



The Gromov-Wasserstein distance
and
distributional invariants of datasets

Facundo Memoli

facundo.memoli@gmail.com

How to **compare** datasets?

How **different** are two given datasets?

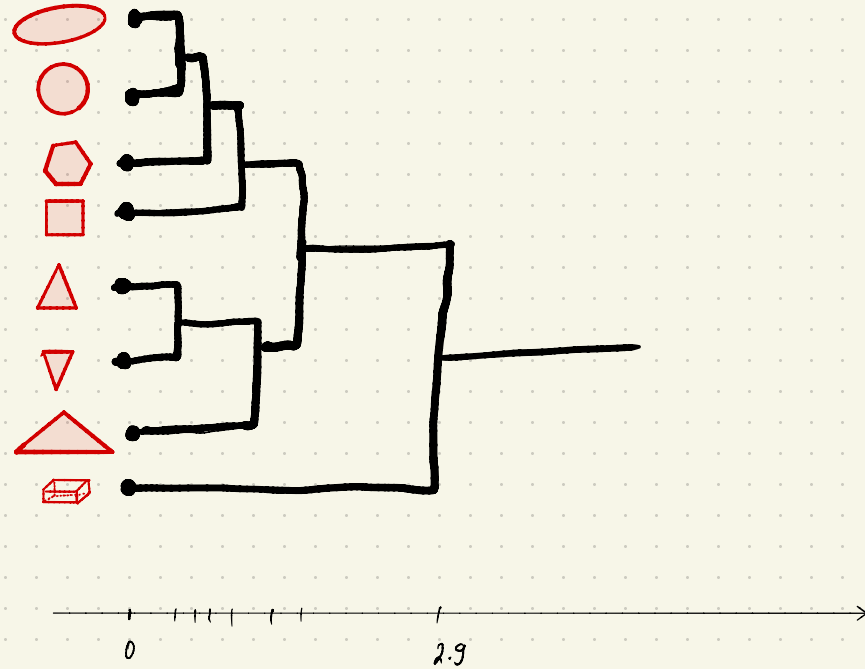
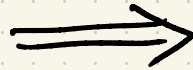
$i \text{ dist}(\triangle, \circ) ?$ $j \text{ dist}(\square, \text{box}) ?$ $i \text{ dist}(\circ, \square) ?$

Applications are clear:

- Clustering
- Classification
- visualization (via cMDS)

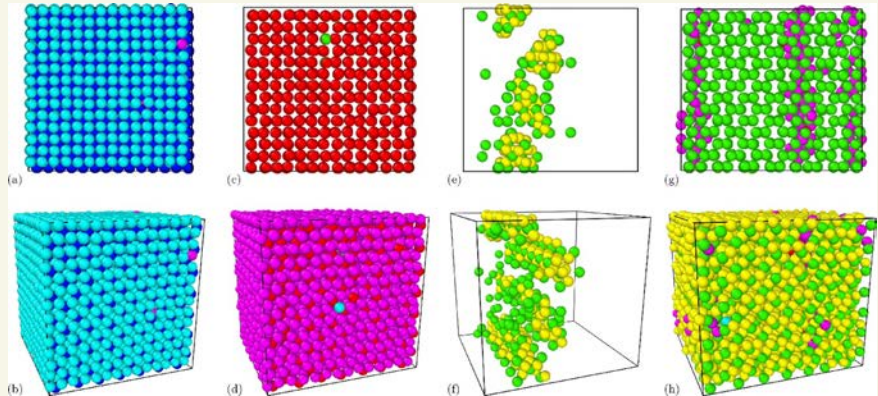
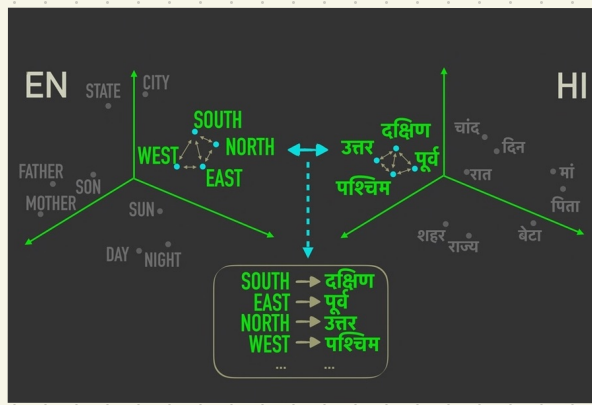
HIERARCHICAL CLUSTERING

	□	◁	○	◌	◡	▤	▷	△
□	0	1.2	1.5	2	0.7	3.1	1.1	2.1
△		0	1.7	2.1	1.9	3.2	0.4	0.5
○			0	0.5	0.2	2.9	1.8	2.1
◌				0	0.6	3.4	2.05	1.1
◡					0	3.1	1.3	0.9
▤						0	3.2	3.4
▷							0	1.2
△								0



Applications

- Shape modeling
(Peyre et al)
- Language translation
(Alvarez-Mellis et al)
- Chemistry.
(Kawano - Mason)
- Metagenomics
(multi-omics)
(Demetci et al
Blumberg et al)

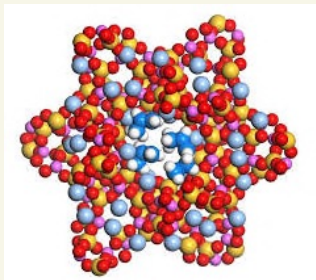


A more basic question:

What is a **dataset** ?

One initial

idea: a dataset is a point cloud in \mathbb{R}^n
modulo rigid transformations



(Standard in Chemistry)



• a dataset is a metric space

• two datasets are considered to be the same iff they are isometric

→ We'll use an enrichment of this representation of datasets.

..... a dataset is a
metric measure space. (m.m. space, for short)

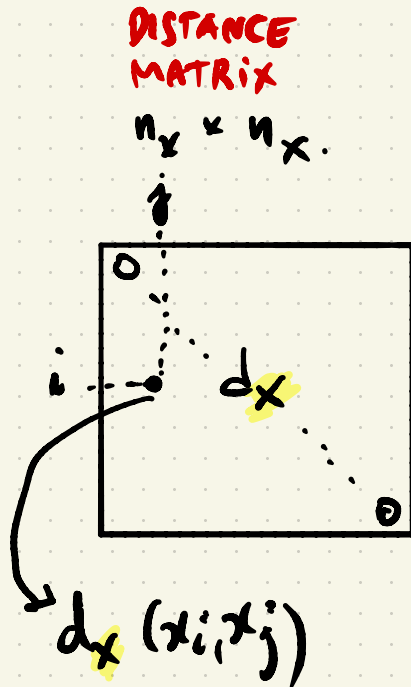
A triple: $\mathcal{X} = (X, d_X, \mu_X)$ where:

- (X, d_X) compact metric space
- μ_X fully supported Borel probability measure on X

\mathcal{M}^w : collection of all mm-spaces.

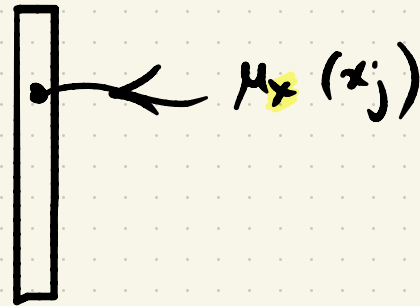
The discrete setting

In the discrete world, $X \in \mathcal{M}^W$ is represented as



WEIGHT VECTOR

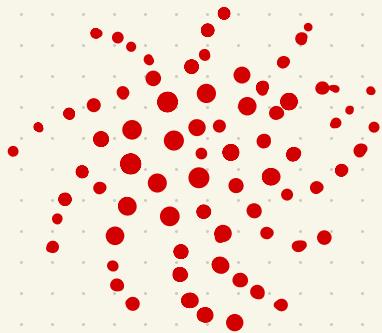
n_x



μ_x

Examples of mm-spaces.

(1) Any point cloud $X \subset \mathbb{R}^d$



$$\mathcal{X} = (X, \|\cdot\|, \underbrace{\left\{ \mu_X(x) = \frac{1}{n} \ \forall x \right\}}_{\text{uniform probability measure}})$$

uniform probability measure.

(2) $X \subset \mathbb{R}^d$ compact \Rightarrow
with $\text{Leb}(X) > 0$

(Lebesgue measure)

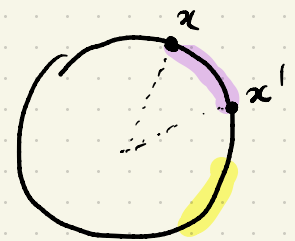
$$\mathcal{X} = (X, \|\cdot\|, \mu_X)$$

normalized
Lebesgue measure



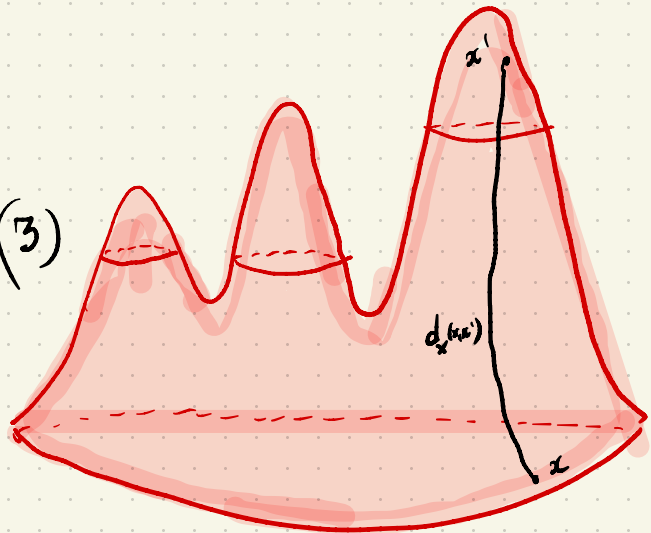
$$(2) \quad X = S^1 \subset \mathbb{R}^2$$

(The circle)



$$\mathcal{X} = (S^1, d_{S^1}, \frac{\text{length}(\cdot)}{2\pi})$$

(3)



$X =$ Riemannian mfd w/ metric tensor g_x

$\rightarrow d_x =$ geodesic distance

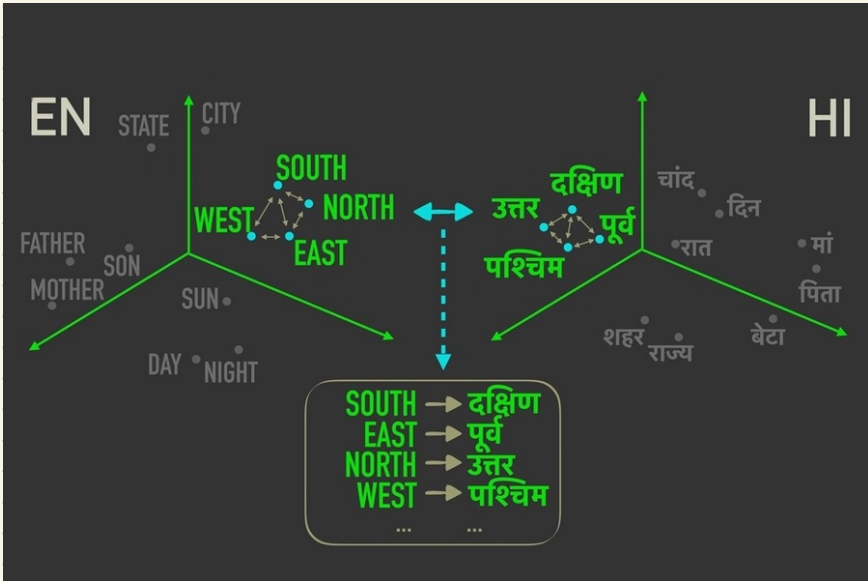
$$\rightarrow \mu_x = \frac{\text{Vol}_x}{\text{Vol}_x(X)}$$

(normalized volume)

(4)

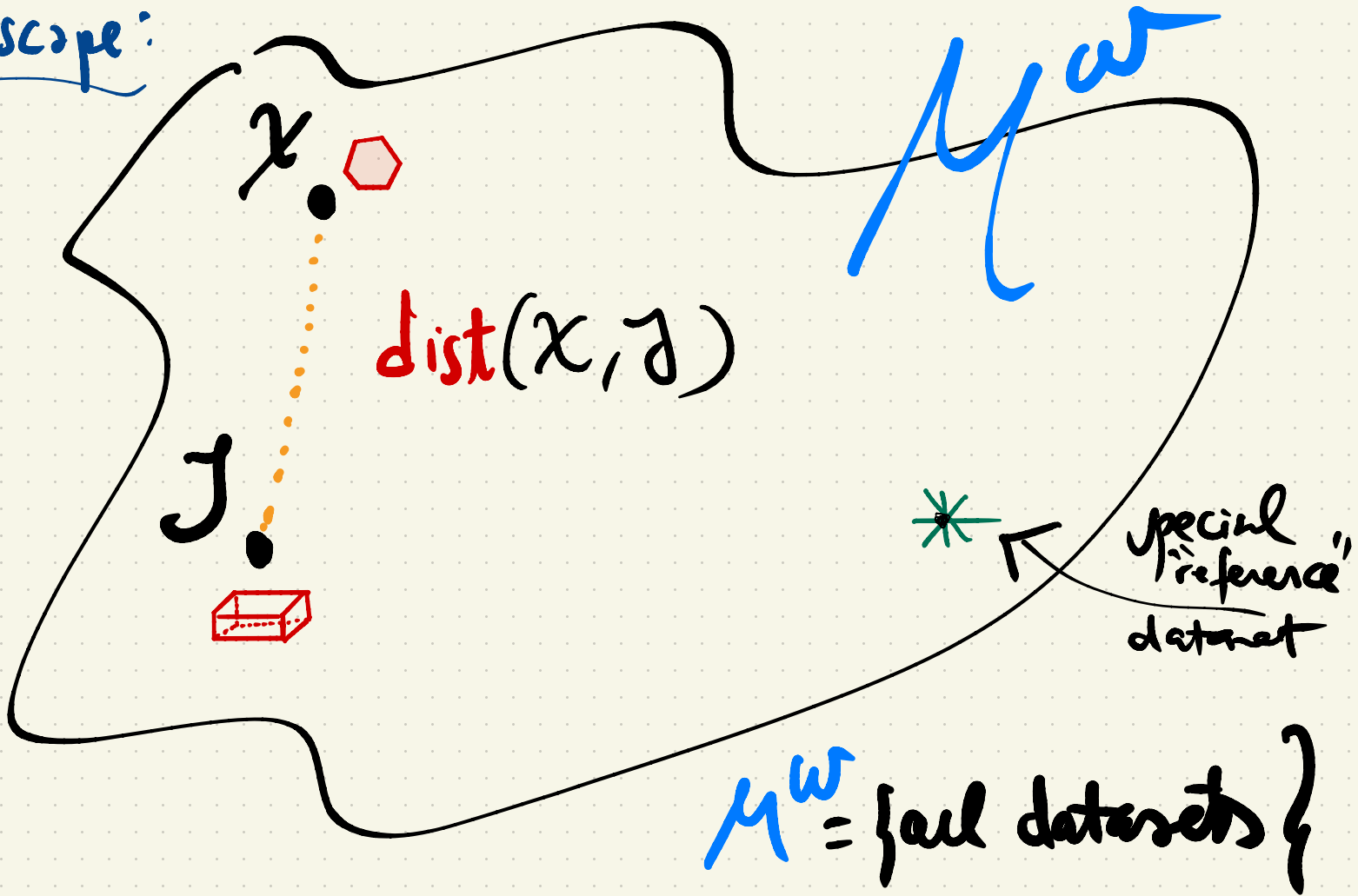
A "language"

$X = \text{lexicon}$
 $d_x \equiv \text{strength of semantic relationship}$
 $\mu_x \equiv \text{relative frequency of word.}$



Automatic language translation
via alignment of
"word embedding spaces"

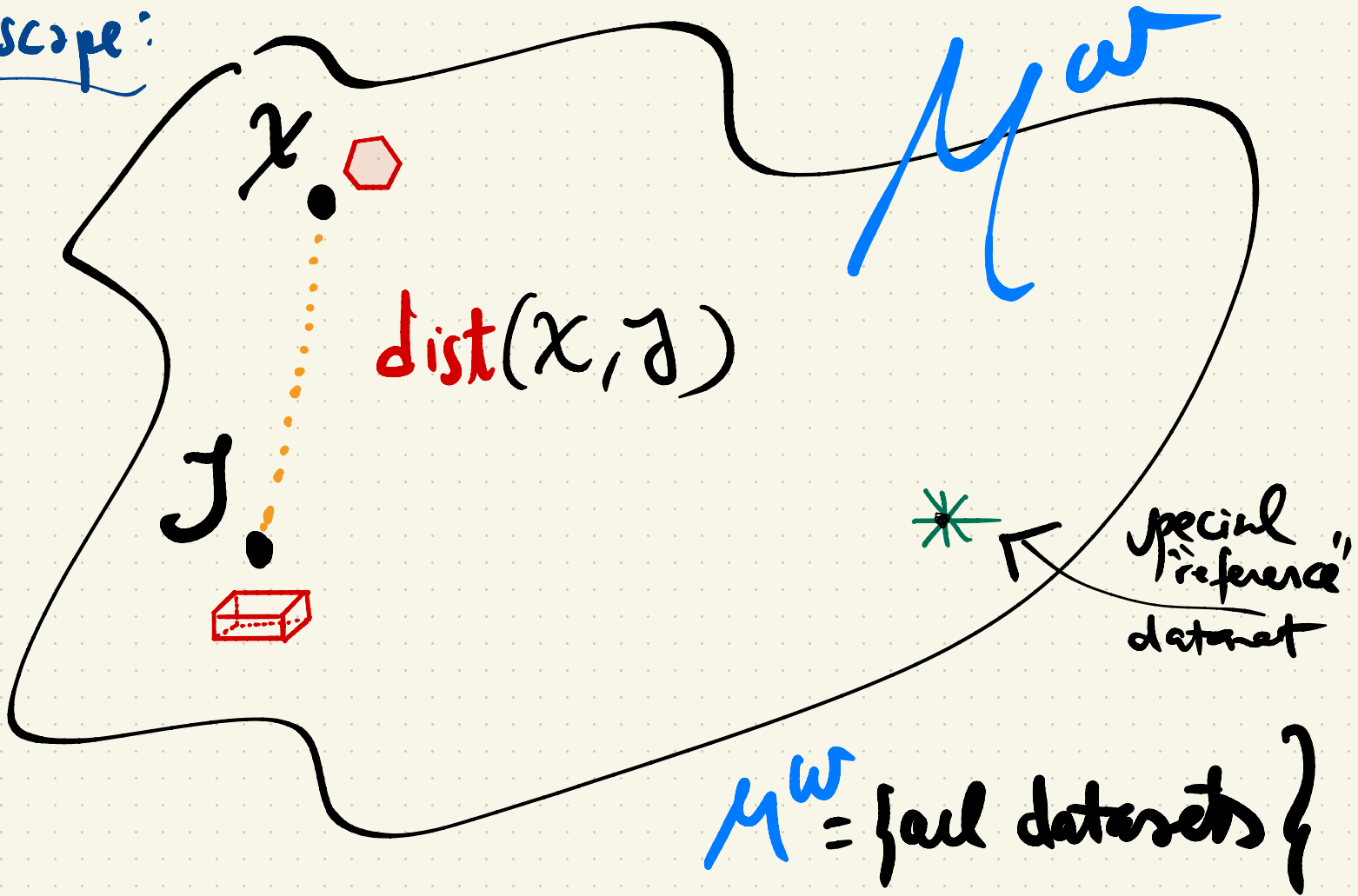
Landscape:



$*$ = $(*, (0), \delta_*)$; the one point dataset

- Serves as "reference" point (like $0 \in \mathbb{R}$)
- distance to $*$ should reflect δ_{z^*} (like $|x-0| = |x|$)
 $x \in \mathbb{R}$)

Landscape:



goal : Construct / define **dist**
on M^u

But before that we need to declare
Equality of datasets

$$\cong : M^u \times M^u \rightarrow \{0, 1\}$$

non-isomorphic isomorphic

Def Two norm. spaces X & Y are isomorphic, denoted $X \cong Y$

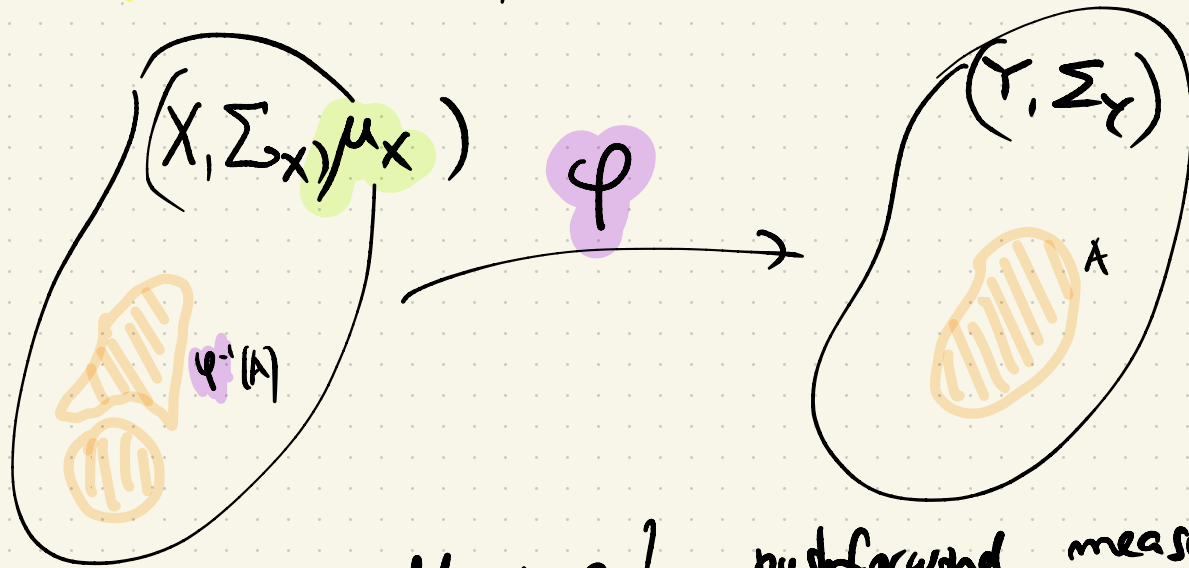
$\exists \Psi: X \rightarrow Y$ isometry. s.t.

\Leftrightarrow

$$\Psi_{\#} \mu_X = \mu_Y$$

(measure preserving isometry)

: the pushforward



φ : measurable map $\left\{ \begin{array}{l} \Rightarrow \text{pushforward measure} \\ \varphi_{\#} \mu_X \text{ is measure on } Y \end{array} \right.$
 μ_X : measure on X defined by: for $A \in \Sigma_Y$

$$(\varphi_{\#} \mu_X)(A) := \mu_X(\varphi^{-1}(A))$$

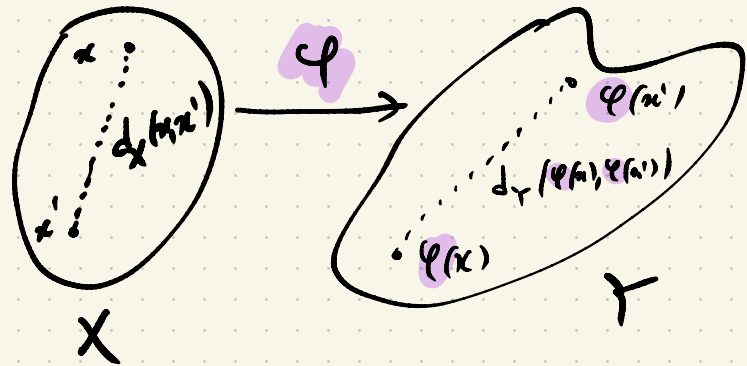
What is an isometry?

Let (X, d_X) , (Y, d_Y) be metric spaces.

A map $\varphi: X \rightarrow Y$ is an **isometry** between X & Y iff:

1. φ is distance preserving: $\forall x, x' \in X \quad d_X(x, x') = d_Y(\varphi(x), \varphi(x'))$

2. φ is surjective.



Def Two mm. spaces X & Y are isomorphic, denoted $X \cong Y$

$\Leftrightarrow \exists \Psi: X \rightarrow Y$ isometry s.t.

$$\Psi_{\#} \mu_X = \mu_Y$$

(measure preserving isometry)

Non-example



X



Y

$\Rightarrow X \not\cong Y$

(no isometry respects the weights)

The construction of

$$\text{dist} : M^w \times M^w \longrightarrow \mathbb{R}_+$$

$$(x, y) \longmapsto \text{dist}(x, y)$$

$$\left(\text{st } x \cong y \Leftrightarrow \text{dist}(x, y) = 0 \right)$$

Main idea: to relate x with y we
"soft maps"

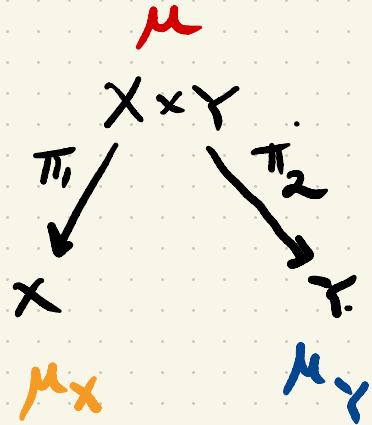
⇓
(stochastic)

Def

Given $X, Y \in \mathcal{M}^{\omega}$, a coupling between X and Y is any μ , probability measure on $X \times Y$ st its marginals are μ_X & μ_Y :

$$(\pi_1)_\# \mu = \mu_X$$

$$(\pi_2)_\# \mu = \mu_Y$$





In probabilistic
jargon μ is a
"joint" distribution
between μ_X & μ_Y

Fact: (1) You can always find at least one coupling:

$\mu = \mu_X \otimes \mu_Y$, the product measure.

(2) When $Y = \{*\}$ $\Rightarrow \mu = \mu_X \otimes \delta_*$ is the unique choice.

(3) If $\Psi: X \rightarrow Y$ is an isomorphism $\Rightarrow \mu_\Psi := (\text{id}_X, \Psi)_\# \mu_X$ is a coupling

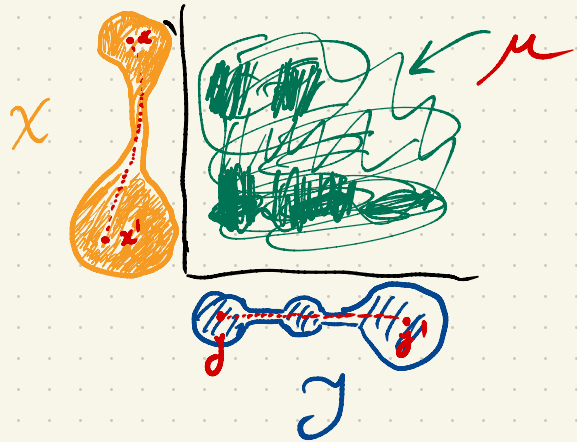
How good is a given μ ?

let $p \geq 1$

Def The p -distortion of μ ,

$\text{dis}_p(\mu) := \left(\begin{array}{l} p\text{-th Average difference} \\ \text{of distances.} \end{array} \right)$

$$= \left[\mathbb{E}_{\mu \otimes \mu} \left(|d_x(x, x') - d_y(y, y')|^p \right) \right]^{1/p}$$



Expanding into a more explicit formula:

$$\text{dis}_p(\mu) = \left[\iint_{X \times Y} \iint_{X \times Y} |d_x(x, x') - d_Y(y, y')|^p \mu(dx \times dy) \mu(dx' \times dy') \right]^{1/p}$$

For later reference, in the finite setting:

$$\text{dis}_p(\mu) = \left[\sum_{i,j,k,l} |d_x(x_i, x_k) - d_Y(y_j, y_l)|^p \mu_{ij} \mu_{kl} \right]^{1/p}$$

Def. The p -th Gromov - Wasserstein distance between X & Y is defined by

$$d_{GW,p}(X,Y) := \frac{1}{2} \min_{\mu \text{ coupling}} \text{dis}_p(\mu)$$

i.e.: one wants to find the best coupling

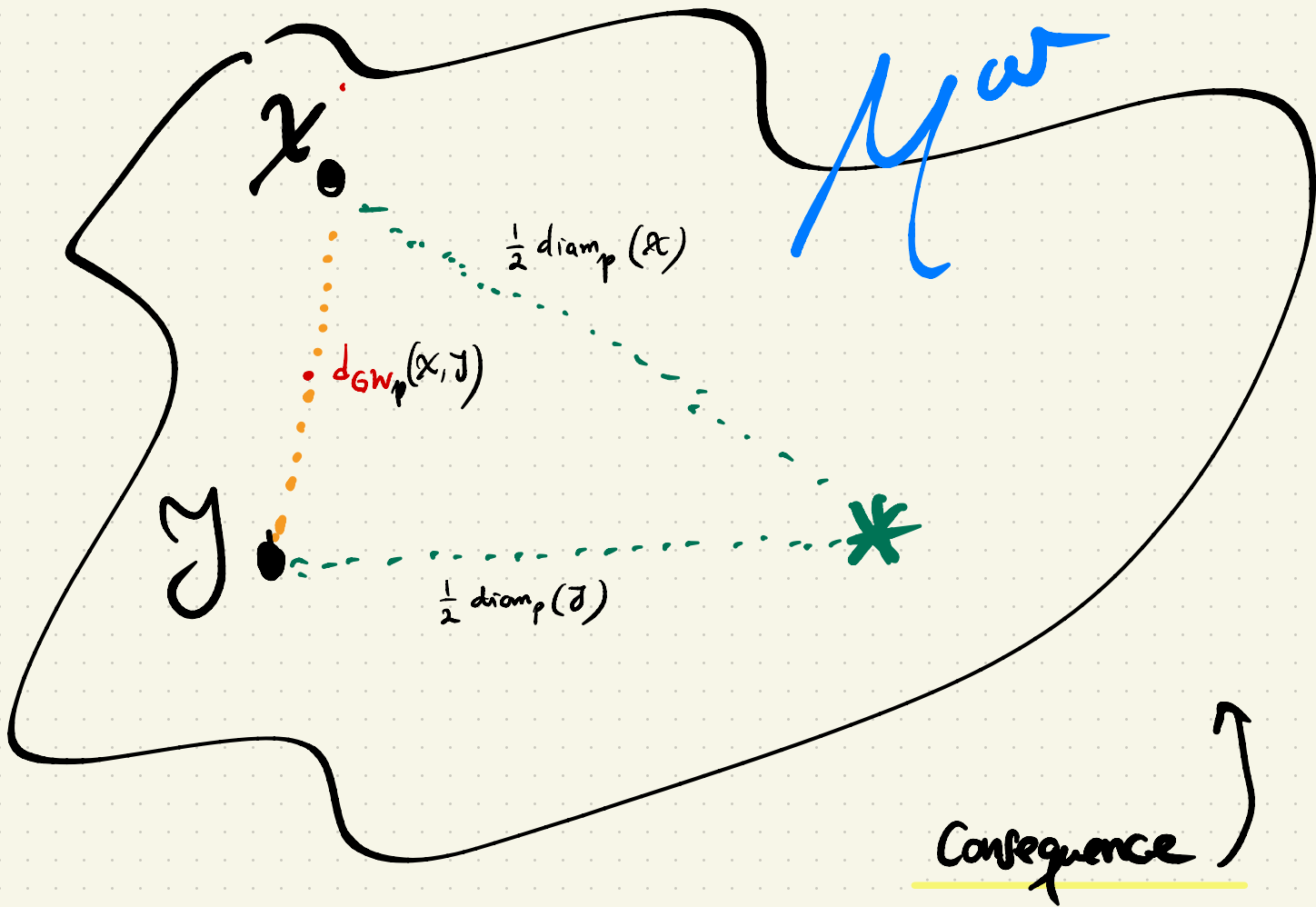
→ our construction of dist is $d_{GW,p}$!!

Example $2 \cdot d_{GW, p}(\mathcal{X}, *) = \left[\iint (d_{\mathcal{X}}(x, x'))^p \mu_{\mathcal{X}}(dx) \mu_{\mathcal{X}}(dx') \right]^{1/p}$
 $=: \text{diam}_p(\mathcal{X})$

$(*, (\cdot|\cdot), \delta_*)$

[the p -diameter of \mathcal{X}]

This is because we have unique coupling $\mu_{\mathcal{X}} \otimes \delta_*$
 between $\mu_{\mathcal{X}}$ & δ_*



Now we have functions:

- $\cong: M^w \times M^w \rightarrow \{0,1\}$

&

- $d_{GWIP}: M^w \times M^w \rightarrow \mathbb{R}_+$

How are they related ?

Is it true that $d_{GWIP}(x,y)=0 \Leftrightarrow x \cong y$?

Main Theorem (Mémoli 2008, Sturm 2012)

For every $p \geq 1$,

$d_{sw,p}$ is a legitimate distance on $\underline{M^w} \cong \mathbb{R}$:

(1) it is symmetric

(2) $d_{sw,p}(x, y) = 0 \iff x \cong y$

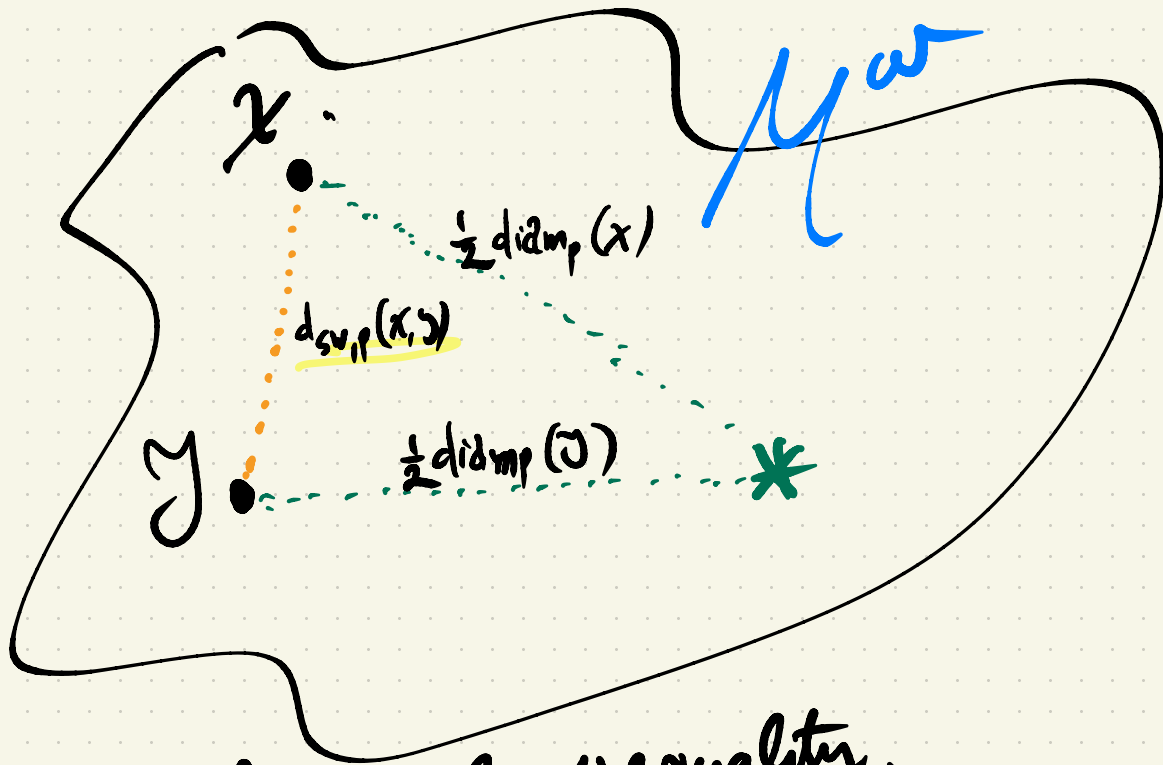
(3) It satisfies the triangle inequality

(3') $(M^w, d_{sw,p})$ is NOT complete.
Furthermore,

(4) It is an intrinsic/geodesic distance.

(5) $(\underline{M^w}, d_{sw,2})$ is Alexandrov with $\text{Cur} \geq 0$.

Example

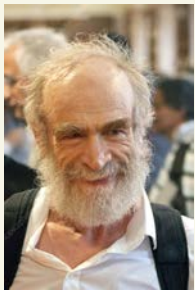


By the triangle inequality,

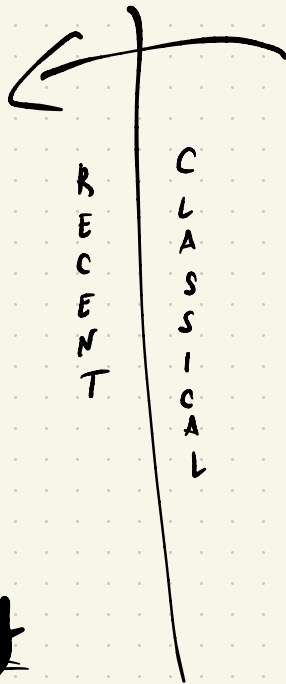
$$\frac{1}{2} |\text{diam}_p(x) - \text{diam}_p(y)| \leq d_{\text{sup}_p}(x, y) \leq \frac{1}{2} (\text{diam}_p(x) + \text{diam}_p(y))$$

A historical Note:

M. Gromov



Metric Geometry



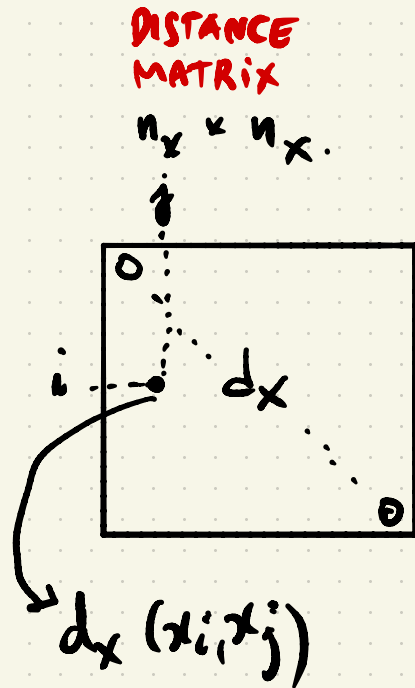
D. Wasserstein
L. Kantorovich
G. Monge.

Optimal Transport

(The Gromov-Wasserstein distance is a generalization of the so called Gromov-Hausdorff distance, a notion which is useful in Metric/Differential/Riemannian Geometry)

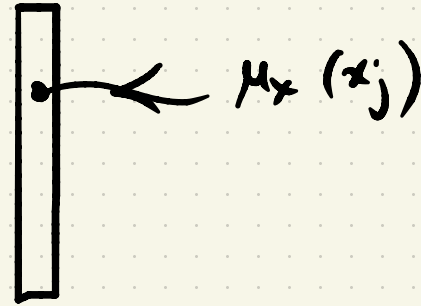
How DO WE COMPUTE $d_{G,1}$?

In the discrete world. $X \in \mathcal{M}^W$ is represented as



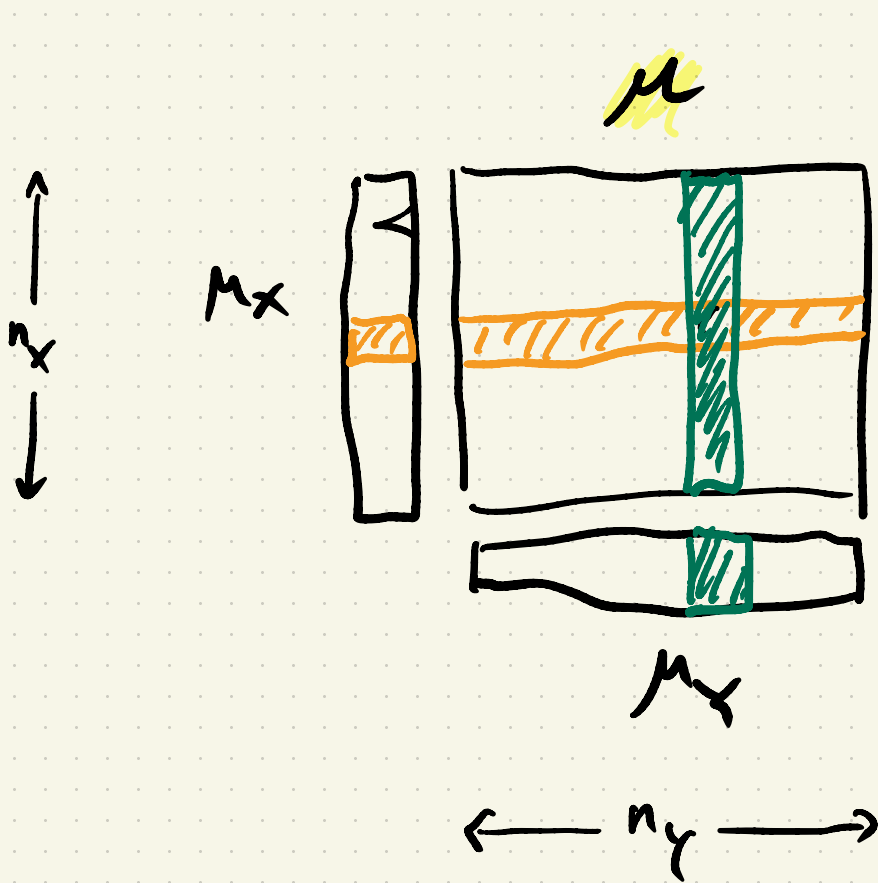
WEIGHT VECTOR

n_x



w_x

Given x, y , finite, a coupling μ is a matrix:



$$\mu_{ij} \geq 0$$

$$\sum_j \mu_{ij} = \mu_x(i) \forall i$$

$$\sum_i \mu_{ij} = \mu_y(j) \forall j$$

Linearly
Constrained

Say $p=1$ for simplicity.

$$\underline{\underline{\text{dis}_1(\mu)}} = \sum_{i,k \ell} |d_X(x_i, x_k) - d_Y(y_i, y_\ell)| \mu_{ij} \mu_{k\ell}$$

Γ_{ijkl}

$$= \sum_{ijkl} \Gamma_{ijkl} \mu_{ij} \mu_{k\ell} = \underline{\underline{U^T \Gamma U}} \quad \text{bilinear form}$$

$$\Rightarrow d_{\text{GW},1}(X,Y) = \frac{1}{2} \min_U \underline{\underline{U^T \Gamma U}} \quad \leftarrow \text{Quadratic functional}$$

↳ linearly constrained

but Γ need NOT be PSD in general

..... \Rightarrow not easy to solve exactly \Rightarrow but have gradient descent!

A number of computational techniques & implementations have been proposed ;

- "Alternate" optimization
- Entropy regularization (Cuturi & Peyré)
- POT (Python OT project)
- See :

Computational Optimal Transport: With Applications to Data Science (Foundations and Trends(r) in Machine Learning)
by Gabriel Peyré (Author), Marco Cuturi (Author)
★★★★☆ 2 ratings

Paperback Other Sellers
\$98.98 - \$99.00 from

Buy new:
\$99.00

Only 4 left in stock (more on the way).

Ships from and sold by Amazon.com.
May be available at a lower price from other sellers, potentially without free Prime shipping.

prime
FREE delivery Wednesday, February 23. Order within 15 hrs 18 mins

Deliver to Faundo - Columbus 43214
Qty: 1

Add to Cart
Buy Now

Any way, given the hardness, it makes sense to look for:

LOWER BOUNDS.

GOAL:

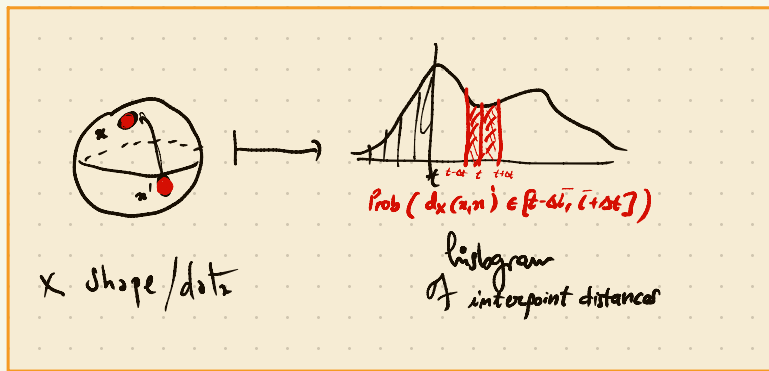
$$\underbrace{\text{down}_P(x, y)}_{\text{difficult}} \geq \underbrace{\text{LB}(x, y)}_{\text{easier}}$$

A simple idea: Global distribution of distances

$$\mathcal{X} = (X, d_X, \mu_X)$$

Fairly classical idea
Popular in Comp. Chemistry
and Shape Analysis
(Osada et al 2002)

$$(\text{dataset}) \quad \mathcal{X} \longmapsto dH_{\mathcal{X}} \quad (\text{prob. measure on } \mathbb{R})$$



Def
(GDD)

$$dH_x = (d_x)_{\#} \mu_x \otimes \mu_x$$

is the global distribution of distances

H_x : cumulative of the measure dH_x

$$H_x(t) = \mu_x \otimes \mu_x \left(\{ (x, x') \mid d_x(x, x') \leq t \} \right)$$

Proposition ($p=1$)

$$d_{GW,1}(x,y) \geq \frac{1}{2} d_{W,1}^{\mathbb{R}}(\mathbb{H}_x, \mathbb{H}_y) =: \underline{SLB}_1(x,y)$$

(second lower bound)

↑
Wasserstein
distance on \mathbb{R}

Remark: The Wasserstein distance on \mathbb{R} has an explicit formula!

$$SLB_1(x,y) = \frac{1}{2} \int_0^{\infty} |H_x(t) - H_y(t)| dt \quad \leftarrow$$

\Rightarrow easily computable

Question How good is SLB ?

i.e. is it true that

$$\text{SLB}(x, y) = 0 \iff x \cong y \quad ?$$

Note $\text{SLB}(x, y) = 0 \iff dH_x = dH_y$

\implies question is whether

$$dH_x = dH_y \iff x \cong y \quad ?$$

I.E. we want to know how strong is the signature $x \mapsto dH_x$

Much can be said about this question...

In general, answer is NO → Consider:

Example by M. Breda & G. Kemper (2002)

Set all weights as $\frac{1}{4}$ ($M_X = M_Y = \text{"uniform"}$).

distances in X	distances in Y
0, 0, 0, 0	0, 0, 0, 0
2, 2	2, 2
$\sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}$	$\sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}$
$\sqrt{10}, \sqrt{10}, \sqrt{10}, \sqrt{10}$	$\sqrt{10}, \sqrt{10}, \sqrt{10}, \sqrt{10}$
4, 4	4, 4

$H_X = H_Y$

Peter Olver's Conjecture



(Extrinsic distances)

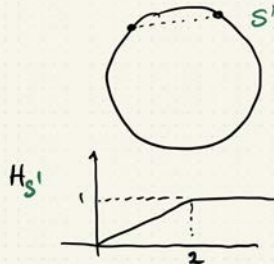
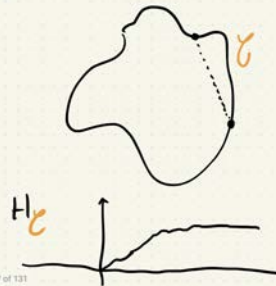
Peter noticed that the curves determined by each set of 4 points, C_X & C_Y , had (via Mottet) $H_{C_X} \neq H_{C_Y}$

Peter Olver's Conjecture

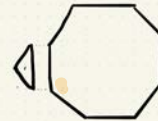
Is it true for planar curves that $H_C = H_{C'} \Leftrightarrow C \cong C'$ (isometric in \mathbb{R}^2)

Some answers (joint with Tom Needham)

(1) if C is s.t. $H_C = H_{\text{circle}} \Rightarrow C \cong \text{circle}$
THEOREM

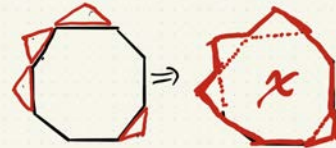


(2) Bad news...

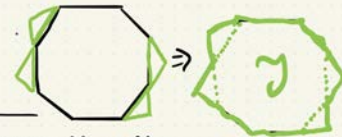


Start with a regular octagon

and an isosceles triangle



H_X



However $H_X = H_Y$

checked by hand (w/ lots of patience!)

Much can be said about this question...

(3) By "smoothing" corners & "rounding" the triangles



\mathcal{H}



Theorem

$\forall \varepsilon > 0 \exists$ curves $\chi_\varepsilon, \gamma_\varepsilon$
within ε of the unit circle
s.t. $\chi_\varepsilon \neq \gamma_\varepsilon$ but
 $H_{\chi_\varepsilon} = H_{\gamma_\varepsilon}$

49 of 131

(4) Similar pattern for embedded surfaces

Proposition $\mathcal{J} \subset \mathbb{R}^3$ embedded closed smooth surface,
Then $H_{\mathcal{J}} = H_{S^2} \iff \mathcal{J} \cong S^2$

(5) This "rigidity" does not extend to arbitrary neighborhoods of S^2



90 of 131

(6) The Riemannian setting \rightsquigarrow Corollary to Chang's rigidity
(cf. Bonnet-Meyer)

Corollary let (M, g_M) be d -dimensional closed Riem. mfd.
with $\text{Ric} \geq d-1$. Then,

1. $\exists \varepsilon = \varepsilon(d) > 0$ s.t. $d_{W,1}^R(dH_M, dH_{S^d}) < \varepsilon$
 $\Rightarrow M$ is diffeo to S^d

2. if $dH_M = dH_{S^d} \Rightarrow M \cong S^d$

(7) H_M contains topological information!

Lemma M closed Riem. mfd.
 $\dim(M) = d$

$$H_M(t) = \frac{\chi(M)}{\text{Vol}(M)} \left(1 - \frac{\int_M \text{scal}_M(p) \text{vol}_M(p)}{\int_M \text{scal}_M(p) \text{vol}_M(p)} t^2 + O(t^4) \right)$$

\swarrow scalar curvature

Corollary when $d=2$

$H_M(t)$ recovers homeomorphism type of M .

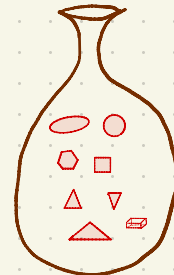
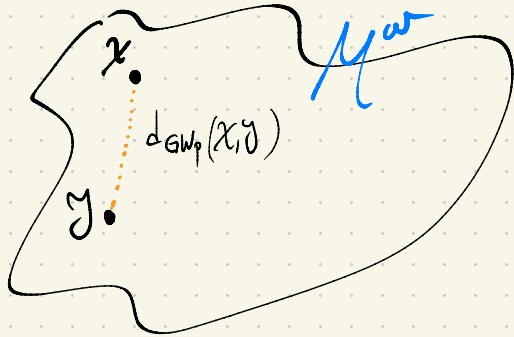
$$\left(\int_M \text{scal}_M(p) \text{vol}_M(p) = C \cdot \chi(M) \right) \quad \text{(via Gauss-Bonnet thm.)}$$

$$\int_{\text{Gauss}} = \chi(M)$$

Discussion

- There are "higher order" distributional invariants (both local & global)
 - ↳ tradeoff between discriminative power & computational cost.
- Connection between GW distance & Weisfeler-Lehman test
 - ↳ applications to GNNs (graph neural networks)
- Instances when (variants of) d_{GW} can be computed/approximated in polynomial time.
- Recent: exact determination of $d_{GW}(S_E^m, S_E^n)$ no benchmarking

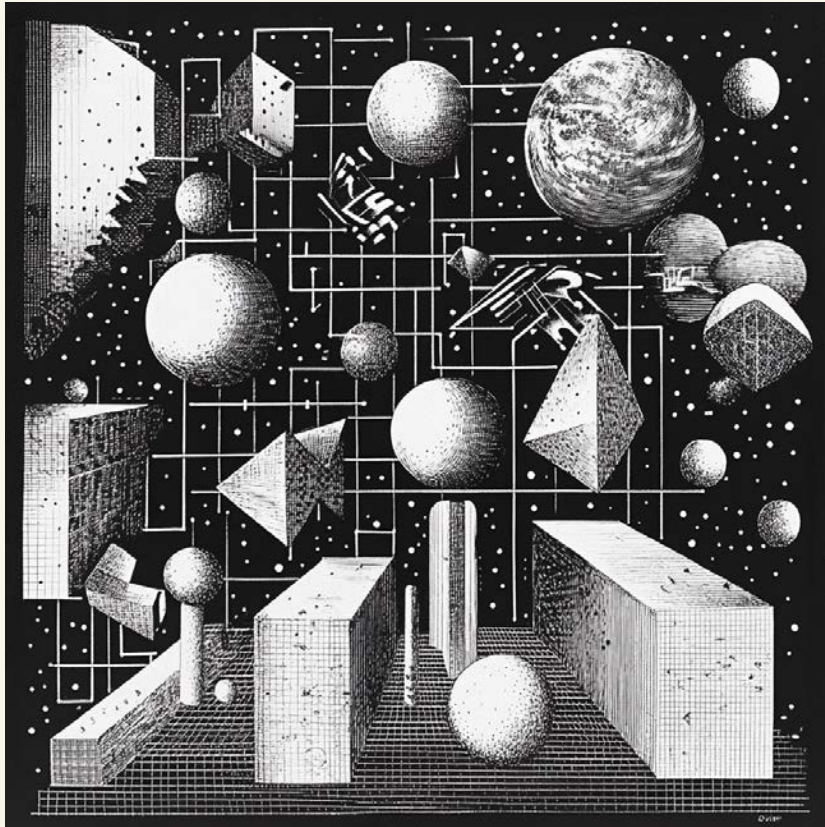
Thank You



dist

DISTANCE MATRIX

	□	△	○	◇	⬢	△		
□	0	1.2	1.5	2	0.7	3.1	1.1	2.1
△	0	4.7	2.1	1.5	3.2	0.4	0.5	
○	0	0.5	0.2	2.3	1.8		2.1	
◇	0	0.6	2.6	2.05	1.1			
⬢	0	3.1	1.3	0.4				
△	0	3.2	3.4					
△								0



References

Published: 30 April 2011

Gromov–Wasserstein Distances and the Metric Approach to Object Matching

Facundo Memoli

Foundations of Computational Mathematics 11, 417–487(2011) | [Cite this article](#)

1617 Accesses | 55 Citations | [Altmetric](#) | [Metrics](#)

Abstract

This paper discusses certain modifications of the ideas concerning the Gromov–Hausdorff distance which have the goal of modeling and tackling the practical problems of object matching and comparison. Objects are viewed as metric measure spaces, and based on ideas from mass transportation, a Gromov–Wasserstein type of distance between objects is defined. This reformulation yields a distance between objects which is more amenable to practical computations but retains all the desirable theoretical properties of this new notion of distance as studied, a strict metric on the collection of isomorphism classes.

AMERICAN MATHEMATICAL SOCIETY BOOKSTORE

MEMOIRS OF THE AMERICAN MATHEMATICAL SOCIETY

The Space of Spaces: Curvature Bounds and Gradient Flows on the Space of Metric Measure Spaces

Karl-Theodor Sturm, University of Bonn, Germany

Softcover ISBN: 978-1-4704-0998-9
 Product Code: MEM02001462
 List Price: \$60.00
 AMS Member Price: \$75.50
 AMS Member Price: \$88.00

Download PDF | eBook | Software culture | See this author

Add to cart

STUDIES IN APPLIED MATHEMATICS

4:48FC16

Probability distributions and inverse problems for metric measure spaces

15

18 August 2022 | <https://doi.org/10.1112/S1046129822000119>

Full text

Abstract

Applications in data science, shape analysis, and object classification frequently require comparison of probability distributions defined on different ambient spaces. To accomplish this, one requires a notion of distance on a given class of metric measure spaces—that is, compact metric spaces endowed with probability measures. Such spaces are typically defined as comparisons between metric measure space invariants, i.e. distance distributions (also referred to as shape distributions, distanceograms, or shape contours in the literature). Generally, distances defined in terms of such distributions are actually pseudometrics, in that they may vanish when comparing nonisomorphic spaces. The goal of this paper is to set up a formal framework assessing the discriminative power of distance distributions, that is, the extent to which these pseudometrics fail to define proper metrics. We formalize several precise problems in terms of these invariants and answer them in several categories of metric measure spaces, including the category of plane curves, where we give a counterexample to the curve histogram conjecture of Benjamini and Ober, the category of embedded and Riemannian manifolds, where we obtain sphere rigidity results, and the category of metric graphs, where we obtain a local injectivity result about the type of

Download:

- PDF
- Other formats

References & Citations

- MathSciNet
- Google Scholar
- Zentralblatt

Export BibTeX Citation

Bookmark

PMLR Proceedings of Machine Learning Research

Volume 162 JMLR MLOSS FAQ Submission Format

[edit]

arXiv.org > math > arXiv:2101.05756

Mathematics > Metric Geometry

(Submitted on 14 Jan 2021 (v1), last revised 7 Jul 2021 (this version, v2))

The ultrametric Gromov-Wasserstein distance

Facundo Memoli, Axel Munk, Zhengchao Wan, Christoph Wethkamp

In this paper, we investigate compact ultrametric measure spaces which form a subset \mathcal{U}^m of the collection of all metric measure spaces \mathcal{M}^1 . Similar as for the ultrametric Gromov-Hausdorff distance on the collection of ultrametric spaces \mathcal{U}^* , we define ultrametric versions of two metrics on \mathcal{U}^m , namely of Sturm's distance of order p and of the Gromov-Wasserstein distance of order p . We study the basic topological and geometric properties of these distances as well as their relation and convergence for $p \rightarrow \infty$ or a polynomial time algorithm for their calculation. Further, several lower bounds for both distances are derived and some of our results are generalized to the case of finite ultra-discontinuity spaces.

Subjects: Metric Geometry (math.MG); Probability and Statistics (math.PR)

Cite as: arXiv:2101.05756 [math.MG]

Weisfeiler-Lehman Meets Gromov-Wasserstein

Samantha Chen, Sunghyuk Lim, Facundo Memoli, Zhengchao Wan, Yusu Wang Proceedings of the 39th International Conference on Machine Learning, PMLR 162:3371-3416, 2022.

Abstract

The Weisfeiler-Lehman (WL) test is a classical procedure for graph isomorphism testing. The WL test has also been widely used both for designing graph kernels and for analyzing graph neural networks. In this paper, we propose the Weisfeiler-Lehman (WL) distance, a notion of distance between labeled measure Markov chains (LMMCs), of which labeled graphs are special cases. The WL distance is polynomial time computable and is also compatible with the WL test in the sense that the former is positive if and only if the WL test can distinguish the two involved graphs. The WL distance captures and compares subtle structures of the underlying LMMCs and, as a consequence of this, it is more discriminating than the distance between graphs used for defining the state-of-the-art Wasserstein Weisfeiler-Lehman graph kernel. Inspired by the structure of the WL distance we identify a neural network architecture on LMMCs which turns out to be universal w.r.t. continuous functions defined on the space of all LMMCs (which includes all graphs) endowed with the WL distance. Finally, the WL distance turns out to be stable w.r.t. a natural variant of the Gromov-Wasserstein (GW) distance for comparing metric Markov chains that we identify. Hence, the WL distance can also be construed as a polynomial time lower bound for the GW distance which is in general NP-hard to compute.

Home > Foundations of Computational Mathematics > Article

The Gromov–Wasserstein Distance Between Spheres

Open access | Published: 16 September 2024
(2024) | [Cite this article](#)

Download PDF | You have full access to this open access article

Shreya Arya, Arnab Auddy, Ranithony A. Clark, Sunghyuk Lim, Facundo Memoli & Daniel Packer

240 Accesses | [Explore all metrics](#)

Foundations of Computational Mathematics

Aims and scope

Submit manuscript

Use our pre-submission checklist

Avoid common mistakes on your manuscript.