

# Géométrie des Données et Apprentissage sur Variétés

## Module 1 – Introduction enrichie avec exercices

Mastère Spécialisé HPC-AI

November 19, 2025

# Plan

- Introduction à la topologie des données
- Variétés riemanniennes
- Transport optimal
- Applications à l'IA générative
- TP : Mapper (KeplerMapper), mini-projet transport optimal (POT library).

- Étudier la forme globale des données : trous, cycles, composantes
- Techniques :
  - ① **Mapper** : visualisation simplifiée d'un nuage de points
  - ② **Persistent Homology** : détecter des features topologiques robustes
- Diagramme de persistance : barres représentant la durée de vie des composantes

# Exemple : Persistent Homology

Points  $\rightarrow$  Graph simplicial  $\rightarrow$  Filtration  $\rightarrow$  Barcodes

- Chaque barre = un trou ou composante connectée
- Longue barre  $\rightarrow$  feature robuste
- Courte barre  $\rightarrow$  bruit

- Une variété lisse  $(\mathcal{M}, g)$  avec métrique  $g$  définit longueur et angles
- Distance géodésique  $d_{\mathcal{M}}(x, y) =$  longueur minimale d'un chemin sur  $\mathcal{M}$
- Exemples : sphère, tore, espace de rotations  $SO(3)$
- Applications : données sur sphère, pose 3D, embedding non-linéaire

# Transport optimal

- Comparer deux distributions  $\mu$  et  $\nu$
- Distance de Wasserstein :

$$W_p(\mu, \nu) = \left( \inf_{\gamma \in \Pi(\mu, \nu)} \int \|x - y\|^p d\gamma(x, y) \right)^{1/p}$$

- Applications :
  - Comparaison distributions réelles vs générées
  - Génération de données réalistes avec structures géométriques

# Applications à l'IA générative

- Score-based diffusion : génération par gradient de log-densité
- Transport optimal : mesurer distances entre distributions générées et réelles
- Topologie persistante : régulariser génération pour préserver cycles/structures
- Graphes et variétés : génération de molécules, maillages 3D, images structurées



- Topologie : Mapper, Persistent Homology  $\rightarrow$  analyser la forme globale des données
- Géométrie avancée : variétés riemanniennes et distances géodésiques
- Transport optimal : distance de Wasserstein pour comparer distributions
- Applications IA générative : score-based diffusion, OT, régularisation topologique