{\#}\mu_0$, we have that

$$d_2(\mu_0,\mu_1)^2 = \int_{\mathbb{R}^n} |DU(x) - x|^2\ d\mu_0(x) = \int_{\mathbb{R}^n} |\eta x|^2 d\mu_0(x) \approx \eta^2|x_0|^2.$$

Then,

$$\begin{aligned} Kd_2(\mu_0,\mu_1)^2 + \frac{1}{n}[e'(0)]^2 &\approx K\eta^2|x_0|^2 + \frac{1}{n}[e'(0)]^2 \\ &\approx K\eta^2|x_0|^2 + e''(0) > e''(0). \end{aligned}$$

By taking η large we derive the appropriate contradiction.

If ϕ is not constant, we need to be a little smarter. We still need to use Jensen's inequality, and so we'll take μ_0 to be a Dirac mass at x_0 as before. Ideally we would take $\epsilon \to 0$, but the $1 + \epsilon^{-1}$ contribution from Young's inequality ruins this. In the above, we were saved since $D\phi$ was zero. Now, we may select u so that $Du(x_0) = 0$. Again, we only need this condition at x_0 since μ_0 is concentrating near x_0. I think $u(x) = \eta/2|x - x_0|^2$ for large/small η (depending on the sign of K) works.

c) Since $0 < N < n$, it suffices to check the following inequality holds

$$\frac{1}{N}[e'(0)]^2 > \frac{1}{n}[e'(0)]^2 > e''(0) - Kd_2(\mu_0,\mu_1)$$

for some choice of μ_0, μ_1 (really, a choice of μ_0 and u, as before). This puts us back in the case when $N = n$. If $K > 0$, then we are done by b). If $K \leq 0$, then we are done by b) if ϕ is not a constant. So, we are reduced to the case when $0 < N < n$, $K \leq 0$, and ϕ is constant. Using similar logic as before, if $u(x) = \eta/2|x|^2$ then $[e'(0)]^2 \approx ne''(0)$ and

$$Kd_2(\mu_0,\mu_1)^2 + \frac{1}{N}[e'(0)]^2 \approx K\eta^2|x_0|^2 + \frac{n}{N}e''(0).$$

We want this to bound $e''(0)$ above, so that

$$K\eta^2|x_0|^2 > \left(1 - \frac{n}{N}\right)e''(0) = \frac{N-n}{N}e''(0).$$

Since $K \leq 0$, by taking η small enough we get the appropriate inequality.

A.3. Problem 3 [New $CD^e(K,N)$ spaces from old].

a) Given function $V_1 : X \to (-\infty,\infty]$ and $V_2 : X \to (-\infty,\infty]$ on a metric space (X,d) with V_i weakly (K_i,N_i)-convex for $i \in \{1,2\}$ and V_2 strongly (K_2,N_2)-convex, show $V := V_1 + V_2$ is weakly (K_1+K_2,N_1+N_2)-convex.

b) Given $(X,d,m) \in CD^e(K_1,N_1)$ with $m(X) < \infty$ and $V : X \to [0,\infty]$ Borel and strongly (K_2,N_2)-convex, show $(X,d,e^{-V}m) \in CD^e(K_1+K_2,N_1+N_2)$.

Solution:

a) We say that $V : X \to (-\infty,\infty]$ is weakly (K,N)-convex if for any $x_0, x_1 \in X$ there exists a geodesic $\gamma : [0,1] \to X$ with $\gamma(0) = x_0, \gamma(1) = x_1$ such that $V \circ \gamma$ is semiconvex and satisfies

$$\frac{d^2}{dt^2}(V(\gamma(t))) \geq \frac{1}{N}\left(\frac{d}{dt}(V(\gamma(t)))\right)^2 + K.$$

We say that V is strongly (K,N)-convex if the above holds for all geodesics γ in X. If $N=\infty$, the first derivative term is omitted.

Let's first show that the sum of two semiconvex functions is semiconvex. Recall that $f:\mathbb{R}\to\mathbb{R}\cup\{\infty\}$ is semiconvex if there exists $\omega:\mathbb{R}^+\to\mathbb{R}^+$ such that $\omega(r)=o(r)$ as $r\to 0$ and, for any constant-speed geodesic $\gamma:[0,1]\to\mathbb{R}$ we have

$$f(\gamma(t))\leq(1-t)f(\gamma_0)+tf(\gamma_1)+t(1-t)\omega(d(\gamma_0,\gamma_1)).$$

(Remark: taking $\omega(r)=\lambda/2r^2$ gives λ-convexity, which is implied by the above differential inequality). Let f,g be semicovex with modulus of continuity ω_f,ω_g. Then,

$$[f+g](\gamma(t))\leq(1-t)[f+g](\gamma_0)+t[f+g](\gamma_1)+t(1-t)[\omega_f+\omega_g](d(\gamma_0,\gamma_1)).$$

All we need to check is that $\omega_f+\omega_g$ is $o(r)$ as $r\to 0$, but that is clear since

$$\lim_{r\to 0}\frac{\omega_f(r)+\omega_g(r)}{r}=\lim_{r\to 0}\frac{\omega_f(r)}{r}+\lim_{r\to 0}\frac{\omega_g(r)}{r}=0.$$

In what follows we interpret $1/N$ as 0 when $N=\infty$, so as to avoid multiple cases. Since V_1 and V_2 are semiconvex, by the above so too is V. We now check the differential inequality holds. Let $\gamma:[0,1]\to X$ be a geodesic such that $V_1\circ\gamma$ is (K,N)-convex. Then, since V_2 is strongly (K,N)-convex we also have $V_2\circ\gamma$ is too. Denoting by $v_i(t)$ the quantity $d/dtV_i(\gamma(t))$ for $i=1,2$ we have

$$\begin{aligned}\frac{d^2}{dt^2}(V(\gamma(t)))&=\frac{d^2}{dt^2}(V_1(\gamma(t)))+\frac{d^2}{dt^2}(V_2(\gamma(t)))\\&\geq\frac{1}{N_1}v_1(t)^2+\frac{1}{N_2}v_2(t)^2+[K_1+K_2].\end{aligned}$$

If either N_1 or N_2 are infinite, then we are done (since $1/(N_1+N_2)=0$, and each first order term above is non-negative). So we assume N_1,N_2 are finite. Notice that

$$\begin{aligned}\frac{1}{N_1+N_2}\left(v_1(t)+v_2(t)\right)^2&\leq\frac{1}{N_1+N_2}\left(v_1(t)^2+2v_1(t)v_2(t)+v_2(t)^2\right)\\&\leq\frac{1}{N_1+N_2}\left((1+\epsilon)v_1(t)^2+(1+\epsilon^{-1})v_2(t)^2\right)\\&=\frac{1+\epsilon}{N_1+N_2}v_1(t)^2+\frac{1+\epsilon^{-1}}{N_1+N_2}v_2(t)^2.\end{aligned}$$

We are done if we can select $\epsilon>0$ such that $(1+\epsilon)/(N_1+N_2)\leq 1/N_1$ and $(1+\epsilon^{-1})/(N_1+N_2)\leq 1/N_2$. The former implies holds whenever

$$\epsilon\leq\frac{N_2}{N_1}$$

while the latter holds if

$$\frac{N_2}{N_1}\leq\epsilon,$$

so we take $\epsilon=N_2/N_1$.

b) Recall since $(X,d,m)\in CD^e(K,N)$, for $\mu_0,\mu_1\in\mathcal{P}_2(X)$ there exists a d_2-geodesic $(\mu_t)_{t\in[0,1]}$ connecting them such that $t\mapsto S(\mu_t\mid m)$ is $(Kd_2(\mu_0,\mu_1)^2,N)$-convex, where

$$S(\mu\mid m)=\begin{cases}\int_X d\mu/dm\log\left(d\mu/dm\right)dm & \mu\ll m\\ \infty & \text{else}\end{cases}.$$

Suppose first that $\mu_0, \mu_1 \ll m$. Then $d\mu_t/(d(e^{-V}m)) = e^V d\mu_t/dm$. Hence,

$$\begin{aligned} S(\mu_t \mid e^{-V}m) &= \int_X e^V \frac{d\mu_t}{dm} \log\left(e^V \frac{d\mu_t}{dm}\right) e^{-V} dm \\ &= \int_X V\ d\mu_t + \int_X \frac{d\mu_t}{dm} \log\left(\frac{d\mu_t}{dm}\right) dm \\ &= \int_X V\ d\mu_t + S(\mu_t \mid m). \end{aligned}$$

Owing to part a), since we know that $S(\mu_t \mid m)$ is $(K_1 d_2(\mu_0,\mu_1)^2, N_1)$-convex, it suffices to show that $I(\mu) = \int_X V\ d\mu$ is strongly (K_2, N_2)-convex (on $\mathcal{P}_2(X)$).

Since V is strongly (K_2, N_2)-convex, if γ_t is a geodesic in X then

$$e^{-V(\gamma_t)/N_2} \geq \sigma^{(1-t)}_{K_2/N_2}(d(\gamma_0,\gamma_1))e^{-V(\gamma_0)/N_2} + \sigma^{(t)}_{K_2/N_2}(d(\gamma_0,\gamma_1))e^{-V(\gamma_1)/N_2}$$

with

$$\sigma^t_k(r) = \begin{cases} s_k(tr)/s_k(r) & \text{if } kr^2 \leq \pi^2 \\ \infty & \text{else} \end{cases}$$

and

$$s_k(r) = \begin{cases} \sin(\sqrt{k}r) & k > 0 \\ r & k = 0 \\ \sinh(\sqrt{-k}r) & k < 0 \end{cases}.$$

We now use Villani's approach of dynamic optimal couplings. Denote by $\mathrm{Geo}(X)$ the space of geodesics on X (where the underlying metric of X is implicit). Note that if e_t is the evaluation map at time t of a geodesic, and $\pi \in \mathcal{P}(\mathrm{Geo}(X))$, then $\Gamma_t := (e_t)_\#\pi \in \mathrm{Geo}(\mathcal{P}_2(X))$. In fact, all geodesics on $\mathcal{P}_2(X)$ arise in this way. So it suffices to use this formulation to check convexity – we want to show that

$$e^{-I(\Gamma_t)/N_2} \geq \sigma^{(1-t)}_{K_2/N_2}(d_2(\Gamma_0,\Gamma_1))e^{-I(\Gamma_0)/N_2} + \sigma^{(t)}_{K_2/N_2}(d_2(\Gamma_0,\Gamma_1))e^{-I(\Gamma_1)/N_2}$$

for any $\Gamma_t \in \mathrm{Geo}(\mathcal{P}_2(X))$. To this end select such a Γ_t so that $\Gamma_t = (e_t)_\#\pi$ for some $\pi \in \mathcal{P}_2(\mathrm{Geo}(X))$. The above inequality, for fixed t, can be viewed in the variable γ. That is,

$$e^{-V(e_t(\gamma))/N_2} \geq \sigma^{(1-t)}_{K_2/N_2}(d(\gamma_0,\gamma_1))e^{-V(e_0(\gamma))/N_2} + \sigma^{(t)}_{K_2/N_2}(d(\gamma_0,\gamma_1))e^{-V(e_1(\gamma))/N_2}$$

where t is fixed and $\gamma \in \mathrm{Geo}(X)$. Taking the logarithm of the above and integrating against π yields

$$\begin{aligned} &-\frac{1}{N_2}\int_{\mathrm{Geo}(X)} V(e_t(\gamma))\ d\pi(\gamma) \geq \\ &\int_{\mathrm{Geo}(X)} \log\left(\sigma^{(1-t)}_{K_2/N_2}(d(\gamma_0,\gamma_1))e^{-V(e_0(\gamma))/N_2} + \sigma^{(t)}_{K_2/N_2}(d(\gamma_0,\gamma_1))e^{-V(e_1(\gamma))/N_2}\right)\ d\pi(\gamma). \end{aligned}$$

The left hand side, owing to the definition of π, is just $-1/N_2 I(\Gamma_t)$ (for variable time Γ_t is a curve in the space $\mathcal{P}_2(X)$, but since t is fixed it is just a particular measure). Exponentiating gives the desired left-hand side. What we aim to show, then, is that

$$\begin{aligned} &\int_{\mathrm{Geo}(X)} \log\left(\sigma^{(1-t)}_{K_2/N_2}(d(\gamma_0,\gamma_1))e^{-V(e_0(\gamma))/N_2} + \sigma^{(t)}_{K_2/N_2}(d(\gamma_0,\gamma_1))e^{-V(e_1(\gamma))/N_2}\right)\ d\pi(\gamma) \geq \\ &\log\left(\sigma^{(1-t)}_{K_2/N_2}(d_2(\Gamma_0,\Gamma_1))e^{-I(\Gamma_0)/N_2} + \sigma^{(t)}_{K_2/N_2}(d_2(\Gamma_0,\Gamma_1))e^{-I(\Gamma_1)/N_2}\right). \end{aligned}$$

This follows from Jensen's inequality, if the functional

$$F_t(x,y,k) = \log\left[\sigma^{(1-t)}_k(1)e^x + \sigma^{(t)}_k e^y(1)\right]$$

is convex, since we apply this with $x = -1/N_2 V(\gamma_0)$, $y = -1/N_2 V(\gamma_1)$, and $k = K_2/N_2 d(\gamma_0,\gamma_1)^2$. (Note that we only need the above arguments of $\sigma^{(t)}_k$ to be 1 since $\sigma^{(t)}_k(d) = \sigma^{(t)}_{kd^2}(1)$). First, the auxiliary function $F(x,y) = \log(e^x + e^y)$ is convex. Next, $a \mapsto F(x+a, y)$ and

$b \mapsto F(x, y+b)$ are increasing. Finally, $k \mapsto \log(\sigma_k^{(t)}(1))$ is convex. Composing these actions gives us that $F_t(x, y, k)$ is convex. (I'm not sure if this works, but it's the best I can do..)

A.4. **Problem 4 [Γ-convergence].** For $k \in \mathbb{N}_\infty := \mathbb{N} \cup \{\infty\}$, let $f_k : X \to (-\infty, \infty]$ be a sequence of lower semi-continuous functions on a compact metric space (X, d). We say f_k has Γ-limit f_∞ if the following two criteria are satisfied:

i) Whenever $\lim_{k\to\infty} d(x_k, x_\infty) = 0$,
$$f_\infty(x_\infty) \le \liminf_{k\to\infty} f_k(x_k);$$
ii) Every $y_\infty \in X$ is the limit of a sequence $y_k \in X$ along which
$$f_\infty(y_\infty) = \lim_{k\to\infty} f_k(y_k).$$

a) Show that if f_k has Γ-limit f_∞ then any sequence $x_k \in \arg\min f_k$ has a subsequence whose limit $x_\infty \in \arg\min f_\infty$.
b) Give an example of Γ-convergence in which there are elements of $\arg\min f_\infty$ not approximated by any such sequence.

Solution:

a) Since (X, d) is compact the sequence $\{x_k\}_{k=1}^\infty$ admits a convergent subsequence. Suppose that $\{x_{k_n}\}_{n=1}^\infty$ is such a subsequence, and that $x_{k_n} \to x_\infty \notin \arg\min f_\infty$. Since f_∞ is lower semi-continuous on the compact space (X, d), it admits a minimizer $\overline{x}_\infty \in \arg\min f_\infty$. Then, by property ii) there exists $\{x_k\}_{k=1}^\infty$ such that
$$f_\infty(\overline{x}_\infty) = \lim_{k\to\infty} f_k(\overline{x}_k) = \lim_{n\to\infty} f_{k_n}(\overline{x}_{k_n}) \ge \liminf_{n\to\infty} f_{k_n}(x_{k_n})$$
where the last inequality follows since $x_k \in \arg\min f_k$ for each k. On the other hand, since $x_{k_n} \to x_\infty$ we have $\lim_{n\to\infty} d(x_{k_n}, x_\infty) = 0$. Thus by property i),
$$f_\infty(x_\infty) \le \liminf_{n\to\infty} f_{k_n}(x_{k_n}).$$
Combining these two inequalities with the fact that $x_\infty \notin \arg\min f_\infty$ while $\overline{x}_\infty$ is, we have
$$\liminf_{n\to\infty} f_{k_n}(x_{k_n}) \le f_\infty(\overline{x}_\infty) < f_\infty(x_\infty) \le \liminf_{n\to\infty} f_{k_n}(x_{k_n}),$$
a contradiction. So, in fact all convergent subsequences of $\{x_k\}_{k=1}^\infty$ converge to a minimizer of f_∞.

b) Let $(X, d) = ([0,1], |\cdot|)$ where $|\cdot|$ is the usual Euclidean distance. Define $f_k : [0,1] \to (-\infty, \infty]$ by
$$f_k(x) = -\frac{x}{k} + 1.$$
We show that the Γ-limit of f_k is $f_\infty \equiv 1$.

i) Suppose that $\{x_k\}_{k=1}^\infty \subset [0,1]$ is such that $d(x_k, x_\infty) \to 0$. Then,
$$f_\infty(x_\infty) = 1 = \left(1 - \frac{x_k}{k}\right) + \frac{x_k}{k} \le f_k(x_k) + \frac{1}{k},$$
from which it is clear that $f_\infty(x_\infty) \le \liminf_{k\to\infty} f_k(x_k)$.

ii) For $y_\infty \in [0,1]$ select the constant sequence $y_k = y_\infty$.

Now, $\arg\min f_k = \{1\}$ while $\arg\min f_\infty = [0,1]$. So, for any $x_\infty \in [0,1)$ it is impossible to find such a sequence.

References

[ACM+21] A. Akdemir, A. Colinet, R. McCann, F. Cavalletti, and F. Santarcangelo. Independence of synthetic curvature dimension conditions on transport distance exponent. Trans. Amer. Math. Soc., 374:5877–5923, 2021.

[AGS14] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. Duke Mathematical Journal, 163(7):1405–1490, 2014.

[BBI01] D. Burago, Yu. Burago, and S. Ivanov. A Course in Metric Geometry, volume 33 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2001.

[CM] F. Cavalletti and E. Milman. The globalization theorem for the curvature dimension condition. submitted.

[EKS15] M. Erbar, K. Kuwada, and K.-T. Sturm. On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. Invent. Mathm., 201:993–1071, 2015.

[GMS15] N. Gigli, A. Mondino, and G. Savaré. Convergence of pointed non-compact metric measure spaces and stability of Ricci curvature bounds and heat flows. Proc. London Math. Soc., 111(5):1071–1129, 2015.

[GP91] K. Grove and P. Petersen. Manifolds near the boundary of existence. J. Diff. Geom., 33:379–394, 1991.

[LV09] J. Lott and C. Villani. Ricci curvature via optimal transport. Ann. of Math., 169:903–991, 2009.

[RS14] T. Rajala and K.-T. Sturm. Non-branching geodesics and optimal maps in strong $CD(K,\infty)$-spaces. Calc. Var. Partial Differential Equations, 50:831–846, 2014.

[San15] F. Santambrogio. Optimal transport for applied mathematicians, volume 87 of Progress in Nonlinear Differential Equations and Their Applications. Birkhäuser, Cham, 2015.

[Stu06a] K.-T. Sturm. On the geometry of metric measure spaces. i. Acta Math., 196(1):65–131, 2006.

[Stu06b] K.-T. Sturm. On the geometry of metric measure spaces. ii. Acta Math., 196(1):133–177, 2006.

[Vil03] C. Villani. Topics in Optimal Transportation, volume 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2003.

[Vil09] C. Villani. Optimal transport, old and new, volume 338 of Grundlehren des mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin-New York, 2009.