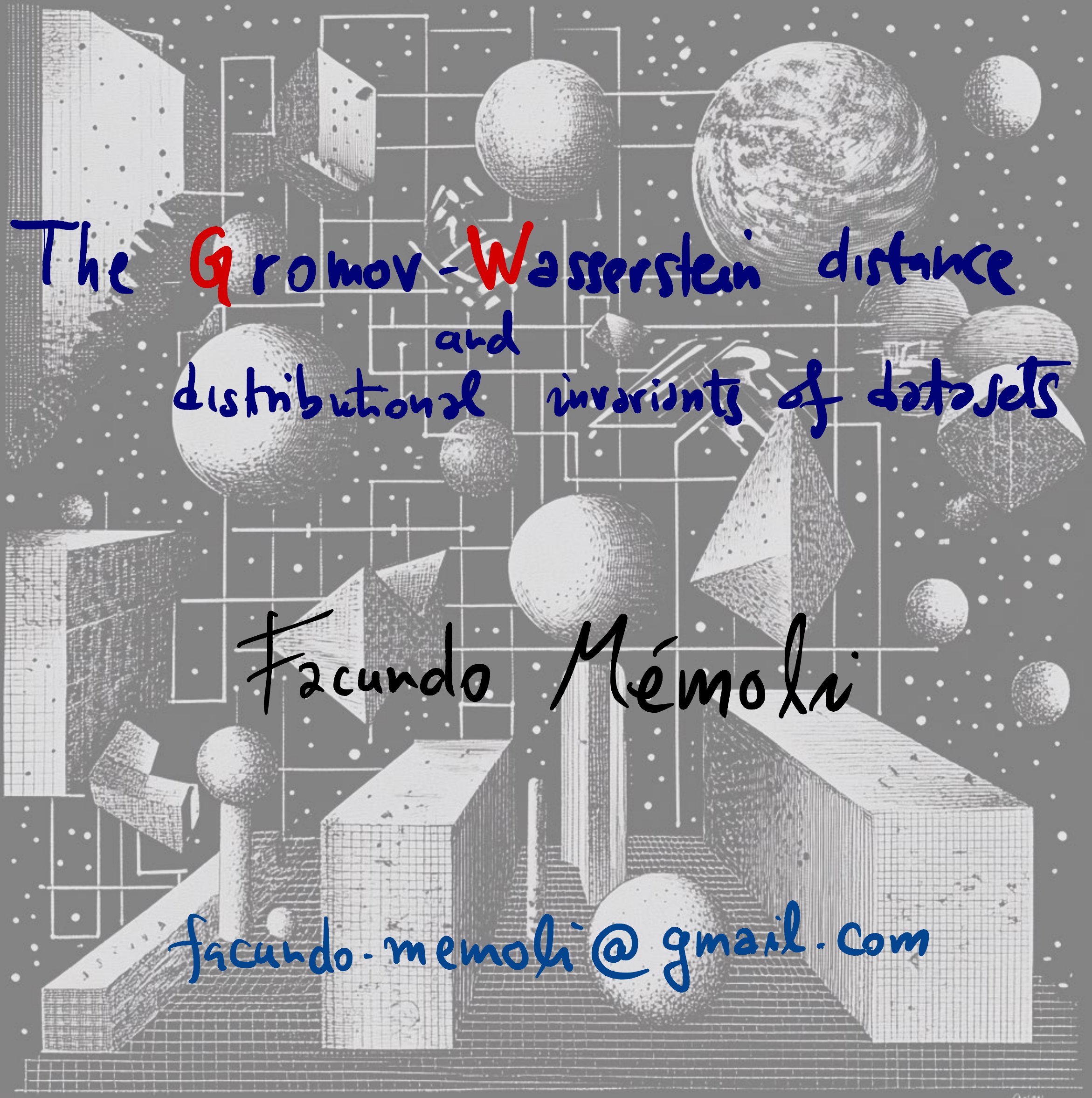


The Gromov-Wasserstein distance and distributional invariants of datasets

Facundo Mémoli

facundo.memoli@gmail.com

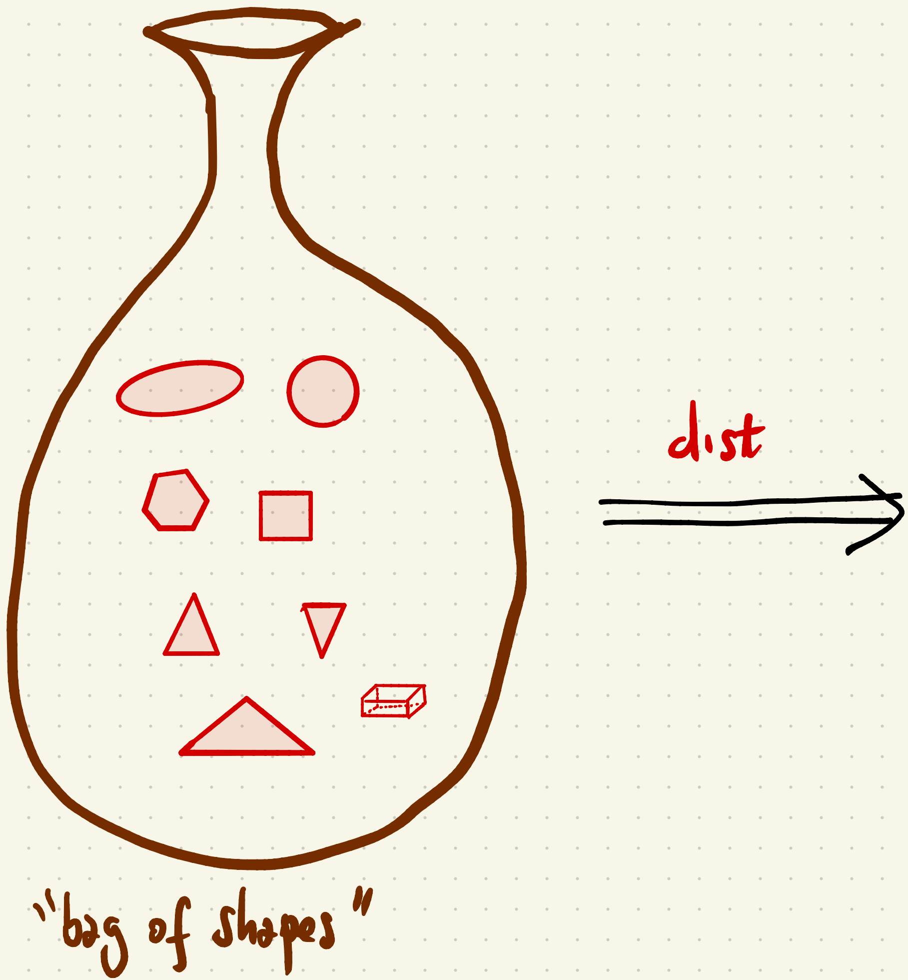
How to compare datasets?

How different are two given datasets?

¿ dist(△, ○) ?   ¿ dist(□, □) ?   ¿ dist(○, □) ?

Applications are clear:

- Clustering
- Classification
- vizualization (via cMDS)



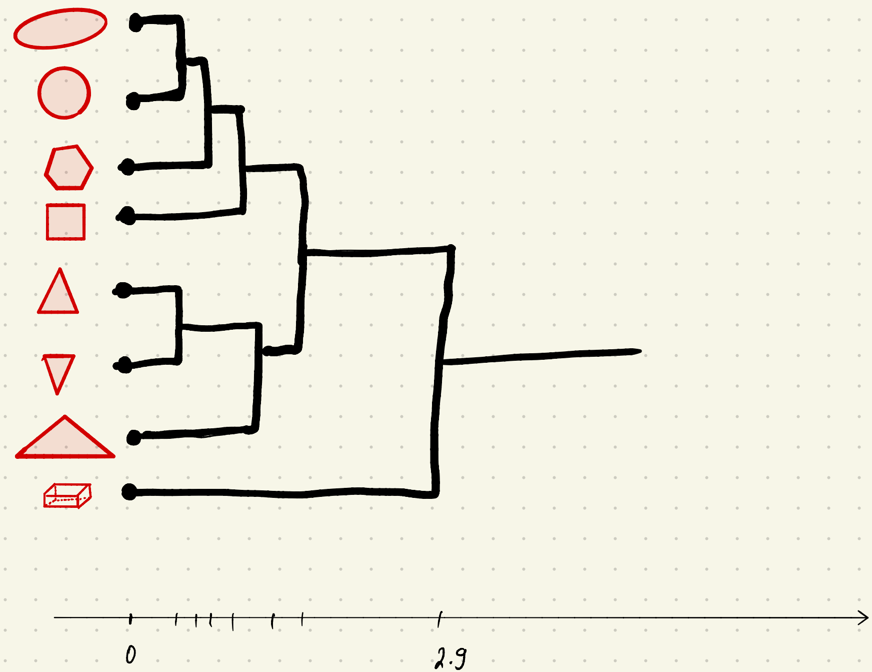
DISTANCE MATRIX

	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	1.9	0.9
						0	3.2	3.4
							0	1.2
								0

HIERARCHICAL CLUSTERING

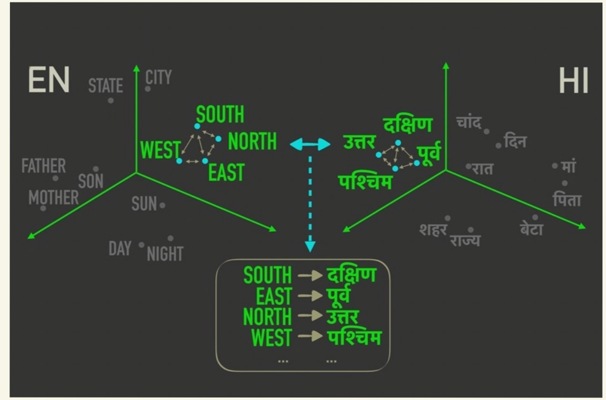
	□	◁	○	⬯	⬡	▭	▷	△
□	0	1.2	1.5	2	0.7	3.1	1.1	2.1
△		0	1.7	2.1	1.9	3.2	0.4	0.5
○			0	0.5	0.2	2.9	1.8	2.1
⬭				0	0.6	3.4	2.05	1.1
⬡					0	3.1	1.9	0.9
▭						0	3.2	3.4
▽							0	1.2
△								0

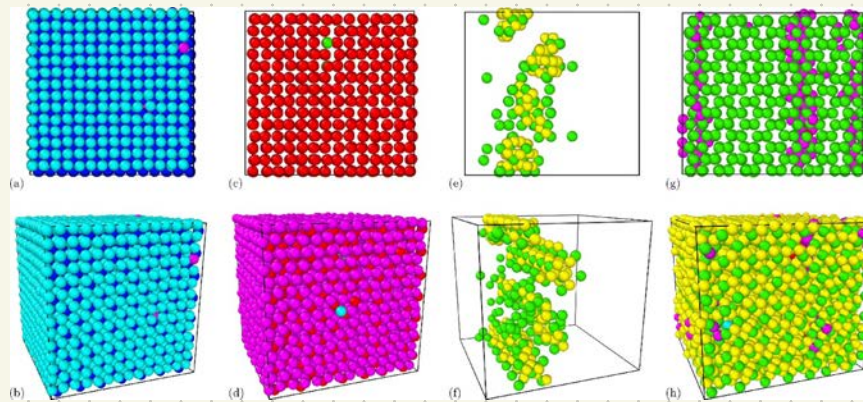
$\Longrightarrow$



Applications

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

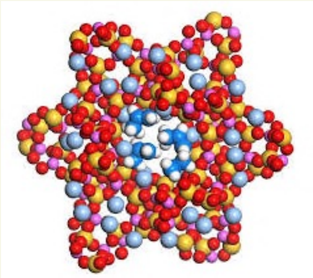




A more basic question:

What is a dataset?

One initial idea: a dataset is a point cloud in $\mathbb{R}^d$ modulo rigid transformations



(Image: Sakai et al)

(Standard in Chemistry)

$\Downarrow$

- a dataset is a metric space
- two datasets are considered to be the same iff they are isometric

$\rightsquigarrow$ We'll use an enrichment of this representation of datasets.

...... a dataset is a

metric measure space. (m.m. space, for short)

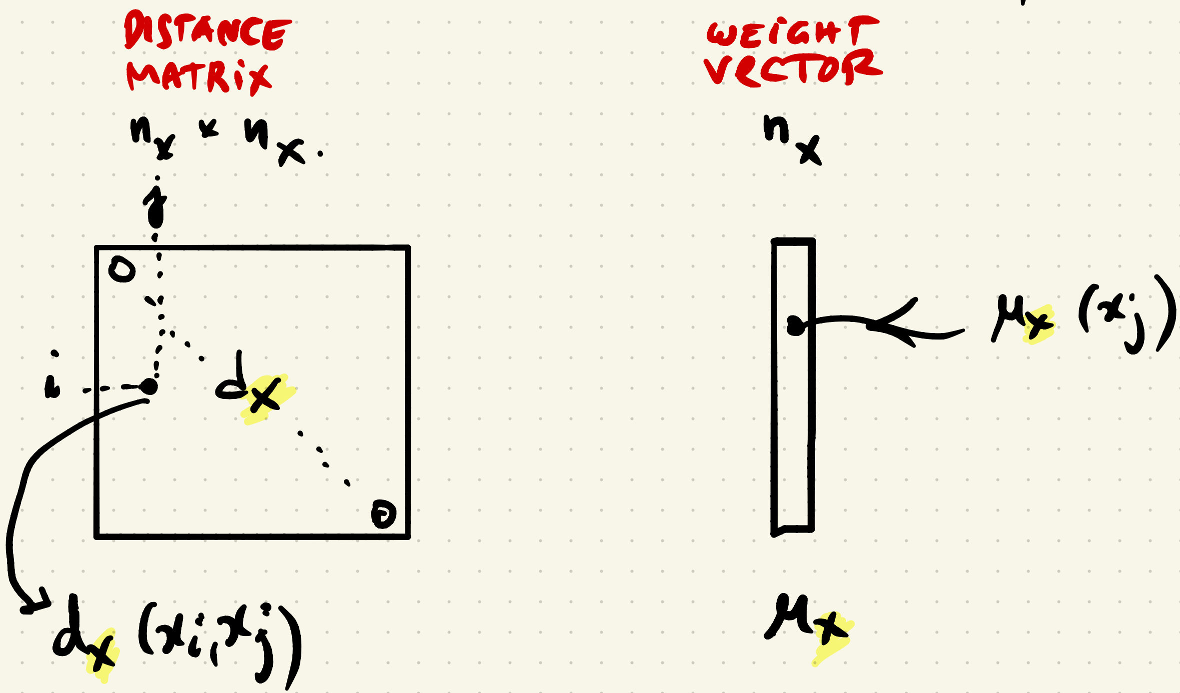
A triple: $\mathcal{X} = (X, d_X, \mu_X)$ where:

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

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

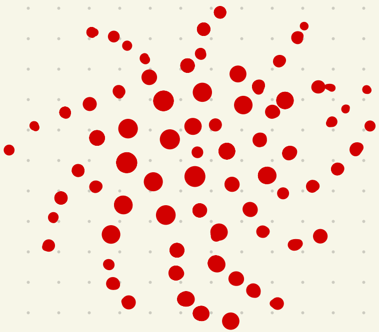
The discrete setting

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



Examples of mm-spaces.

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



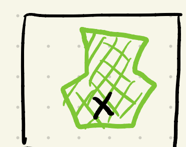
$$\mathcal{X} = \left(X, \|\cdots\|, \underbrace{\left\{\mu_X(x) = \tfrac{1}{n} \ \forall x\right\}}_{\text{uniform probability measure.}}\right)$$

normalized Lebesgue measure

(2) $X \subset \mathbb{R}^d$ compact $\Rightarrow$ $\mathcal{X} = (X, \|\cdots\|, \mu_X)$

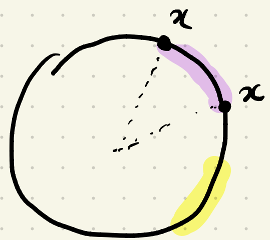
with $\mathrm{Leb}(X) > 0$

(Lebesgue measure)



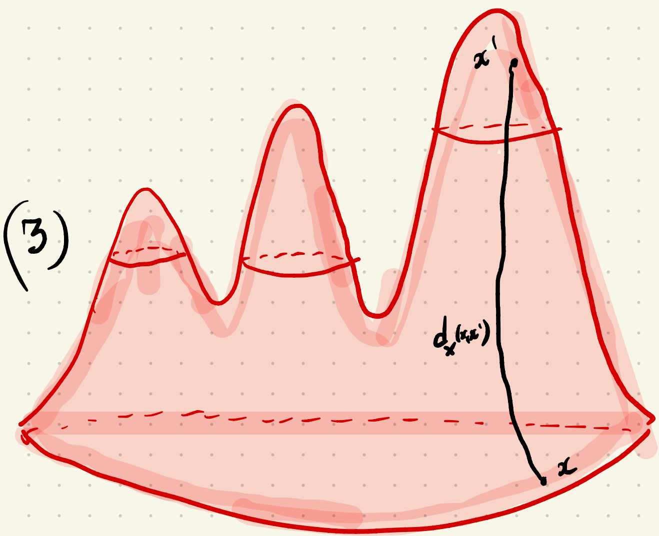
(2) $X = S^1 \subset \mathbb{R}^2$

(the circle)



$\mathcal{X} = \left(S^1, \; d_{S^1}, \; \frac{\text{length}(\ldots)}{2\pi} \right)$

(3)



X = Riemannian mfld w/ metric tensor g_X

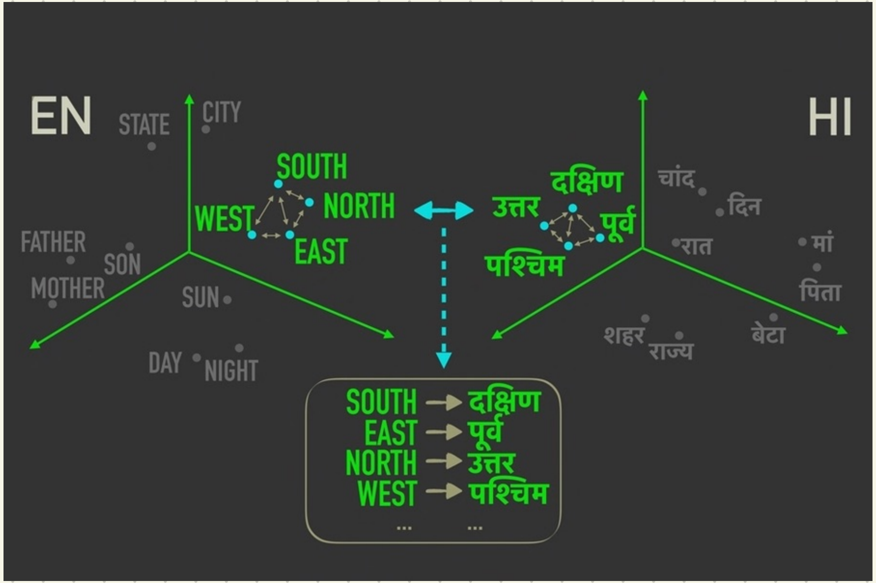
$\longrightarrow$ d_X = geodesic distance

$\longrightarrow$ $\mu_X = \dfrac{vol_X}{vol_X(X)}$

(normalized volume)

(4) A "language"

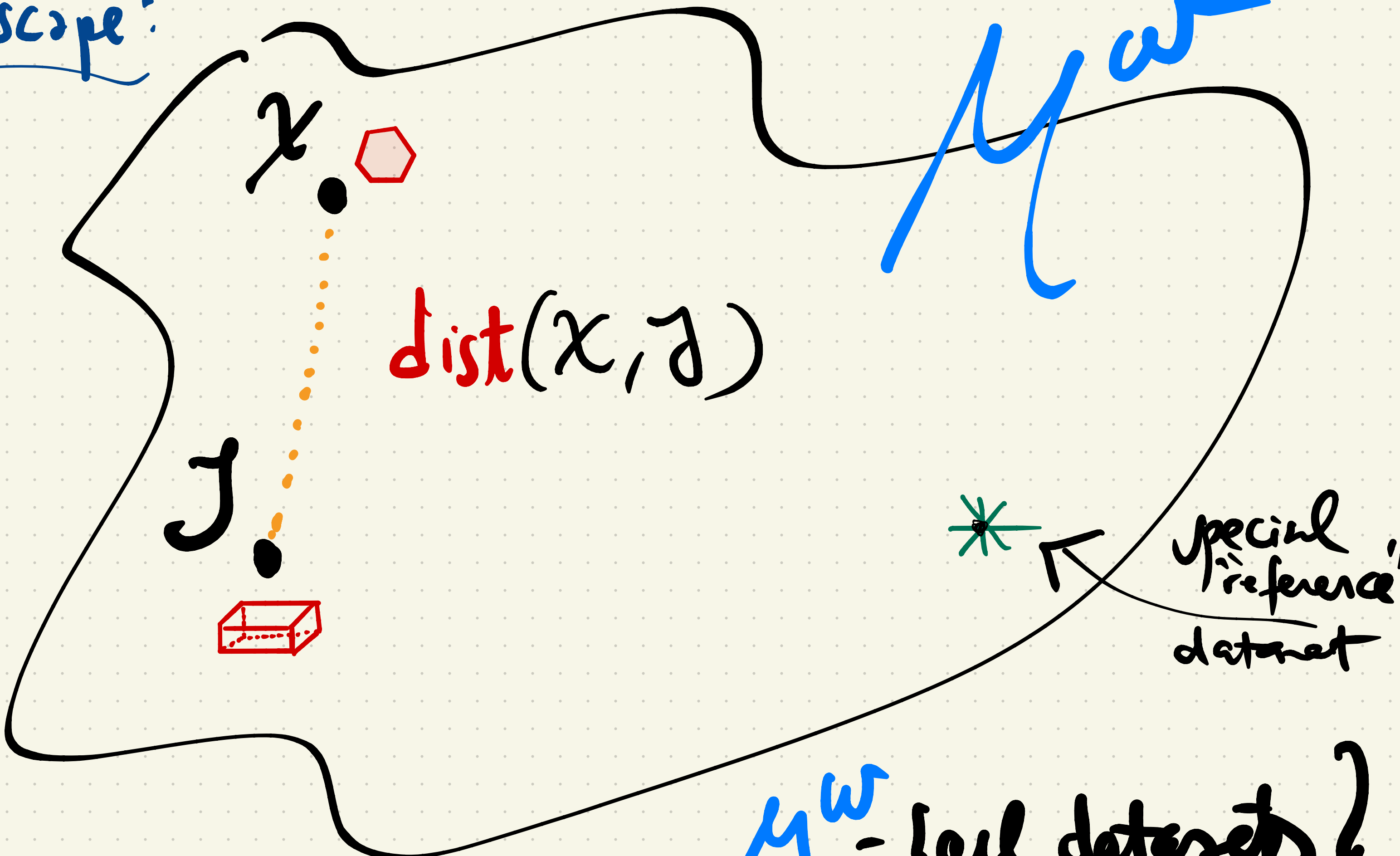
$$\left(\begin{array}{l} X = \text{lexicon} \\ d_X \cong \text{strength of semantic relationship} \\ \mu_X \cong \text{relative frequency of word.} \end{array} \right)$$



Automatic language translation via alignment of "word embedding spaces"

Gromov-Wasserstein Alignment of Word Embedding Spaces
David Alvarez-Melis, Tommi S. Jaakkola
EMNLP'18: Empirical Methods in Natural Language Processing. 2018.

Landscape:

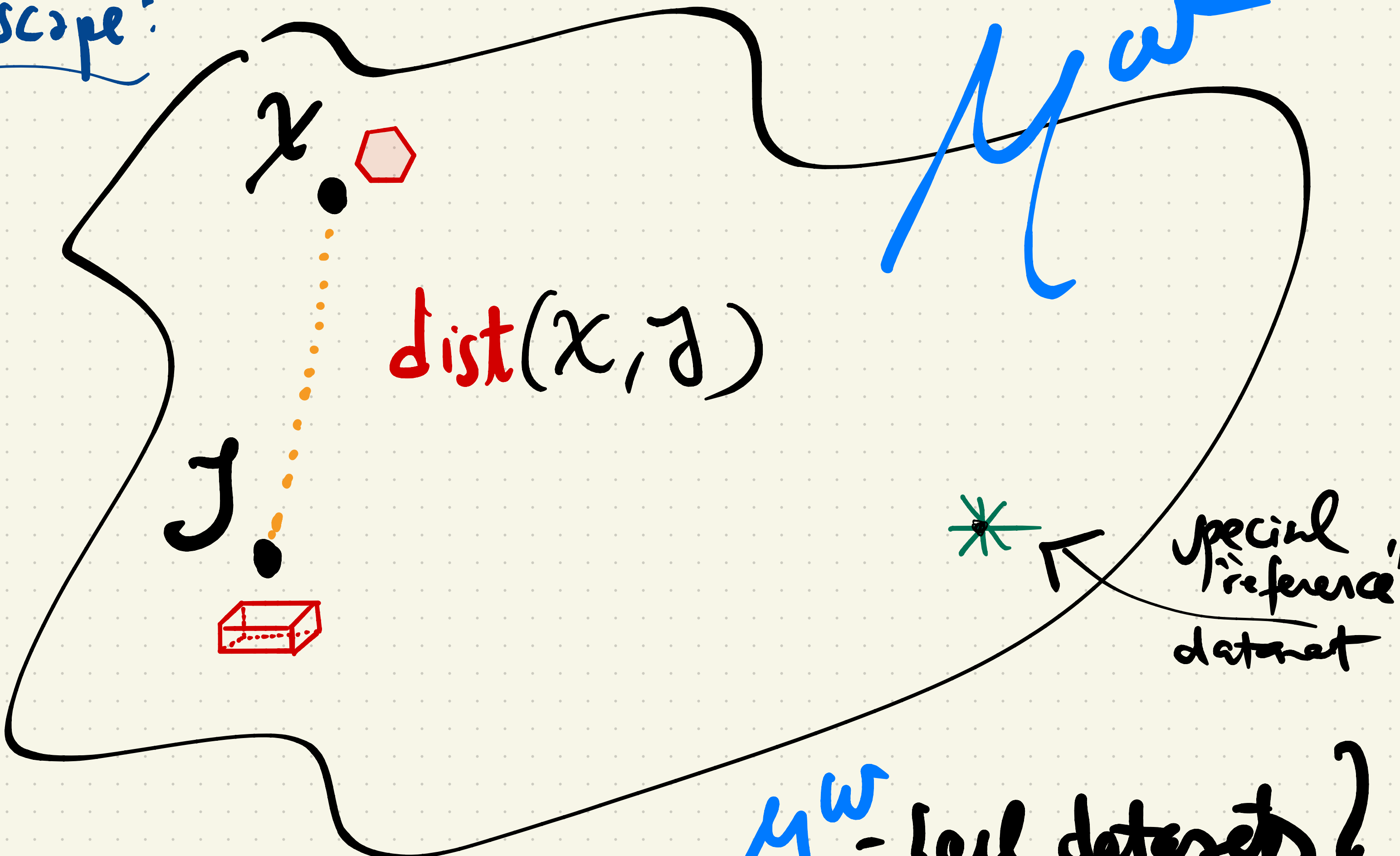


$M^w = \{\text{all datasets}\}$

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

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

Landscape:



$\mathcal{M}^w = \{\text{all datasets}\}$

goal: construct / define dist on $\mathcal{M}^w$ .....

But before that we need to declare

Equality of datasets

$$\cong : \mathcal{M}^w \times \mathcal{M}^w \longrightarrow \{0, 1\}$$

non-isomorphic isomorphic

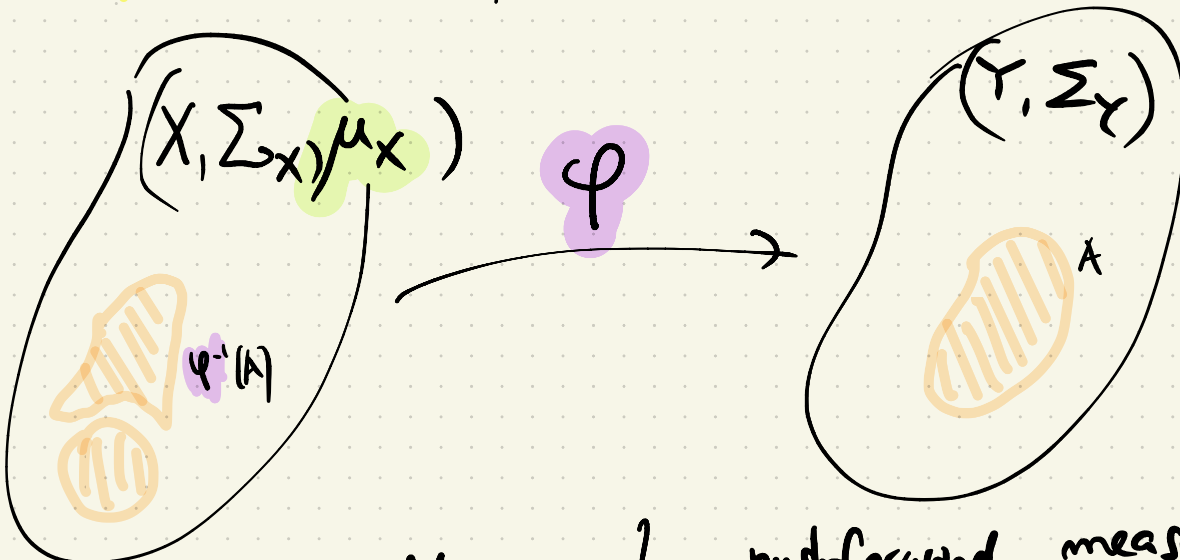
Def Two mm. spaces $\mathcal{X}$ & $\mathcal{Y}$ are **isomorphic**, denoted $\mathcal{X} \cong \mathcal{Y}$

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

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

(measure preserving isometry)

$\#$: the pushforward



φ : measurable map
μ_X : measure on X

$\Rightarrow$ pushforward measure $\varphi_\# \mu_X$ is measure on Y

defined by : for $A \in \Sigma_Y$

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

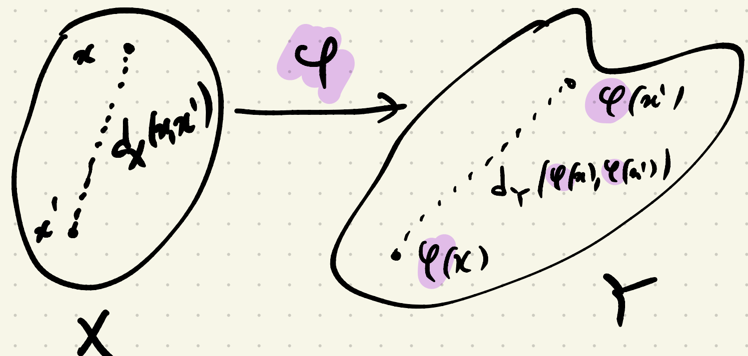
What is an *isometry*?

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

A map $\varphi : X \longrightarrow Y$ 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.



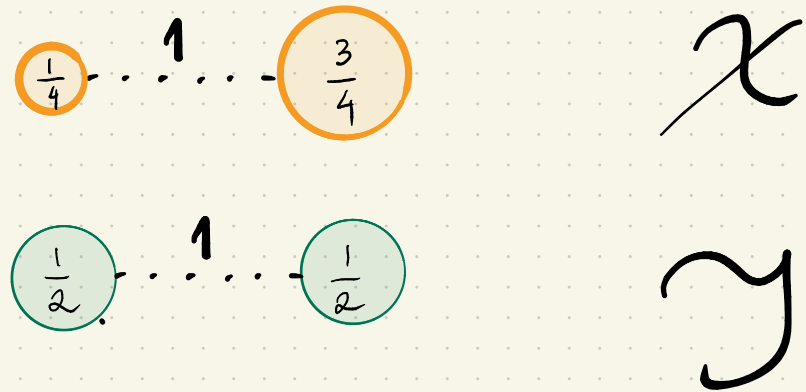
Def Two mm. spaces $\mathcal{X}$ & $\mathcal{Y}$ are *isomorphic*, denoted $\mathcal{X} \cong \mathcal{Y}$

$\Longleftrightarrow$ $\exists\, \Psi : X \longrightarrow Y$ isometry s.t.

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

(measure preserving isometry)

Non - example



$$\Rightarrow \quad \mathcal{X} \not\cong \mathcal{Y}$$

(no isometry respects the weights)

The construction of

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

$$(\mathcal{X}, \mathcal{Y}) \longmapsto \mathrm{dist}(\mathcal{X}, \mathcal{Y})$$

$$\left(\text{st} \quad \mathcal{X} \cong \mathcal{Y} \iff \mathrm{dist}(\mathcal{X}, \mathcal{Y}) = 0 \right)$$

Main idea: to relate $\mathcal{X}$ with $\mathcal{Y}$ use "soft maps".

$\Downarrow$

(stochastic)

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

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

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

$$\begin{tikzcd}
 & \mu \\
 & X \times Y \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\
X & & Y \\
\mu_X & & \mu_Y
\end{tikzcd}$$

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

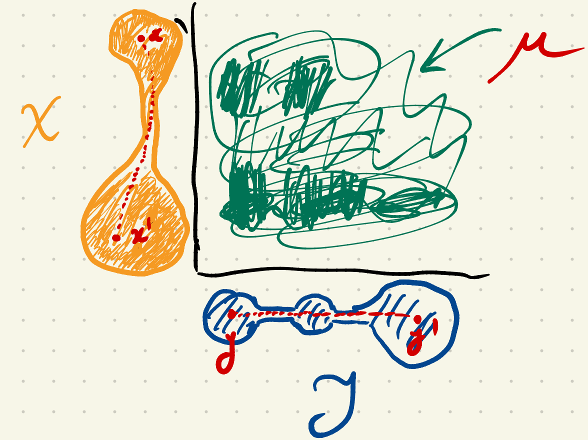
Fact: (1) You can always find at least <u>one</u> coupling:

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

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

(3) If $\Psi : X \to Y$ is an isomorphism $\Rightarrow \mu_\Psi := (id_X, \Psi)_\# \mu_X$ is a coupling

How good is a given μ ?



Let $p \geq 1$

Def The p-distortion of μ,

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

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

Expanding into a more explicit formula:

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

(For later reference, in the finite setting:

$$\mathrm{dis}_p(\mu) = \left[\sum_{ijkl} \left| d_X(x_i, x_k) - d_Y(y_j, y_l) \right|^p \mu_{ij}\, \mu_{kl} \right]^{1/p}$$
)

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

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

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

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

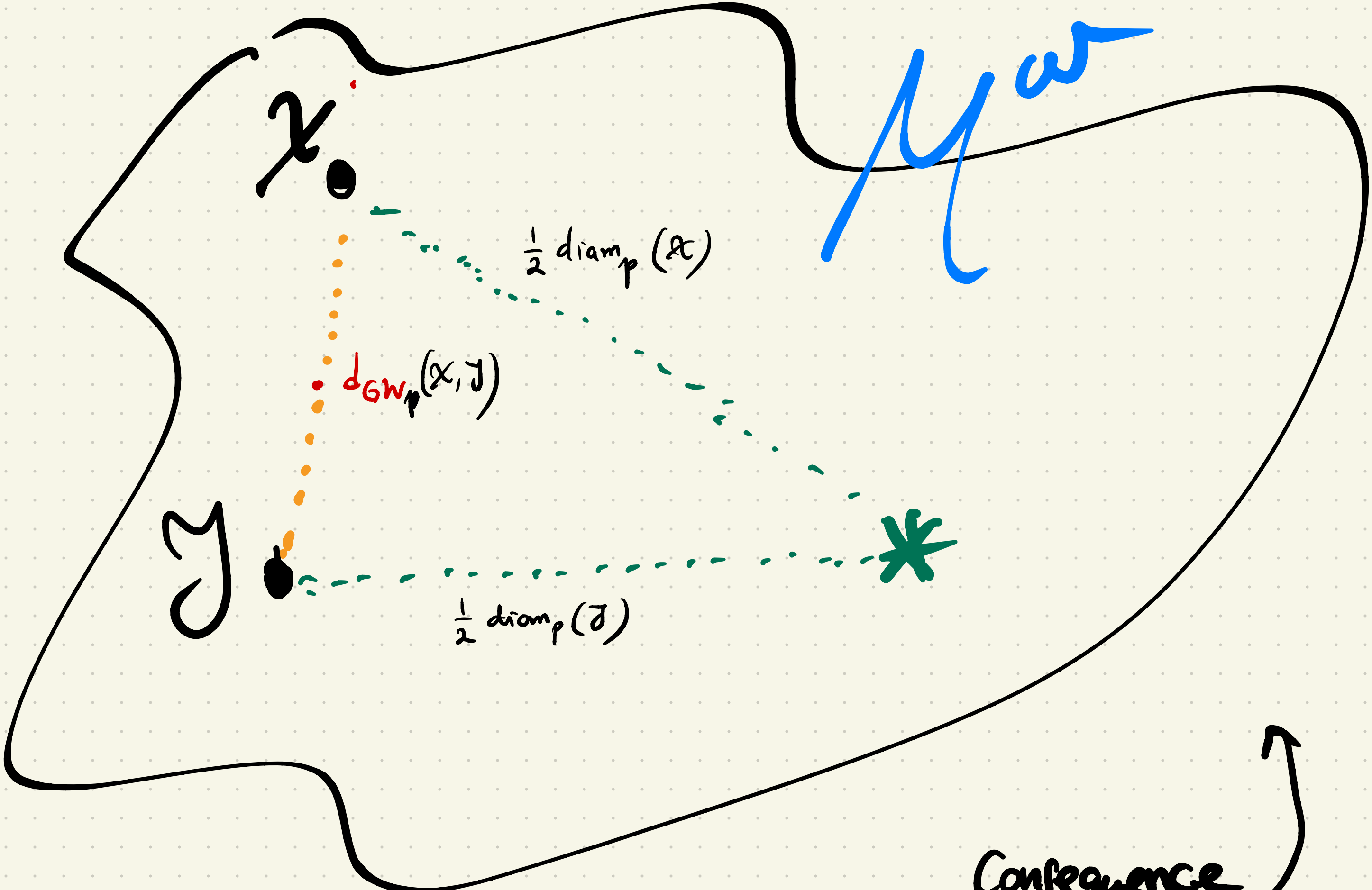
Example $\quad 2\cdot d_{GW,p}(X, *) = \left[\iint \left(d_X(x,x')\right)^p \mu_X(dx)\,\mu_X(dx')\right]^{1/p}$

$$=: \mathrm{diam}_p(X)$$

$(*, ((0)), \delta_*)$

[the p-diameter of X]

This is because we have unique coupling $\mu_X \otimes \delta_*$ between μ_X & δ_*



Consequence

Now we have functions :

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

&

- $d_{GW,p} : \mathcal{M}^w \times \mathcal{M}^w \longrightarrow \mathbb{R}_+$

How are they related ?

Is it true that $\boxed{d_{GW,p}(X,Y)=0 \Leftrightarrow X \cong Y}$?

Main Theorem (Mémoli 2008, Sturm 2012) For every $p \geq 1$,

$d_{SW,p}$ is a legitimate distance on $\underline{\mathcal{M}^w} \backslash \cong$:

(1) it is symmetric

(2) $d_{SW,p}(X,Y) = 0 \iff X \cong Y$

(3) It satisfies the triangle inequality

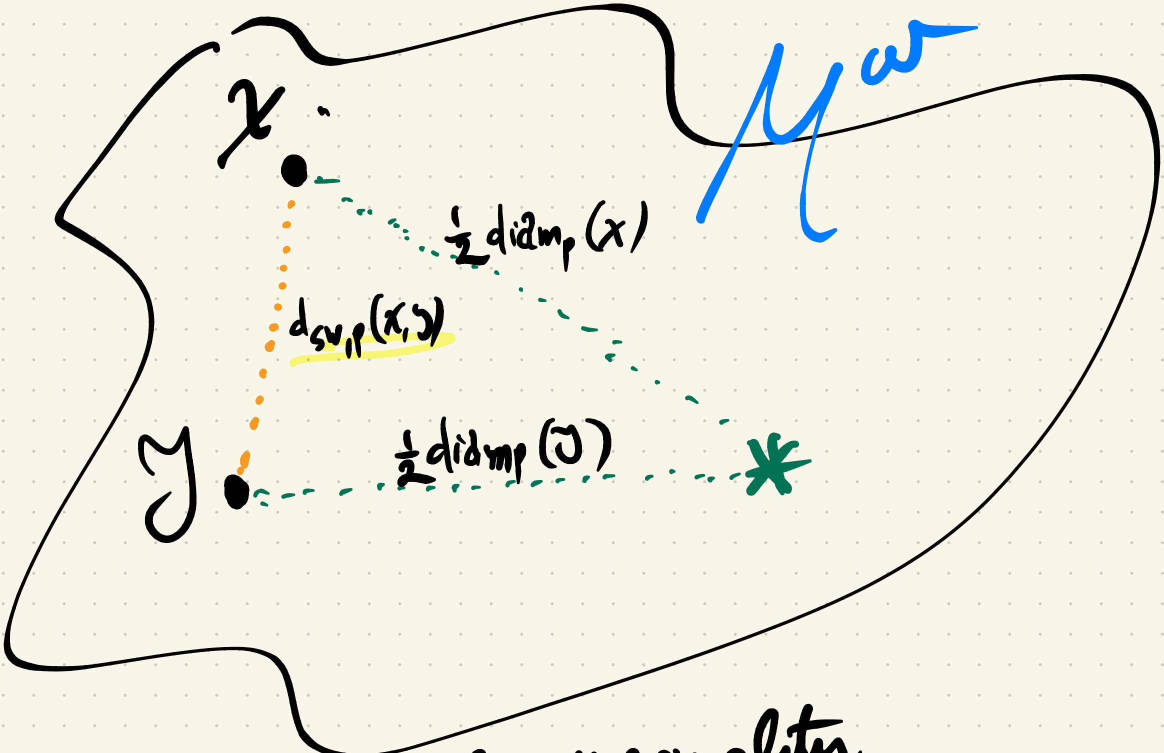
(3') $(\mathcal{M}^w, d_{SW,p})$ is NOT complete.

Furthermore,

(4) It is an intrinsic/geodesic distance.

(5) $(\overline{\mathcal{M}^w}, d_{GW,2})$ is Alexandrov with Curv ≥ 0.

Example

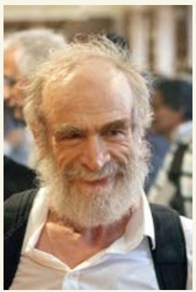


By the triangle inequality,

$$\frac{1}{2}\left|\mathrm{diam}_p(x) - \mathrm{diam}_p(y)\right| \leq d_{GW,p}(x,y) \leq \frac{1}{2}\left(\mathrm{diam}_p(x) + \mathrm{diam}_p(y)\right)$$

A historical Note:

M. Gromov



RECENT

CLASSICAL

D. Wasserstein

L. Kantorovich

G. Monge.

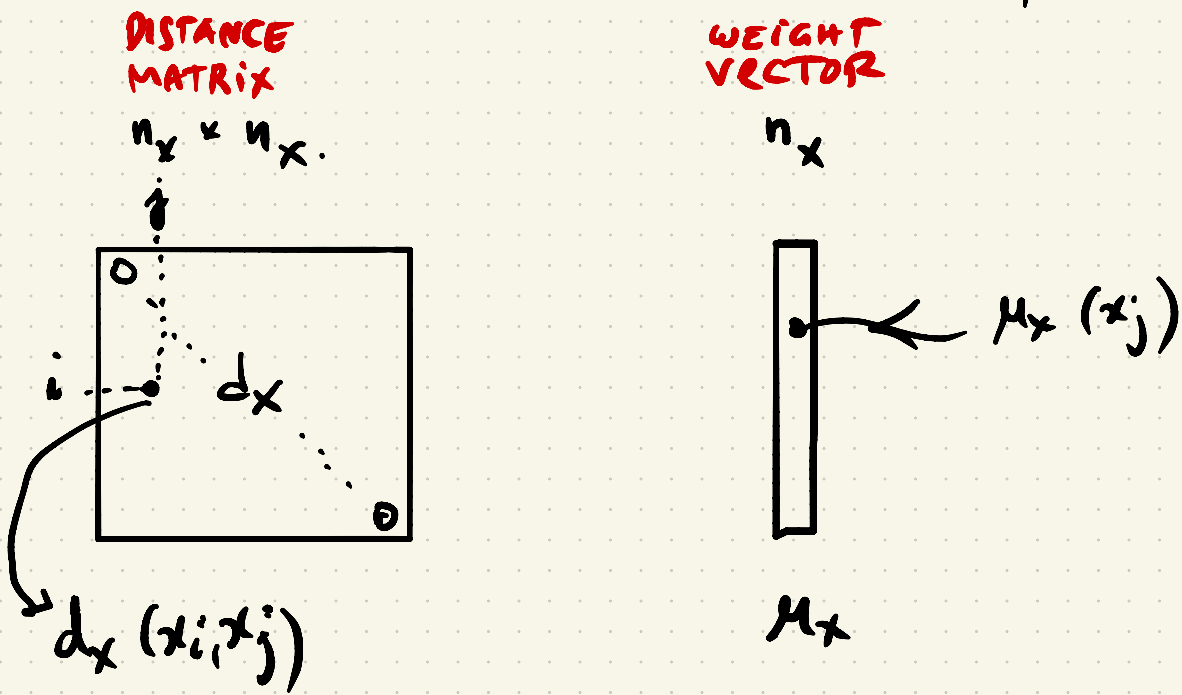
Metric Geometry

Optimal Transport

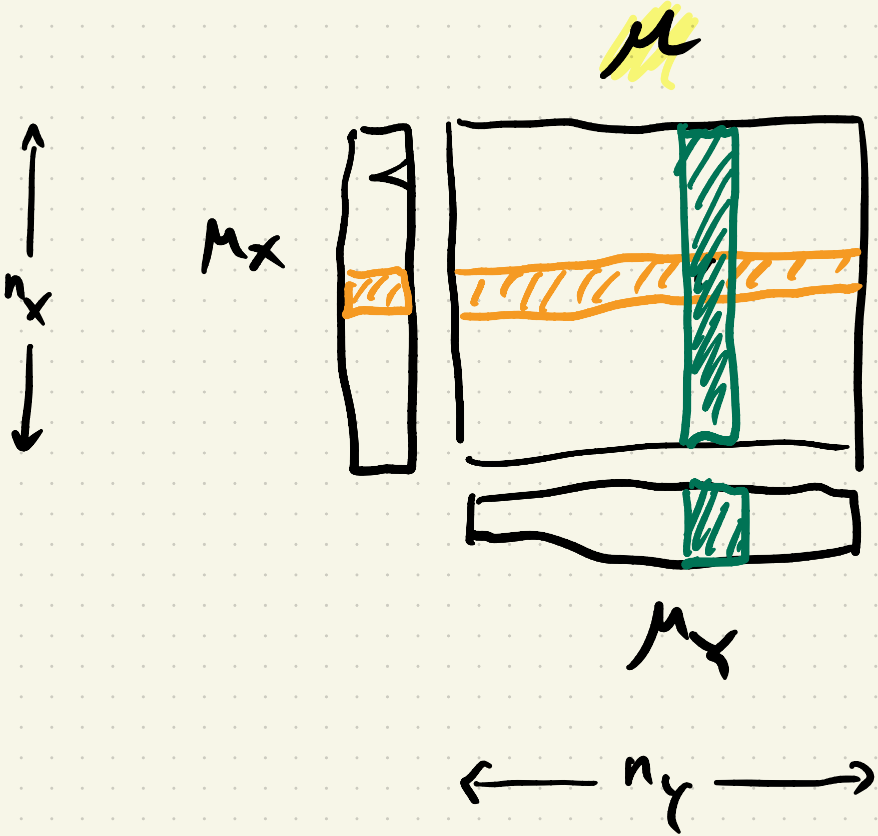
(The Gromov-Wasserstein distance is a generalization of the so called Gromov-Hausdorff distance, a notion which is useful in Metric/Differential/Riemannian Geometry)

HOW DO WE COMPUTE $dGW_{1,p}$?

In the discrete world. $X \in M^w$ is represented as



Given X, Y finite, a coupling μ is a matrix:



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

$$\begin{cases} \sum_j \mu_{ij} = \mu_X(i) \ \forall i \\ \sum_i \mu_{ij} = \mu_Y(j) \ \forall j \end{cases}$$

$\Downarrow$

Linearly Constrained

Say $p=1$ for simplicity.

$$=\Gamma_{ijkl}$$

$$\underline{\underline{\mathrm{dis}_1(\mu)}} = \sum_{ijkl} \left| d_X(x_i, x_k) - d_Y(y_j, y_l) \right| \mu_{ij} \mu_{kl}$$

$$= \sum_{ijkl} \Gamma_{ijkl} \, \mu_{ij} \mu_{kl} = \underline{U^T \Gamma U}$$

bilinear form

$$\Rightarrow \boxed{d_{GW,1}(X, Y) = \frac{1}{2} \min_{U} \underline{\underline{U^T \Gamma U}}}$$

Quadratic functional

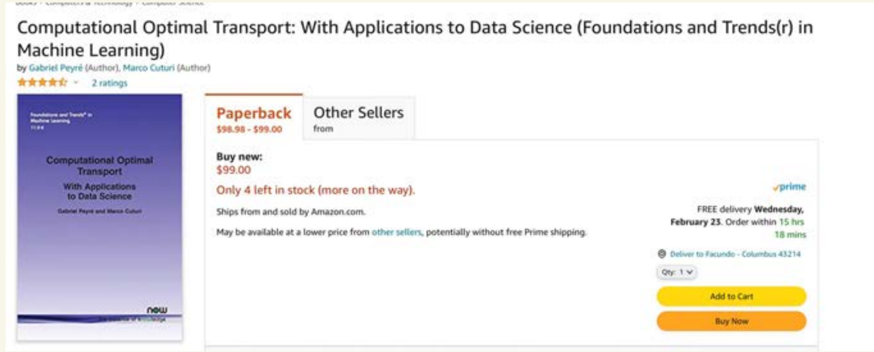
linearly constrained

.... but Γ need NOT be PSD in general

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

A number of computational techniques & implementations have been proposed:

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



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

LOWER BOUNDS.

GOAL:

$$\underbrace{d_{GW,p}(X,Y)}_{\text{difficult}} \geq \underbrace{LB(X,Y)}_{\text{easier}}$$

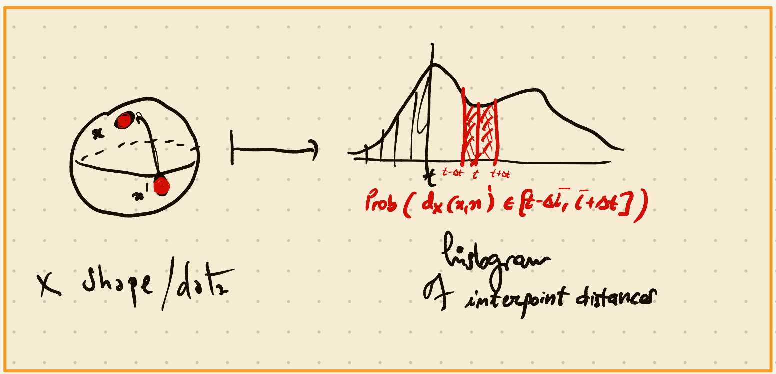
A simple idea: Global distribution of distances

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

Fairly classical idea

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

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



Def (GDD) $\boxed{dH_X = (d_X)_\# \mu_X \otimes \mu_X}$ is the global distribution of distances

H_X : cumulative of the measure dH_X

$$H_X(t) = \mu_X \otimes \mu_X\left(\left\{(x,x') \mid d_X(x,x') \leq t\right\}\right)$$

Proposition $(p=1)$

$$d_{GW,1}(X,Y) \geq \frac{1}{2}\, d_{W,1}^{\mathbb{R}}(dH_X, dH_Y) =: SLB_1(X,Y)$$

(Second lower bound)

Wasserstein distance on $\mathbb{R}$

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

$$SLB_1(X,Y) = \frac{1}{2}\int_0^\infty |H_X(t) - H_Y(t)|\, dt$$

$\Longrightarrow$ easily computable

Question How good is SLB ?

i.e. is it true that

$$\boxed{SLB(X,Y) = 0 \iff X \cong Y}\ ?$$

Note $SLB(X,Y) = 0 \iff dH_X = dH_Y$

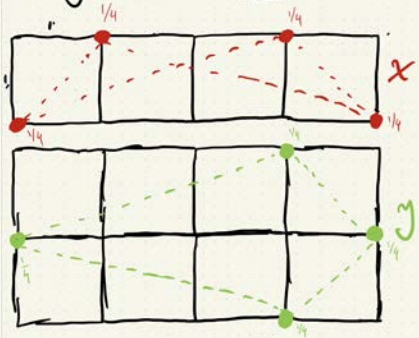
$\Longrightarrow$ question is whether

$$\boxed{dH_X = dH_Y \iff X \cong Y}\ ?$$

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 $\longrightarrow$ Consider :



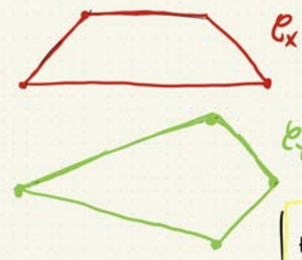
Example by M. Boutin & G. Kemper (2002)

Say all weights are $\frac{1}{4}$ ($\mu_X = \mu_Y =$ "uniform").

distances in $\mathcal{X}$	distances in $\mathcal{Y}$
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.

$H_{\mathcal{X}} = H_{\mathcal{Y}}$

Peter Olver's Conjecture



(Extrinsic distances)

Peter noticed that the curves determined by each set of 4 points, $\mathcal{C}_X$ & $\mathcal{C}_Y$, had (via Matlab) $H_{\mathcal{C}_X} \neq H_{\mathcal{C}_Y}$

Peter Olver's Conjecture

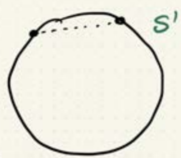
Is it true for planar curves that $H_{\mathcal{C}} = H_{\mathcal{C}'} \iff \mathcal{C} \cong \mathcal{C}'$

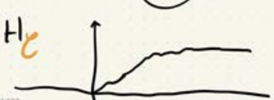
isometric (in $\mathbb{R}^2$)

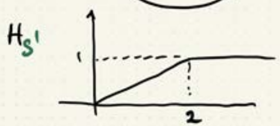
Some answers (joint with Tom Needham)

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

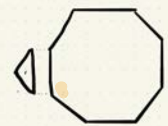
THEOREM





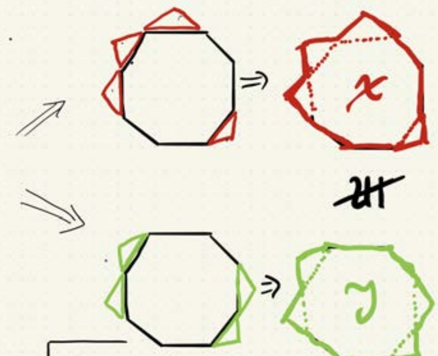


(2) Bad news...



Start with a regular OCTAGON

and an isosceles triangle



However $H_{\mathcal{X}} = H_{\mathcal{Y}}$

checked by hand (w/ lots of patience!)

Much can be said about this question ...

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

Theorem

$\forall\, \varepsilon > 0$ $\exists$ curves $\mathcal{X}_\varepsilon, \mathcal{Y}_\varepsilon$ within ε of the unit circle s.t. $\mathcal{X}_\varepsilon \not\cong \mathcal{Y}_\varepsilon$ but

$$H_{\mathcal{X}_\varepsilon} = H_{\mathcal{Y}_\varepsilon}$$

(4) Similar pattern for embedded surfaces

Proposition $\mathcal{S} \subset \mathbb{R}^3$ embedded closed smooth surface,

Then $H_{\mathcal{S}} = H_{S^2} \iff \mathcal{S} \cong S^2$

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

(6) The Riemannian Setting $\sim$ Corollary to Cheng's rigidity (cf. Bonnet-Myers)

Corollary Let (M, g_M) be d-dimensional closed Riem. mfld with $\mathrm{Ric} \geq d-1$. Then,

1. $\exists$ $\varepsilon = \varepsilon(d) > 0$ s.t. $d^R_{W,1}(dH_M, dH_{S^d}) < \varepsilon$ $\Rightarrow$ 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. mfld. $\dim(M) = d$

scalar curvature

$$H_M(t) = \frac{\omega_d(t)}{\mathrm{Vol}(M)}\left(1 - \frac{\int_M \mathrm{Scal}_M(p)\,\mathrm{vol}_M(dp)}{6(d+2)\,\mathrm{vol}(M)}\, t^2 + O(t^4)\right)$$

Corollary when $d = 2$

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

(in dim = 2 case. $\mathrm{Scal}_M(p)$ = Gauss Curvature at p.
$\int_M \mathrm{Scal}_M(p)\,\mathrm{vol}_M(dp) = C \cdot \chi(M)$)

(via Gauss-Bonnet thm)

$\int \mathrm{gauss} = \chi(M)$

Discussion

- There are "higher order" distributional invariants (both local & global)

  $\longrightarrow$ tradeoff between discriminative power & computational ease.

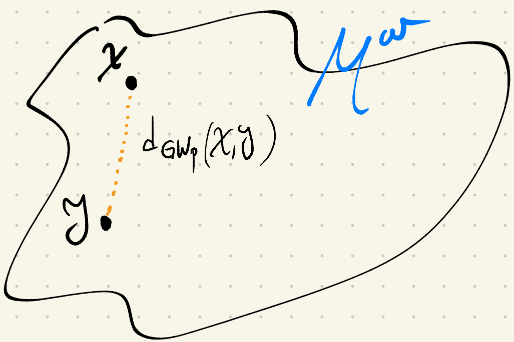
- Connection between GW distance & Weisfeiler-Lehman test

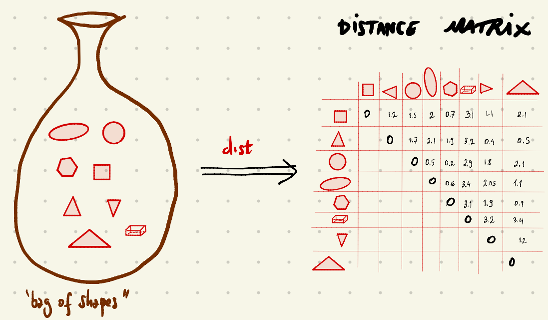
  $\longrightarrow$ applications to GNNs (graph neural networks)

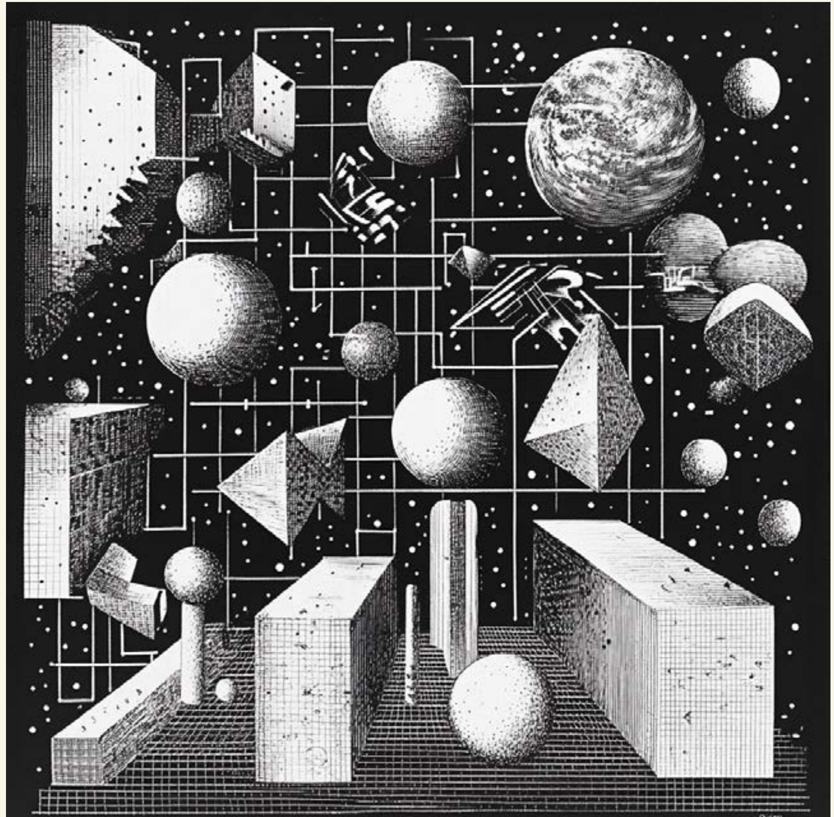
- Instances when (variants of) d_{GW} can be computed/approximated in polynomial time.

- Recent: exact determination of $d_{GW}(S^m_E, S^n_E) \rightsquigarrow$ benchmarking

Thank You


x
y
$d_{GW_p}(x,y)$
$\mathcal{M}^{\omega}$


dist
DISTANCE MATRIX
"bag of shapes"



References

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

Karl-Theodor Sturm: University of Bonn, Bonn, Germany

MEMOIRS of the American Mathematical Society

Published: 30 April 2011

Gromov–Wasserstein Distances and the Metric Approach to Object Matching

Facundo Mémoli

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

1617 Accesses | 55 Citations | 9 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 ...

STUDIES IN APPLIED MATHEMATICS

...nce distributions and inverse problems for metric measure ...es

arXiv.org > math > arXiv:2101.05756

Mathematics > Metric Geometry

The ultrametric Gromov-Wasserstein distance

Facundo Mémoli, Axel Munk, Zhengchao Wan, Christoph Weitkamp

PMLR Proceedings of Machine Learning Research

Weisfeiler-Lehman Meets Gromov-Wasserstein

Samantha Chen, Sunhyuk 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

Shreya Arya, Arnab Auddy, Ranthony A. Clark, Sunhyuk Lim, Facundo Mémoli & Daniel Packer

240 Accesses