An Introduction to Gromov-Wasserstein Distances

Dongming (Merrick) Hua

September 20th, 2023

These notes were written for a seminar on optimal transport and particle physics in Summer 2023 at UCSB, organized by Prof. Katy Craig. Most of the contents within are based off of Sturm's paper "On the geometry of metric spaces. I" [2] and Mémoli's paper "Gromov–Wasserstein Distances and the Metric Approach to Object Matching" [1]. Basic knowledge of metric spaces and measure theory are assumed, as well as some familiarity with the Wasserstein p-distances.

1 Introduction

A few weeks ago Haoqing told us about the Wasserstein p-distances, which give us a way to compare how close two probability measures on a metric space are. But what if we want to compare measures on different metric spaces? This isn't just an issue of mathematical curiosity—indeed, many problems in object matching can be viewed in such a context, as seen in Memoli's paper [1]

The Gromov-Wasserstein distances give us a way to do this. The main idea is that since the Wasserstein distances allow us to compare measures on the same metric space, we should turn measures on different metric spaces into measures living in the same metric space.

2 Background

Our basic objects of study are metric measure spaces:

Definition 2.1 (Metric Measure Space)**.** *A* ***metric measure space*** *(mm-space for short) is a triple (X, d, μ), where (X, d) is a metric space and μ is a Borel measure on (X, d) (a measure on the Borel σ-algebra of (X, d)) satisfying the following properties:*

- *The metric space (X, d) is complete (Cauchy sequences converge) and separable (contains a countable dense subset).*

- *The measure μ is locally finite: for any $x \in X$ and sufficiently small $r > 0$ we have $\mu(B_r(x)) < \infty$ (small enough balls have finite measure).*

For our purposes we will require that our metric measure spaces are **normalized**, i.e., that $\mu(X) = 1$. Another key concept is that of the support:

Definition 2.2. *Given a metric measure space (X, d, μ), the **support** of μ, denoted $\operatorname{supp}[\mu]$, is the set*

$$\operatorname{supp}[\mu] = \{x \in X \mid \textit{any open set containing } x \textit{ has positive measure}\}.$$

Note that the support is always closed, because its complement is the union of open sets of measure zero. Essentially, the support is where we actually have measure-theoretic information in a metric measure space.

Example 2.3. *Euclidean space $\mathbb{R}^n$ with the usual Euclidean distance and the Lesbesgue measure is a metric measure space, but unfortunately this measure cannot be normalized. This measure has support all of $\mathbb{R}^n$.*

Example 2.4. *Given any (nonempty) metric space (X, d), we can make it a metric measure space by picking a point $x \in X$ and putting a Dirac mass there. In other words, we give it the measure δ_x which returns 1 when a measurable set contains x, and 0 otherwise. Note that this is normalized, and has support $\operatorname{supp}[\delta_x] = \{x\}$.*

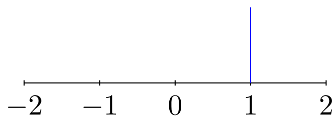

Figure 1: The metric space $[-2, 2]$ with a Dirac mass at 1.

Example 2.5. *One standard example (which we will see later) is to take $\mathbb{R}^n$ and put Dirac masses of weight $\frac{1}{n+1}$ on each of the vertices of the standard n-simplex. The support of this measure is simply $n+1$ points all at distance 1 from each other, each with a weight of $\frac{1}{n+1}$. Note that this is normalized.*

Example 2.6. *A compact Riemannian manifold with metric and renormalized volume measure both induced by the Riemannian metric is a normalized metric measure space. For example, take the rescaled volume measure on S^n.*

Whenever we have an object with some structure, we always want to know what it means for two such structures to be the same:

Definition 2.7. *Two metric measure spaces (X_1, d_1, μ_1) and (X_2, d_2, μ_2) are isomorphic (as mm-spaces) if there exists a map $f\colon \operatorname{supp}[\mu_1] \to \operatorname{supp}[\mu_2]$ which is an isometry and satisfies*

$$\mu_2 = f_{\#}\mu_1.$$

Here $f_{\#}\mu_1$ is the pushforward: we have $f_{\#}\mu_1(B) = \mu_1(f^{-1}(B))$ for any Borel set B in X_2.

Finally, an important isomorphism invariant of metric measure spaces is the variance:

Definition 2.8. *Let (X, d, μ) be a metric measure space. The* ***variance*** *of (X, d, μ) is*

$$\mathrm{Var}(X, d, \mu) = \inf \int_{X'} d'(x, z)^2 d\mu',$$

where the infimum is taken over all metric measure spaces (X', d', μ') isomorphic to (X, d, μ), and all choices of basepoint $z \in X'$.

Remark: Excuse the abuse of notation in using d for both a metric and for integration; this will continue throughout these notes.

Remark: Note that the quantity $\int_{X'} d'(x, z)^2 d\mu'$ is just the second moment of (X', d', μ'), which Haoqing discussed in his talk.

Finiteness of the variance is an important condition to ensure finiteness of the Gromov-Wasserstein 2-distance, as we will see later on; we will assume our metric measure spaces to have finite variance.

3 Gromov-Wasserstein Distances

We will work with the Gromov-Wasserstein 2-distance from here on, but keep in mind that we can define the Gromov-Wasserstein p-distance for $1 \leq p < \infty$ analogously, and that many of the following proofs and properties follow similarly. Furthermore, from now on we require that our mm-spaces be normalized and have finite variance.

First we formalize what it means to make two measures on different metric spaces live in the same space.

Definition 3.1. *Let (X_1, d_1, μ_1) and (X_2, d_2, μ_2) be metric measure spaces. A measure μ on the product space $X_1 \times X_2$ is a* ***measure coupling*** *of μ_1 and μ_2 if*

$$\mu(B \times X_2) = \mu_1(B),\ \mu(X_1 \times B') = \mu_2(B')$$

for all measurable $B \subseteq X_1$ and $B' \subseteq X_2$.

Remark: When $(X_1, d_1, \mu_1) = (X_2, d_2, \mu_2)$ we recover the notion of coupling Haoqing discussed in his talk.

Definition 3.2. *Let (X_1, d_1, μ_1) and (X_2, d_2, μ_2) be metric measure spaces. A metric d on the disjoint union $X_1 \sqcup X_2$ is a* ***metric coupling*** *of d_1 and d_2 if $d(x, y) = d_1(x, y)$ and $d(x', y') = d_2(x', y')$ for all $x, y \in \mathrm{supp}[\mu_1]$ and $x', y' \in \mathrm{supp}[\mu_2]$.*

Now we can define the Gromov-Wasserstein 2-distance:

Definition 3.3. *Let (X_1, d_1, μ_1) and (X_2, d_2, μ_2) be metric measure spaces. The **Gromov-Wasserstein 2-distance** between the two spaces is defined to be*

$$d_{GW,2}\left((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)\right) = \inf \left(\int_{X_1 \times X_2} d(x, y)^2 d\mu \right)^{\frac{1}{2}},$$

where the infimum is taken over all measure couplings μ of μ_1, μ_2 and metric couplings d of d_1, d_2.

We will drop the 2 in the subscript from here on out. The careful reader may note that this process doesn't quite involve a common metric measure space that both (X_1, d_1, μ_1) and (X_2, d_2, μ_2) live in, because the metric coupling d is not a metric on $X_1 \times X_2$. However, the following fact takes care of this:

Proposition 3.4. *We have*

$$d_{GW}\left((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)\right) = \inf d_W(\phi_{\#}\mu_1, \phi'_{\#}\mu_2),$$

where the infimum is taken over all metric spaces (X, d) with isometric embeddings $\phi\colon \operatorname{supp}[\mu_1] \to X$ and $\phi'\colon \operatorname{supp}[\mu_2] \to X$, and d_W is the Wasserstein 2-distance.

Proof. See Lemma 3.3 of [2]. □

Note that d_{GW} is the infimum of nonnegative numbers, hence is nonegative. Furthermore, it is the infimum of quantities which are symmetric with respect to (X_1, d_1, μ_1) and (X_2, d_2, μ_2), hence it itself is symmetric. By the alternate definition given in Proposition 3.4., we also see that isomorphic metric measure spaces have Gromov-Wasserstein distance zero. The triangle inequality follows from placing a clever choice of metric on the three spaces involved (see Theorem 3.6 in [2]). The distance being nonzero when the spaces involved are *nonisomorphic* follows from a comparison with Gromov's box metric (see Lemma 3.7 in [2]). Finally, the fact that the Gromov-Wasserstein distance is finite follows from the triangle inequality and one of the examples in the next section. All this comes together to show that

Theorem 3.5. *The Gromov-Wasserstein distance d_{GW} is a metric on the set of equivalence classes of normalized metric measure spaces with finite variance.*

4 Examples

Lemma 4.1. *Let (X, d, μ) be a normalized metric measure space with finite variance, and let $(X', d', \delta_{x'})$ be a metric measure space with measure given by a Dirac mass at a point $x' \in X'$. Then we have*

$$d_{GW}\left((X, d, \mu), (X', d', \delta_{x'})\right)^2 = \operatorname{Var}(X, d, \mu).$$

Proof. Let $\varepsilon_1 > 0$ be arbitrary, and assume without loss of generality that $z \in X$ achieves $\int_X d(x,z)^2 d\mu < \mathrm{Var}(X,d,\mu) + \varepsilon_1$ (since we only have to consider isomorphism classes, we can take (X,d,μ) to be a representative achieving a value close to the actual variance). Furthermore, we can assuming without loss of generality that $X = \mathrm{supp}[\mu]$.

Now take the isometrically embed X in X via the identity id, and $\mathrm{supp}[\delta_{x'}] = \{x'\}$ in X via ϕ', $x' \mapsto z$. It follows that

$$d_W(\mathrm{id}_\# \mu, \phi'_\# \delta_{x'})^2 = d_W(\mu, \delta_z)^2 \leq \int_X d(x,z)^2 d\mu < \mathrm{Var}(X,d,\mu) + \varepsilon_1.$$

Since $\varepsilon_1 > 0$ is arbitrary, we have

$$d_{GW}\left((X,d,\mu),(X',d',\delta_{x'})\right)^2 \leq d_W(\mathrm{id}_\# \mu, \phi_\# \delta_{x'})^2 \leq \mathrm{Var}(X,d,\mu).$$

On the other hand, any coupling with a delta distribution must end up being a product measure, so any integral for $d_W(\phi_\# \mu, \phi'_\# \delta_{x'})$ must end up being of the form $\int_Y d(y,z)^2 d(\phi_\# \mu)^*$, where $(Y, d_Y, \phi_\# \mu)$ is isomorphic to (X,d,μ). Therefore, $\mathrm{Var}(X,d,\mu) \leq (d_W(\mathrm{id}_\# \mu, \phi'_\# \delta_{x'}) + \varepsilon_2)^2$ for any $\varepsilon_2 > 0$. It follows that

$$\mathrm{Var}(X,d,\mu) \leq d_W(\mathrm{id}_\# \mu, \phi'_\# \delta_{x'})^2,$$

and the desired equality follows.

*Technically we should be taking the integral over a space Y containing an isometrically embedded copy of X and a distinguished point y corresponding to x', but it is clear that having y outside of the copy of X always results in a larger integral (the distances we are integrating are larger), so we can just assume y is inside the copy of X, and disregard everything outside. □

Given the triangle inequality, one consequence of this lemma is that d_{GW} is bounded for our class of metric measure spaces, since we have

$$\begin{aligned} d_{GW}\left((X,d,\mu),(X',d',\mu')\right) &\leq d_{GW}\left((X,d,\mu),(X',d',\delta_{x'})\right) + d_{GW}\left((X',d',\delta_{x'}),(X',d',\mu')\right) \\ &= \sqrt{\mathrm{Var}(X,d,\mu)} + \sqrt{\mathrm{Var}(X',d',\mu')}, \end{aligned}$$

and the variances are finite by assumption.

The following two examples are from Sturm's paper [2], and involve the notion of **Gromov-Wasserstein convergence**: a sequence (X_n, d_n, μ_n) of metric measure space Gromov-Wasserstein converges to a metric measure space (X,d,μ) if the numbers $d_{GW}((X_n,d_n,\mu_n),(X,d,\mu))$ converge to 0. It turns out that the space of all isomorphism classes of normalized, finite variance metric measure spaces is actually complete under the Gromov-Wasserstein distance.

Example 4.2. *For each n, let $X_n = (\frac{1}{n}\mathbb{Z} \cap [0,1])^k$, with the usual metric and normalized counting measure. In other words, X_n is a $(n+1) \times (n+1)$ grid of evenly spaced points in the unit square, each with a weight of $\frac{1}{(n+1)^2)}$. We have $(X_n, d_n, \mu_n) \xrightarrow{GW} (X,d,\mu)$, where (X,d) is $[0,1]^k$ with the usual metric and μ is the Lesbesgue measure.*

This is an example of a dimensional increase: the spaces in the limit are all discrete and thus 0-dimensional, but the limit is k-dimensional.

Example 4.3. *Let X be a finite graph living in $\mathbb{R}^3$, let d be the graph metric on X, and give X the normalized one-dimensional Lesbesgue measure μ. Now, for each n, let X_n be the closed $\frac{1}{n}$-neighborhood of X, and give X_n the geodesic metric and three-dimensional Lesbesgue measure. Then $(X_n, d_n, \mu_n) \xrightarrow{GW} (X, d, \mu)$.*

The same holds if we take X_n to be the points at distance exactly $\frac{1}{n}$, and give this shell the induced geodesic metric and two-dimensional Lesbesgue measure.

This is an example of dimensional collapse: the spaces in the limit are all three (or two) dimensional, but the limit is one dimensional.

5 Applications

Example 5.1. *As mentioned previously, Gromov-Wasserstein distance has applications to object matching: objects can be approximated by finite metric measure spaces, and then the Gromov-Wasserstein distance can tell us how close the resulting approximations are to each other. See Mémoli's paper [1] for more.*

Example 5.2. *Another potential application (as communicated to me by Prof. Katy Craig) is the example of jets in particle colliders Jessica spoke about in her talk. Since the orientation of the jet is not so important, Gromov-Wasserstein can be more useful here than regular Wasserstein, because regular Wasserstein would assign a cost to rotating a jet. More generally, Gromov-Wasserstein can be better in a "change-of-coordinates" situation, where we don't care if two objects are different via a measure-preserving isometry of the same space.*

The issue with both these (more practical) examples is that Gromov-Wasserstein distance is very hard to compute: even for finite metric measure spaces it is NP-hard. Mémoli has an alternate formulation (not quite equivalent) of the distance in his paper [1] which seems more amenable to computation.

Finally, an application to other mathematics.

Example 5.3. *In geometry lower bounds on Ricci curvature are quite useful, and show up in a lot of big theorems (for example Gromov's precompactness theorem). One can define a notion of curvature for metric measure spaces involving Wasserstein distance, which shares nice properties with the Ricci curvature. In particular, lower bounds on this synthetic notion of Ricci curvature agree with lower bounds on Ricci curvature in the case of Riemannian manifolds, and can be used in place of Ricci curvature in more general settings. It turns out that Gromov-Wasserstein is in some sense a "correct" notion of distance for this notion of curvature, because Gromov-Wasserstein convergence preserves lower bounds for this curvature (under suitably nice conditions). See Sturm's paper [2] for more.*

References

[1] Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Found. Comput. Math.*

[2] Karl-Theodor Sturm. On the geometry of metric spaces. i. *Acta Math.*