Gromov-Wasserstein Bound between Reeb and Mapper Graphs

XVII CLAPEM

Ziyad Oulhaj [1] Bertrand Michel [1] Mathieu Carrière [2]

[1]Laboratoire de Mathématiques Jean Leray
Nantes Université - École Centrale Nantes

[2]DataShape - Centre Inria d'Université Côte d'Azur

March 2, 2026

Background

Reeb Graph

Let X be a topological space and let $f : X \to \mathbb{R}$ be a continuous function called *filter function*.

Let $\sim_f$ be the equivalence relation between two elements x and y in X defined by:
$x \sim_f y$ if and only if x and y are in the same connected component of $f^{-1}(z)$ for some z in $f(X)$.

The Reeb graph R of (X, f) is defined as the quotient space $X/ \sim_f$.

Reeb Graph example

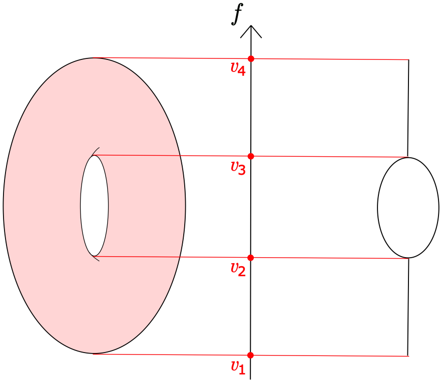


Figure: Example of a Reeb graph of a torus with the height function.

Mapper Algorithm

Given a sample $\mathbb{X}_n = \{X_1, \ldots, X_n\}$, taken from a topological space X as well as a filter function f :

1. Cover the range of values $Y_n = f(\mathbb{X}_n)$ with a set of consecutive intervals $\mathrm{I}_1, \ldots, \mathrm{I}_r$ that overlap.
2. Group the points that fall in the same connected component of each preimage $f^{-1}(\mathrm{I}_j)$. This defines a *pullback cover* $\mathcal{C}$ of $\mathbb{X}_n$.
3. The Mapper graph $\mathbb{M}_n$ is defined as the *nerve* of $\mathcal{C}$.

Mapper Example

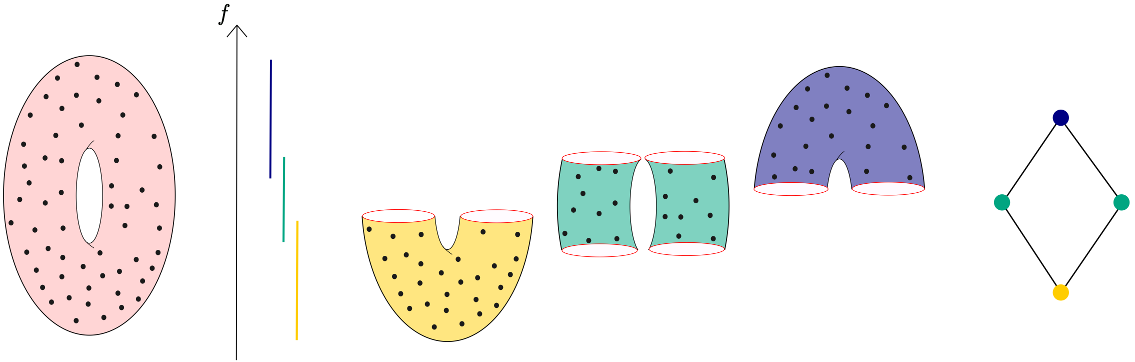


Figure: Example of a Mapper graph built on a torus.

Question

As n(sample size), r(resolution) $\to \infty$:

$$\text{"d"}\ (\mathrm{R}, \mathbb{M}_n) \to 0.$$

Question

As n(sample size), r(resolution) $\to \infty$:

$$\text{"d"}\ (\mathrm{R}, \mathbb{M}_n) \to 0.$$

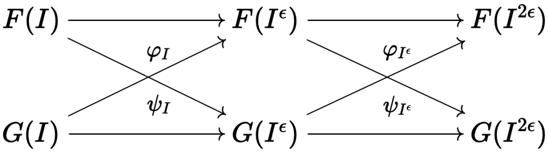
- "d"?
- Relationship between r and n ?
- X_n is random. Probabilistic result ?

Existing approach: Cosheaves [De Silva et al., 2016]

A reeb graph is a covariant functor $F : \textbf{Int} \to \textbf{Set}$.

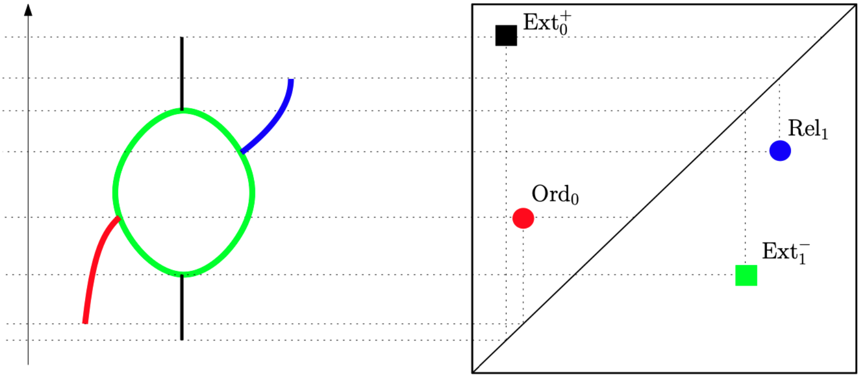
- $F(I) = \pi_0(f^{-1}(I))$
- $F(I \subseteq J) = \pi_0(f^{-1}(I) \subseteq f^{-1}(J))$

$\epsilon-$interleaving:



- Stable w.r.t. function perturbations
- Similar approach possible for the Mapper

Existing approaches 2: Homology [Carrière and Michel, 2022]



Goal

Metric that incorporates measure information directly. (Also computable if possible...)

Metric Measure Space geometry

Metric measure spaces [Sturm, 2006]

Metric Measure Space

A metric measure space (or mm-space for short) is a triple $(X, \mathrm{d}, \mathrm{m})$ where:

- (X, d) is a polish metric space, that is complete and separable,
- m is a measure on the Borel $\sigma-$algebra of (X, d) that is locally finite.

Gromov-Wassertein distance

Gromov-Wasserstein distance

The Gromov-Wasserstein $p-$distance between two mm-spaces $(X_1, \mathrm{d}_1, \mathrm{m}_1)$ and $(X_2, \mathrm{d}_2, \mathrm{m}_2)$ is defined as:

$$\mathrm{GW}_p((X_1, \mathrm{d}_1, \mathrm{m}_1), (X_2, \mathrm{d}_2, \mathrm{m}_2)) = \left(\inf_{d,\mu} \int_{X_1 \times X_2} d(x,y)^p \, d\mu(x,y) \right)^{\frac{1}{p}},$$

where the infimum is taken over all measures μ on $X_1 \times X_2$ that have marginals m_1 and m_2, and on all metric couplings d of d_1 and d_2.

Gromov-Wassertein distance

Isomorphy between mm-spaces

If $\mathrm{GW}_2((X_1, \mathrm{d}_1, \mathrm{m}_1), (X_2, \mathrm{d}_2, \mathrm{m}_2)) = 0$ (and under some assumptions), there exists a map $\Phi\colon \mathrm{supp}(\mathrm{m}_1) \to \mathrm{supp}(\mathrm{m}_2)$ such that:

- Φ is an isometry of metric spaces,
- $\mathrm{m}_2 = \mathrm{m}_1 \circ \Phi^{-1}$.

Reeb graphs as MM-spaces

Framework

In all of the following M is a compact connected Riemannian manifold and f is a Morse function.
We want to look at the Reeb graph $\mathrm{R} := \mathrm{R}(\mathrm{M})_f$ as a metric measure space

Why a Morse function ?

A smooth map $f : \mathrm{M} \to \mathbb{R}$ is called a Morse function if its critical points are non-degenerate.

Why a Morse function ?

A smooth map $f : \mathrm{M} \to \mathbb{R}$ is called a Morse function if its critical points are non-degenerate.

Morse Lemma

Let $f : \mathrm{M} \to \mathbb{R}$ be a Morse function and $c \in \mathrm{Crit}(f)$. Then there exists a chart $\varphi : U \subseteq \mathrm{M} \to \mathbb{R}^d$ containing c such that for every $p \in U$:

$$f(p) = f(c) - \sum_{j=1}^{i} x_j^2 + \sum_{j=i+1}^{d} x_j^2,$$

where $x = \varphi(p)$ and the integer i depends only on the signature of the Hessian at the critical point c.

Why a Morse function ?

A smooth map $f : \mathrm{M} \to \mathbb{R}$ is called a Morse function if its critical points are non-degenerate.

Morse Lemma

Let $f : \mathrm{M} \to \mathbb{R}$ be a Morse function and $c \in \mathrm{Crit}(f)$. Then there exists a chart $\varphi : U \subseteq \mathrm{M} \to \mathbb{R}^d$ containing c such that for every $p \in U$:

$$f(p) = f(c) - \sum_{j=1}^{i} x_j^2 + \sum_{j=i+1}^{d} x_j^2,$$

where $x = \varphi(p)$ and the integer i depends only on the signature of the Hessian at the critical point c.

- $\mathrm{Crit}(f)$ is finite.
- Level sets are locally path connected.

Cylinders

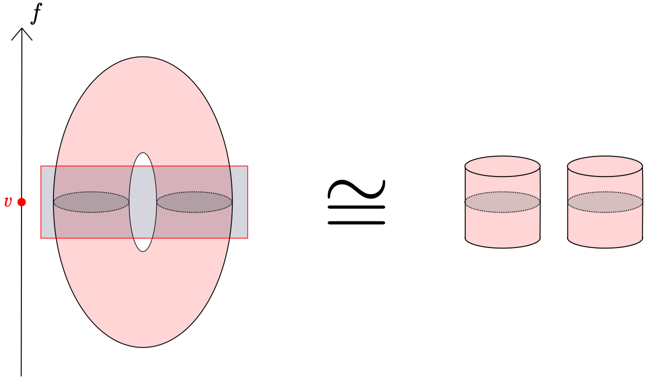


Figure: Around a non-critical value v, the manifold is homeomorphic to a finite collection of cylinders, whose faces are given by the level set $f^{-1}(\{v\})$.

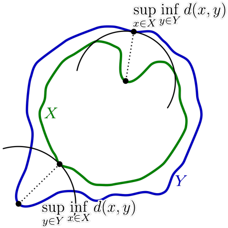
Hausdorff distance

Definition

Let A, B be two subsets of a metric space (X, d), the Hausdorff distance between A and B is defined as:

$$\mathrm{d_H}(A, B) = \max\left(\sup_{x \in A} \mathrm{d}(x, B), \sup_{y \in B} \mathrm{d}(y, A)\right),$$

where $\mathrm{d}(x, B) = \inf_{v \in B} \mathrm{d}(x, v)$ and $\mathrm{d}(y, A) = \inf_{u \in A} \mathrm{d}(u, y)$.



Hausdorff distance property

Proposition

When (X, d) is complete (resp. compact), then $(\mathrm{C}(X), \mathrm{d_H})$ is complete (resp. compact).

For Reeb graphs

- The elements of R are closed in M. $(\mathrm{R}, \mathrm{d_H})$ is therefore a metric space.
- $(\mathrm{R}, \mathrm{d_H})$ is however not closed, and hence not polish...

Not closed

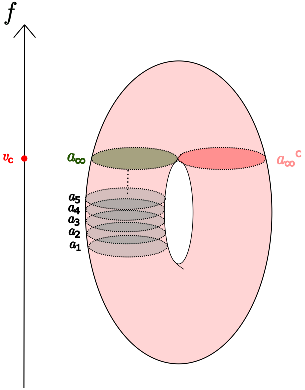


Figure: Example of a Reeb graph sequence $(a_n)_{n\in\mathbb{N}}$ that has a Hausdorff limit a_∞ outside of the Reeb graph. Notice that this occurs when a_∞ is inside the level set of a critical value v_c.

Topology of $(\mathrm{R}, \mathrm{d}_{\mathrm{H}})$

Proposition

Let a be a compact subset of M that is an adherent point to R. Then a is connected and f is constant on a.

Topology of $(\mathrm{R}, \mathrm{d_H})$

Proposition

Let a be a compact subset of M that is an adherent point to R. Then a is connected and f is constant on a.

Proposition

$\pi \colon x \longmapsto [x]_{\sim_f}$ is continuous at every point $x \in \mathrm{M}$ such that $f(x)$ is not critical.

Topology of $(\mathrm{R}, \mathrm{d_H})$

Proposition

Let a be a compact subset of M that is an adherent point to R. Then a is connected and f is constant on a.

Proposition

$\pi : x \longmapsto [x]_{\sim_f}$ is continuous at every point $x \in \mathrm{M}$ such that $f(x)$ is not critical.

Corollary

π is Borel measurable.

Topology of $(\mathrm{R}, \mathrm{d_H})$

Proposition

Let a be a compact subset of M that is an adherent point to R. Then a is connected and f is constant on a.

Proposition

$\pi : x \longmapsto [x]_{\sim_f}$ is continuous at every point $x \in \mathrm{M}$ such that $f(x)$ is not critical.

Corollary

π is Borel measurable.

Proposition

$\bar{\mathrm{R}} \setminus \mathrm{R}$ is finite.

Comparison with the Mapper graph - Main result

Setting

What we showed

$(\bar{\mathrm{R}}, \mathrm{d_H}, \mathrm{m} \circ \pi^{-1})$ is a metric measure space, for m a Borel measure on M.

Aim

Do the same for $\mathbb{M}_n$ and give an upper bound on $\mathrm{GW}_p(\bar{\mathrm{R}}, \mathbb{M}_n)$.

Mapper

We construct an injection between the Mapper (as a simplicial complex) and $\mathrm{C}(\mathbb{M})$, through a map $\blacktriangle : \mathrm{X}_n \to \mathbb{M}_n$.
We then consider $(\mathbb{M}_n, \mathrm{d_H}, \mathrm{m}_n \circ \blacktriangle^{-1})$ where m_n is the empirical measure

$$\mathrm{m}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}.$$

Assumptions

We make two assumptions:

- The Ricci curvature Ric of M is lower bounded:

$$\mathrm{Ric} \geq (d-1) \cdot k$$

  where $k \in \mathbb{R}$.

- The measure m is absolutely continuous with respect to the volume measure Vol and admits an upper bounded density, i.e.,

$$\sup \frac{d\mathrm{m}}{d\mathrm{Vol}} < +\infty.$$

  Furthermore, we assume that m is fully supported on M.

Cover

We will assume a *homogeneous covering of resolution* $r \in \mathbb{N}$, i.e., the intervals $\mathbb{I} = \{\mathrm{I}_1, ..., \mathrm{I}_r\}$ are all of the same length and consecutive intervals have a percentage $0 < g < 1/2$ of their length in common.

We call *maximal width* δ_r of the cover $\mathbb{I}$ the maximal length of an elementary block in $\mathbb{I}$.

Main result

Theorem [Oulhaj et al., 2025]

Consider the two metric measure spaces $(\overline{\mathrm{R}}, \mathrm{d_H}, \mathrm{m} \circ \pi^{-1})$ and $(\mathbb{M}_n, \mathrm{d_H}, \mathrm{m}_n \circ \mathbb{\Delta}^{-1})$ corresponding to the Reeb graph and to the Mapper graph respectively. For any $\alpha > 0$, $p \geq 1$ and a resolution $r(n)$ of the Mapper satisfying $r(n) \underset{n \to \infty}{\sim} n^{\frac{1}{d+\alpha}}$, we then have, for any large enough sample size n:

$$\mathbb{E}\left(\mathrm{GW}_p(\overline{\mathrm{R}}, \mathbb{M}_n)\right) \lesssim n^{-\frac{\nu}{d+\alpha}}$$

where

$$\nu = \min\left\{\frac{1}{2}, \frac{d}{p(d+1)}\right\}.$$

Proof idea

Estimation-Approximation bound

$$\mathrm{GW}_p(\overline{\mathrm{R}}, \mathbb{M}_n) \leq \mathrm{E}_n + \mathrm{A}_n,$$

where

- $\mathrm{E}_n = \mathrm{W}_p(\mathrm{m} \circ \pi^{-1}, \mathrm{m}_n \circ \pi^{-1})$, and
- $\mathrm{A}_n = \left(\frac{1}{n} \sum_{i=1}^{n} \mathrm{d_H}([x_i]_{\sim_f}, \mathbb{A}(x_i))^p\right)^{\frac{1}{p}}$.

Introduce $(\overline{\mathrm{R}}, \mathrm{d_H}, \mathrm{m}_n \circ \pi^{-1})$ as an intermediate space.

Estimation term

$\mathrm{E}_n = \mathrm{W}_p(\mathrm{m} \circ \pi^{-1}, \mathrm{m}_n \circ \pi^{-1})$ is the Wasserstein distance between a measure and an associated empirical measure in a polish metric space ($(\bar{\mathrm{R}}, \mathrm{d_H})$ here).
This has been studied in [Weed and Bach, 2019] where a sharp upper bound is given. The main ingredient is a control on the *covering number* of $(\bar{\mathrm{R}}, \mathrm{d_H})$.

- We relate the covering number of $(\bar{\mathrm{R}}, \mathrm{d_H})$ to that of M.
- We use volume comparison (Bishop-Gromov Theorem) to control the covering number of M.

Covering number for R

Proposition

There exists $\varepsilon' > 0$ such that for every $\varepsilon \leq \varepsilon'$ we have

$$\mathcal{N}_{\varepsilon}(\mathrm{R}) \leq N_c + \left(\frac{\beta}{\varepsilon}\right)^{2d},$$

where N_c is the number of elements in R associated to critical values, d is the dimension of M and β is a constant depending on f and M.

Approximation term

$\mathrm{A}_n = \left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{d_H}([x_i]_{\sim_f}, \mathbb{\Delta}(x_i))^p\right)^{\frac{1}{p}}$ is bounded explicitly.

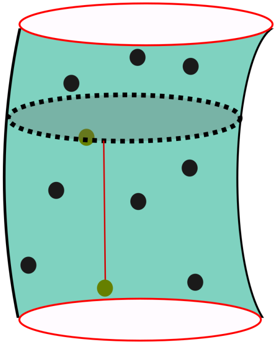
- Away from critical values, the gradient flow of f can be used.
- Otherwise...

Step 1

For any large enough resolution r and for every $x_j \in \mathbb{A}(x_i)$:

$$\mathrm{d}(x_j, [x_i]_{\sim_f}) \leq \max_{c \in \mathrm{Crit}(f)} \mu(c) \cdot \sqrt{\frac{1-g}{g}} \cdot \sqrt{\delta_r}$$

$\mu(c)$ is the square root of the spectral radius of the metric at c.



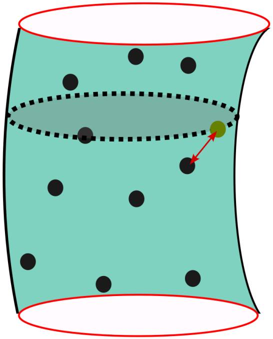
Step 2

Let $\lambda > 0$ such that $\mathrm{d_H}(\mathrm{X}_n, \mathrm{M}) \leq \lambda$ and

$$\omega_f(\lambda) := \sup_{\mathrm{d}(x,y)\leq\lambda} |f(x) - f(y)| \leq \frac{g}{1-g}\frac{\delta_r}{4}.$$

For any large enough resolution r and for every $y \in [x_i]_{\sim_f}$:

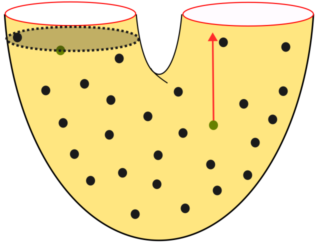
$$\mathrm{d}(y, \mathbb{A}(x_i)) \leq \lambda + \frac{1}{2} \cdot \max_{c\in\mathrm{Crit}(f)} \mu(c) \cdot \sqrt{\frac{1-g}{g}} \cdot \sqrt{\delta_r}.$$



Step 3

Close to critical values. We need to "bruteforce" the problem.
For any large enough resolution r, we have:

$$\max_{\mathrm{J}\in\mathbb{J}} \mathrm{Vol}\left(f^{-1}(\mathrm{J})\right) \leq \eta \cdot \delta_r^{\frac{d}{d+1}},$$



where

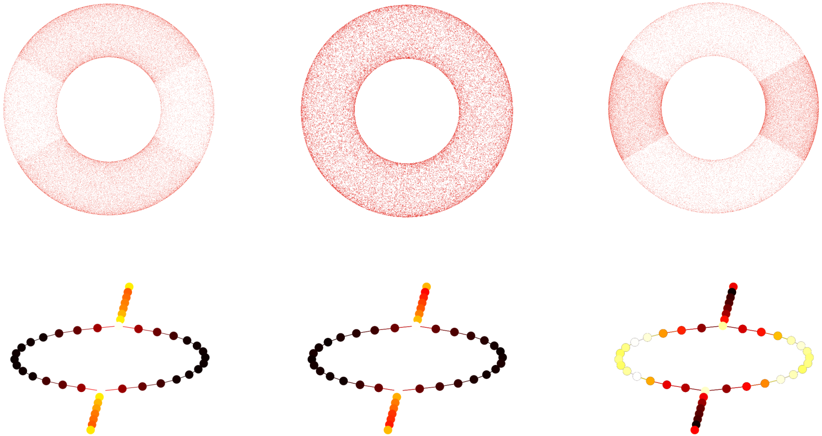
$$\eta = \max_{c\in\mathrm{Crit}(f)} \mu(c)^2 \cdot \sup_{v\notin f(\mathrm{Crit}(f))} \mathrm{Vol}_{d-1}\left(f^{-1}(\{v\})\right) + 2\alpha_d \left|\mathrm{Crit}(f)\right|,$$

and where $\mathrm{Vol}_{d-1}\left(f^{-1}(\{v\})\right)$ is the $(d-1)$-dimensional volume of $f^{-1}(\{v\})$ seen as a submanifold of M and α_d is the volume of the unit ball in $\mathbb{R}^d$.

Unfortunately

This last step gives the slowest term that dictates the final convergence rate :(

In practice



In practice

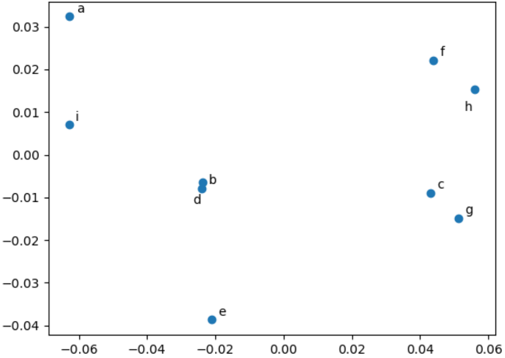


Figure: MDS visualization of the Gromov-Wasserstein distance matrix between the different Mapper graphs.

Label	left	right
a	$1/12$	$1/12$
b	$1/12$	$1/6$
c	$1/12$	$1/3$
d	$1/6$	$1/12$
e	$1/6$	$1/6$
f	$1/6$	$1/3$
g	$1/3$	$1/12$
h	$1/3$	$1/6$
i	$1/3$	$1/3$

Table: Correspondence between the labels and the parameters of the measure.

Thank you for listening !

Carrière, M. and Michel, B. (2022).
Statistical analysis of mapper for stochastic and multivariate filters.
Journal of Applied and Computational Topology, 6(3):331–369.

De Silva, V., Munch, E., and Patel, A. (2016).
Categorified reeb graphs.
Discrete & Computational Geometry, 55(4):854–906.

Oulhaj, Z., Carrière, M., and Michel, B. (2025).
Gromov-wasserstein bound between reeb and mapper graphs.

Sturm, K.-T. (2006).
On the geometry of metric measure spaces. i.
Acta Math, 196:65–131.

Weed, J. and Bach, F. (2019).
Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance.
Bernoulli, 25(4A):2620–2648.