

Article

# The Gromov–Wasserstein Distance: A Brief Overview

Facundo Mémoli

Department of Mathematics, The Ohio State University, Columbus, OH, USA;

E-Mail: memoli@math.osu.edu; Tel.: +1-614-292-5585.

Received: 1 May 2014; in revised form: 12 August 2014 / Accepted: 22 August 2014 /

Published: 2 September 2014

---

**Abstract:** We recall the construction of the Gromov–Wasserstein distance and concentrate on quantitative aspects of the definition.

**Keywords:** metric geometry; graph theory; shape recognition; optimal transportation

---

## 1. Introduction

Modeling datasets as metric spaces seems to be natural for some applications and concepts revolving around the Gromov–Hausdorff distance—a notion of distance between compact metric spaces—provide a useful language for expressing properties of data and shape analysis methods. In many situations, however, this is not enough, and one must incorporate other sources of information into the model, with “weights” attached to each point being one of them. This gives rise to the idea of representing data as metric measure spaces, which are metric spaces endowed with a probability measure. In terms of a distance, the Gromov–Hausdorff metric is replaced with the Gromov–Wasserstein metric.

### 1.1. Notation and Background Concepts

The book by Burago, *et al.* [1] is a valuable source for many concepts in metric geometry. We refer the reader to that book for any concepts not explicitly defined in these notes.

We let  $\mathcal{M}$  denote the collection of all compact metric spaces and by  $\mathcal{M}^{\text{iso}}$  the collection of all isometry classes of  $\mathcal{M}$ . Recall that for a given metric space  $(X, d_X) \in \mathcal{M}$ , its diameter is defined as  $\text{diam}(X) := \max_{x, x' \in X} d_X(x, x')$ . Similarly, the radius of  $X$  is defined as  $\text{rad}(X) := \min_{x \in X} \max_{x' \in X} d_X(x, x')$ .

For a fixed metric space  $(Z, d_Z)$ , we let  $d_{\mathcal{H}}^Z$  denote the Hausdorff distance between (closed) subsets of  $Z$ .

We will often refer to a metric space  $(X, d_X)$  by only  $X$ , but the notation for the underlying metric will be implicitly understood to be  $d_X$ . Recall that a map  $\varphi : X \rightarrow Y$  between metric spaces  $(X, d_X)$  and  $(Y, d_Y)$  is an isometric embedding if  $d_Y(\varphi(x), \varphi(x')) = d_X(x, x')$  for all  $x, x' \in X$ . The map  $\varphi$  is an isometry if it is a surjective isometric embedding.

Recall that given measurable spaces  $(X, \Sigma_X)$  and  $(Y, \Sigma_Y)$ , a measure  $\mu$  on  $(X, \Sigma_X)$  and a measurable map  $f : X \rightarrow Y$ , the push-forward measure  $f_{\#}\mu$  on  $(Y, \Sigma_Y)$  acts according to  $f_{\#}\mu(B) = \mu(f^{-1}(B))$  for any  $B \in \Sigma_Y$ .

A metric measure space (mm-space for short) is a triple  $(X, d_X, \mu_X)$  where  $(X, d_X)$  is a compact metric space and  $\mu_X$  is a Borel probability measure with full support:  $\text{supp}(\mu_X) = X$ . We denote by  $\mathcal{M}^w$  the collection of all mm-spaces. An isomorphism between  $X, Y \in \mathcal{M}^w$  is any isometry  $\Psi : X \rightarrow Y$ , such that  $\Psi_{\#}\mu_X = \mu_Y$ .

## 2. The Gromov–Hausdorff Distance

One says that a subset  $R \subset X \times Y$  is a correspondence between sets  $X$  and  $Y$  whenever  $\pi_1(R) = X$  and  $\pi_2(R) = Y$ , where  $\pi_1 : X \times Y \rightarrow X$  and  $\pi_2 : X \times Y \rightarrow Y$  are the canonical projections. Let  $\mathcal{R}(X, Y)$  denote the set of all correspondences between  $X$  and  $Y$ .

The Gromov–Hausdorff (GH) distance between compact metric spaces  $(X, d_X)$  and  $(Y, d_Y)$  is defined as:

$$d_{GH}(X, Y) := \frac{1}{2} \inf_R \sup_{(x,y),(x',y') \in R} |d_X(x, x') - d_Y(y, y')| \tag{1}$$

where  $R$  ranges over  $\mathcal{R}(X, Y)$ .

**Example 1.** The GH distance between any compact metric space  $X$  and the space with exactly one point is equal to  $\frac{1}{2} \text{diam}(X)$ .

It turns out that  $(\mathcal{M}, d_{GH})$  is a nice space in that it has many compact subclasses.

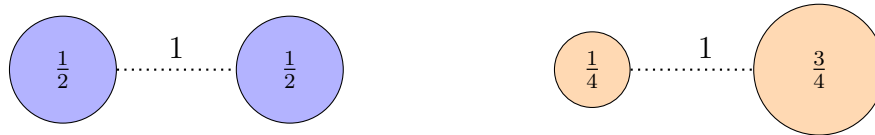
**Theorem 1.** ([1]) *Let  $N : [0, +\infty) \rightarrow \mathbb{N}$  be a bounded function and  $D > 0$ . Let  $\mathcal{F}(N, D) \subset \mathcal{M}$  be any family of compact metric spaces, such that  $\text{diam}(X) \leq D$  for all  $X \in \mathcal{F}(N, D)$ , such that for any  $\varepsilon > 0$ , any  $X \in \mathcal{F}(N, \varepsilon)$  admits an  $\varepsilon$ -net with at most  $N(\varepsilon)$  elements. Then,  $\mathcal{F}(N, D)$  is pre-compact in the Gromov–Hausdorff topology.*

**Example 2.** An important example of families, such as the above, is given by those closed  $n$ -dimensional Riemannian manifolds  $(X, g_X) \in \mathcal{M}(n, \kappa, D)$  with the diameter bounded by  $D > 0$  and the Ricci curvature bounded below by  $\kappa$ .

**Theorem 2** ([2]). *The space  $(\mathcal{M}^{\text{iso}}, d_{GH})$  is complete.*

It then follows from the two theorems above that classes  $\mathcal{F}(N, D)$ , such as above, are totally bounded for the Gromov–Hausdorff distance. This means that such classes are easy to organize in the sense of clustering or databases.

In many practical applications, one would like to take into account “weights” attached to points in a dataset. For example, the two metric spaces with the weights below are isometric, but not isomorphic in the sense that no isometry respects the weights:



The idea is that weights represent how much we trust a given “measurement” in practical applications. This leads to considering a more general collection of datasets and, in turn, an adapted notion of equality and a compatible metric over them. This naturally leads to regarding datasets as mm-spaces and then finding a notion of distance on  $\mathcal{M}^w$  compatible with isomorphism of mm-spaces.

**3. A Metric on  $\mathcal{M}^w$**

Let  $(X, d_X, \mu_X)$  and  $(Y, d_Y, \mu_Y)$  be two given mm-spaces. In our path to defining a distance between mm-spaces, we emulate the construction of the Gromov–Hausdorff distance and start by identifying a notion of correspondence between mm-spaces.

A probability measure  $\mu$  over  $X \times Y$  is called a coupling between  $\mu_X$  and  $\mu_Y$  if  $(\pi_1)_\# \mu = \mu_X$  and  $(\pi_2)_\# \mu = \mu_Y$ . We denote by  $\mathcal{U}(\mu_X, \mu_Y)$  the collection of all couplings between  $\mu_X$  and  $\mu_Y$ .

**Example 3.** When  $Y = \{p\}$ ,  $\mu_Y = \delta_p$ , and thus, there is a unique coupling between  $X$  and  $Y$ :  $\mathcal{U}(\mu_X, \mu_Y) = \{\mu_X \otimes \delta_p\}$ .

**Example 4.** Consider for example the spaces with two points each that we depicted above. In that case,  $\mu_X$  can be identified with the vector  $(\frac{1}{2}, \frac{1}{2})$  and  $\mu_Y$  with the vector  $(\frac{3}{4}, \frac{1}{4})$ . In this case, one sees that the matrix:

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & 0 \end{bmatrix}$$

induces a valid coupling.

Now, given  $p \geq 1$ , consider the function

$$(x, y, x', y') \mapsto |d_X(x, x') - d_Y(y, y')|^p$$

and pick any  $\mu \in \mathcal{U}(\mu_X, \mu_Y)$ . One then integrates this function against the measure  $\mu \otimes \mu$  and infimizes over the choice of  $\mu \in \mathcal{U}(\mu_X, \mu_Y)$  to define the Gromov–Wasserstein distance of order  $p$  [3]:

$$d_{GW,p}(X, Y) := \frac{1}{2} \inf_{\mu} \left( \iint |d_X(x, x') - d_Y(y, y')|^p \mu(dx \times dy) \mu(dx' \times dy') \right)^{1/p}$$

**Remark 1.** This is an  $L^p$  analogue of Equation (1).

**Theorem 3** ([3]). *The Gromov–Wasserstein distance of order  $p \geq 1$  defines a proper distance on the collection of isomorphism classes of mm-spaces.*

By standard compactness arguments, one can prove that the infimum above is always attained [3]. Let  $\mathcal{U}_p^{\text{opt}}(X, Y)$  denote the set of all the couplings in  $\mathcal{U}(\mu_X, \mu_Y)$  that achieve the minimum. The structure of the former set depends not only on  $\mu_X$  and  $\mu_Y$ , but also on  $d_X, d_Y$  and  $p$ .

**Example 5.** Consider the mm-space with exactly one point:  $(\{*\}, (0), \delta_*)$ . Then,

$$d_{G\mathcal{W},p}(X, \{*\}) = \frac{1}{2} \left( \iint (d_X(x, x'))^p \mu_X(dx) \mu_X(dx') \right)^{1/p}$$

and we define  $\text{diam}_p(X)$ —the  $p$ -statistical diameter of  $X$ —as twice the right-hand side. Notice that  $\lim_{p \rightarrow \infty} \text{diam}_p(X)$  is equal to the usual diameter of  $X$  (as a metric space).

**Question 1.** *To what extent are we able to replicate the nice properties of  $(\mathcal{M}, d_{G\mathcal{H}})$  in the context of  $(\mathcal{M}^w, d_{G\mathcal{W},p})$ ? In particular, it is of interest to investigate whether this new space of datasets is complete and whether one can easily identify rich pre-compact classes.*

### 3.1. Pre-Compactness

**Theorem 4** ([3]). *For a function non-decreasing  $\rho : [0, \infty) \rightarrow [0, 1]$ , such that  $\rho(\varepsilon) > 0$  for  $\varepsilon > 0$  and  $D > 0$ , let  $\mathcal{F}^w(\rho, D) \subset \mathcal{M}^w$  denote the set of all mm-spaces  $X$ , such that  $\text{diam}(X) \leq D$  and  $\inf_x \mu_X(B_\varepsilon(x)) \geq \rho(\varepsilon)$  for all  $\varepsilon > 0$ . Then,  $\mathcal{F}^w(\rho, D)$  is pre-compact for the Gromov–Wasserstein topology, for any  $p \geq 1$ .*

**Remark 2.** Recall Example 2, where closed  $n$ -dimensional Riemannian manifolds were regarded as metric spaces. One can, all the same, embed closed Riemannian manifolds into  $\mathcal{M}^w$  via  $(X, g_X) \mapsto (X, d_X, \mu_X)$ , where  $d_X$  is the geodesic distance induced by the metric tensor  $g_X$  and  $\mu_X$  stands for the normalized volume measure on  $X$ . It is well known [4] that for  $\varepsilon > 0$  small,  $\mu_X(B_\varepsilon(x)) = \frac{c_n}{\text{vol}(X)} \varepsilon^n (1 - \frac{s_X(x)}{6(n+1)} \varepsilon^2 + O(\varepsilon^4))$ , where  $s_X(x)$  is the scalar curvature of  $X$  at  $x$ , and  $\text{vol}(X)$  is the total volume of  $X$ . Thus, a lower bound on  $\mu_X(B_\varepsilon(x))$  plays the role of a proxy for an upper bound on curvature.

### 3.2. Completeness

The Space  $\mathcal{M}^w$  with any  $p$ -Gromov–Wasserstein distance is not complete. Indeed, consider the following family of mm-spaces:  $\Delta_n \in \mathcal{M}^w$ , where  $\Delta_n$  consists of  $n \in \mathbb{N}$  points at distance one from each other, and all with weights  $1/n$ .

**Claim 1.** *For all  $n, m \geq 1$ ,  $d_{G\mathcal{W},p}(\Delta_n, \Delta_m) \leq \frac{1}{2} (n^{-1/p} + m^{-1/p})$ .*

The claim will follow from the following claim and triangle inequality for  $d_{G\mathcal{W},p}$ :

**Claim 2.** *For all  $n, m \geq 1$ ,  $d_{G\mathcal{W},p}(\Delta_n, \Delta_{n \cdot m}) \leq \frac{1}{2} n^{-1/p}$ .*

In order to verify the claim, we denote by  $\{x_1, x_2, \dots, x_n\}$  the points of  $\Delta_n$  and label the points in  $\Delta_{n \cdot m}$  by  $\{y_{11}, \dots, y_{1m}, y_{21}, \dots, y_{2m}, \dots, y_{n1}, \dots, y_{nm}\}$ . Consider the following coupling between  $\mu_n$  and  $\mu_{n \cdot m}$ , the reference measures on  $\Delta_n$  and  $\Delta_{n \cdot m}$ :

$$\mu(x_i, y_{kj}) := \frac{1}{n \cdot m} \delta_{ik}, \text{ for all } i, k \in \{1, \dots, n\} \text{ and } j \in \{1, \dots, m\}$$

It is clear that this defines a valid coupling between  $\mu_n$  and  $\mu_{n \cdot m}$ .

Now, note that

$$\begin{aligned}
 J(\mu) &:= \sum_{i,i'} \sum_{(k,j),(k',j')} |d_{\Delta_n}(x_i, x_{i'}) - d_{\Delta_{n \cdot m}}(y_{kj}, y_{k'j'})|^p \mu(x_i, y_{kj}) \mu(x_{i'}, y_{k'j'}) \\
 &= \frac{1}{(n \cdot m)^2} \sum_{i,i'} \sum_{j,j'} |d_{\Delta_n}(x_i, x_{i'}) - d_{\Delta_{n \cdot m}}(y_{ij}, y_{i'j'})|^p \\
 &= \frac{1}{(n \cdot m)^2} \sum_i \sum_{j,j'} |1 - \delta_{jj'}|^p \\
 &= \frac{m - 1}{n \cdot m} \\
 &\leq n^{-1}
 \end{aligned}$$

Now, by definition,  $d_{\mathcal{GW},p}(\Delta_m, \Delta_{n \cdot m}) \leq \frac{1}{2} (J(\mu))^{1/p}$ , so the claim follows.

Claim 1 indicates that  $\{\Delta_n\}_{n \in \mathbb{N}}$  constitutes a Cauchy sequence in  $\mathcal{M}^w$ . However, a potential limit object for this sequence will have countably infinitely many points at distance one from each other. This space is not compact, thus  $d_{\mathcal{GW},p}$  is not a complete metric.

### 3.3. Other Properties: Geodesics and Alexandrov Curvature

Very recently, Sturm [5] pointed out that  $\mathcal{M}^w$  is a geodesic space when endowed with any  $d_{\mathcal{GW},p}$ ,  $p \geq 1$ . This means that given any two spaces  $X_0, X_1$  in  $\mathcal{M}^w$ , one can find a curve  $[0, 1] \ni t \mapsto X_t \in \mathcal{M}^w$ , such that  $d_{\mathcal{GW},p}(X_t, X_s) = |t - s| d_{\mathcal{GW},p}(X_0, X_1)$ ,  $s, t \in [0, 1]$ .

**Proposition 1 ([5]).** *For each  $p \geq 1$ , the space  $(\mathcal{M}^w, d_{\mathcal{GW},p})$  is geodesic. Furthermore, for  $p > 1$ , the following curves on  $\mathcal{M}^w$  define geodesics between  $(X_0, d_0, \mu_0)$  and  $(X_1, d_1, \mu_1)$  in  $\mathcal{M}^w$ :*

$$[0, 1] \ni t \mapsto (X_0 \times X_1, d_t, \mu)$$

where  $d_t((x_0, x_1), (x'_0, x'_1)) := (1 - t) d_0(x_0, x'_0) + t d_1(x_1, x'_1)$  for  $(x_0, x_1), (x'_0, x'_1) \in X_0 \times X_1$  and  $\mu \in \mathcal{U}_p^{\text{opt}}(X, Y)$ . Furthermore, for  $p > 1$ , all geodesics are of this form.

Sturm further proved that the completion  $\overline{\mathcal{M}^w}$  of the space  $\mathcal{M}^w$  with metric  $d_{\mathcal{GW},2}$  satisfies:

**Theorem 5 ([5]).** *The metric space  $(\overline{\mathcal{M}^w}, d_{\mathcal{GW},2})$  is an Alexandrov space of curvature  $\geq 0$ .*

Amongst the consequences of this property is the fact that one can conceive of gradient flows on the space of all mm-spaces [5].

### 3.4. The Metric $d_{\mathcal{GW},p}$ in Applications

Applications of the notion of Gromov-Wasserstein distance arise in shape and data analysis. In shape analysis, the main application is shape matching under invariances. Many easily computable lower bounds for the GW distance have been discussed in [3,6]. All of them lead to solving linear programming optimization problems (for which there are polynomial time algorithms) or can be computed via elicited formulas. As an example, consider the following invariant of an mm-space  $(X, d_X, \mu_X)$ :

$$H_X : [0, \infty) \rightarrow [0, 1], \quad t \mapsto \mu_X \otimes \mu_X(\{(x, x') \mid d_X(x, x') \leq t\})$$

This invariant simply encodes the distribution of pairwise distances on the dataset  $X$ , and it is defined by analogy with the so-called shape distributions that are well known in computer graphics [7]. Then, one has:

**Proposition 2** ([3,6]). *Let  $X, Y \in \mathcal{M}^w$  be any two mm-spaces and  $p \geq 1$ . Then,*

$$d_{\mathcal{GW},p}(X, Y) \geq \frac{1}{2} \int_0^\infty |H_X(t) - H_Y(t)| dt$$

**Remark 3.** This invariant is also related to the work of Boutin and Kemper [8] and Brinkman and Olver [9].

Other lower bounds which can be computed in time polynomial in the number of points in the underlying mm-spaces have been reported in [3]. As a primary example, one has that the *local shape distributions* of shapes provide a lower bound which is strictly stronger than the ones in the Proposition above. In more detail, consider for a given mm-space  $(X, d_X, \mu_X)$  the invariant:

$$h_X : X \times [0, \infty) \rightarrow [0, 1], (x, t) \mapsto \mu_X(\overline{B_t(x)}).$$

Then, for mm-spaces  $X$  and  $Y$  consider the *cost* function  $c_{X,Y} : X \times Y \rightarrow \mathbb{R}^+$  given by:

$$c_{X,Y}(x, y) := \int_0^\infty |h_X(x, t) - h_Y(y, t)| dt.$$

One then has:

**Proposition 3** ([3,6]). *Let  $X, Y \in \mathcal{M}^w$  be any two mm-spaces and  $p \geq 1$ . Then,*

$$d_{\mathcal{GW},p}(X, Y) \geq \frac{1}{2} \inf_{\mu} \iint c_{X,Y}(x, y) \mu(dx \times dy),$$

where  $\mu$  ranges in  $\mathcal{U}(\mu_X, \mu_Y)$ .

**Remark 4.** Solving for the infimum above leads to a mass transportation problem for which there exists efficient linear programming techniques.

**Remark 5.** It is possible to define a notion of *spectral* Gromov-Wasserstein distance which operates at the level of compact Riemannian manifolds without boundaries, and is based on the comparison of heat-kernels. This notion permits inter-relating many pre-existing shape matching methods and suggests some others [12].

#### 4. Discussion and Outlook

The Gromov–Hausdorff distance offers a useful language for expressing different tasks in shape and data analysis. Its origins are in the work of Gromov on synthetic geometry. For finite metric spaces, the Gromov–Hausdorff distance leads to solving NP-hard combinatorial optimization problems. Related to construction is Gromov–Wasserstein distances that operate on metric measure spaces [3,10]. In contrast to the Gromov–Hausdorff distance, the computation of Gromov–Wasserstein distances leads to solving quadratic optimization problems on continuous variables. The space of all metric measures

spaces endowed with a certain variant of the Gromov–Wasserstein distance [3] enjoys nice theoretical properties [5]. It seems of interest to develop provably correct approximations to these distances when restricted to some suitable subclasses of finite metric spaces. Other aspects of the Gromov–Wasserstein distance are discussed in [3,5,10–12].

### Conflicts of Interest

The author declares no conflict of interest.

### References

1. Burago, D.; Burago, Y.; Ivanov, S. A Course in Metric Geometry. In *AMS Graduate Studies in Math*; American Mathematical Society: Providence, RI, USA, 2001; Volume 33.
2. Petersen, P. Gromov-Hausdorff convergence of metric spaces. In *Differential Geometry: Riemannian Geometry*, Proceedings of the Symposium in Pure Mathematics, Los Angeles, CA, USA, 8–18 July 1990.
3. Mémoli, F. Gromov-Wasserstein distances and the metric approach to object matching. *Found. Comput. Math.* **2011**, *11*, 417–487.
4. Sakai, T. Riemannian geometry. In *Translations of Mathematical Monographs*; American Mathematical Society: Providence, RI, USA, 1996; Volume 149.
5. Sturm, K.-T. The space of spaces: Curvature bounds and gradient flows on the space of metric measure spaces. *Mathematics* **2012**, arXiv:1208.0434.
6. Mémoli, F. On the use of Gromov-Hausdorff distances for shape comparison. In *Proceedings of the Point Based Graphics 2007*, Prague, Czech Republic, 2–3 September 2007.
7. Osada, R.; Funkhouser, T.; Chazelle, B.; Dobkin, D. Shape distributions. *ACM Trans. Graph.* **2002**, *21*, 807–832.
8. Boutin, M.; Kemper, G. On reconstructing  $n$ -point configurations from the distribution of distances or areas. *Adv. in Appl. Math.* **2004**, *32*, 709–735.
9. Brinkman, D.; Olver, P.J. Invariant histograms. *Am. Math. Mon.* **2012**, *119*, 4–24.
10. Sturm, K.-T. On the geometry of metric measure spaces. I. *Acta Math.* **2006**, *196*, 65–131.
11. Mémoli, F. Gromov-Hausdorff distances in Euclidean spaces. In *Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2008*, Anchorage, AK, USA, 23–28 June 2008.
12. Mémoli, F. A spectral notion of Gromov-Wasserstein distances and related methods. *Appl. Comput. Harmon. Anal.* **2011**, *30*, 363–401.