Persistent Homology-based Projection Pursuit

Oleg Kachan
Skolkovo Institute of Science and Technology
Nobel st. 3, Moscow, Russia 121205
oleg.kachan@skoltech.ru

Abstract

Dimensionality reduction problem is stated as finding a mapping $f : X \in \mathbb{R}^m \to Z \in \mathbb{R}^n$, where $n \ll m$ while preserving some relevant properties of the data. We formulate topology-preserving dimensionality reduction as finding the optimal orthogonal projection to the lower-dimensional subspace which minimizes discrepancy between persistent diagrams of the original data and the projection. This generalizes the classic projection pursuit algorithm which was originally designed to preserve the number of clusters, i.e. the 0-order topological invariant of the data. Our approach further allows to preserve k-th order invariants within the principled framework. We further pose the resulting optimization problem as the Riemannian optimization problem which allows for a natural and efficient solution.

1. Introduction

We reformulate the classic projection pursuit [17, 24] algorithm using persistent homology-based loss function. The projection pursuit seeks for "interesting" representations of data, by finding the data's low-dimensional projection maximizing the certain loss function, called *projection index*.

While there can be a lot of possible indices, the projection pursuit's original loss function [17], being the trade-off between the data spread and local density is designed to preserve clusters in data. This could be naturally formulated as minimizing the Wasserstein distance d_{W_2} between 0-th persistence diagrams $D(X)$ and $D(P^T X)$ – the discrete measures, quantifying the topology of the data $X \in \mathbb{R}^m$ and its orthogonal projection $P^T X \in \mathbb{R}^n$, in terms of their 0-order topological invariants – connected components or clusters

$$\min_{P \in St(n,m)} d_{W_2}(D(X), D(P^T X)), \tag{1}$$

where $St(n, m)$ is the set of orthonormal $m \times n$ matrices, such that $n < m$.

The benefit of the new formulation is that it is more general, further allowing to preserve the higher-order topology invariants of the data with nontrivial topology, with retaining the intrinsic clusters present in the data being the simplest case. More engaging case is the preservation of k-dimensional topological holes with $(k-1)$-dimensional loops bounding them for $k \geq 1$, as this pattern is present in periodic data. Examples of such data are natural images under different lighting conditions, 3D shapes under rotations, and sliding window embeddings of periodic time series [33].

There are various examples of dimensionality reduction algorithms which are designed to preserve intrinsic geometric properties of the dataset such as local or global geodesic [38] distances, affine connections [36], heat kernels [5], and tangent spaces [43]. We propose to preserve a weaker invariant of the dataset under consideration, namely its topology and explicitly encode topology preservation in the loss function. Unlike local-to-global methods which may preserve topology to some extent, we consider directly minimizing the discrepancy between the discrete measures summarizing the global topological invariants of the original and compressed data. We show the example on model data for which our algorithm and PCA give two different low-dimensional representations, preserving the cycle in the former case and recovering the maximum variance in the latter.

We are aware of two works [41, 31] studying the same problem of global structure preservation, relying on persistent homology and discrete Morse theory. Our approach is conceptually simpler and selects the data transformation from the class of linear transformations, which have the benefits of interpretability and explicit out-of-sample mapping, compared to the algorithm [41] based on landmark Isomap. While [31] use topology loss as a regularization, we show that the preservation of the global structure of the data may be of interest on its own and could lead to meaningful results if paired with local structure preserving transformation.

Our work relies on the results of differentiablity of optimal transport [18] and its' approximations, particularly entropy-regularized Wasserstein distance [34] and recently introduced notion of differentiability of the persistent homology mapping [20, 35, 19] for the finite metric spaces. Our model is end-to-end differentiable, thus optimized by gradient descent, and is composed of three differentiable mappings, for one of which the parameters lie on nonlinear matrix manifold, parameterizing the low-dimensional linear subspace, projection to which maximally preserves the global topology of the dataset.

Selection of the proper emdedding dimension n is as important as the projection itself. The intrinsic dimension estimation methods [10] could be used to provide the initial guess to find the dimension of low topology distortion.

The rest of the paper is organized as follows: in section 2 we give our formulation of the topology-preserving dimensionality reduction problem, write down optimization problem and give the solution of it. In section 3 we give the required background on topological inference , persistent homology, Riemannian optimization, and computational optimal transport to keep the paper self-contained, followed by section 4 which summarizes the algorithm. In section 5 we present the results of experiments on model data. Section 6 concludes the paper.

2. Problem formulation

Given a dataset $X \in \mathbb{R}^m$ find a mapping $f_\theta : X \in \mathbb{R}^m \to Z \in \mathbb{R}^n$, where $n \ll m$ while preserving the topological properties of the data. The topology-preserving loss function is defined as Wasserstein distance [40, 37] between the discrete measures quantifying the topology of the data, namely persistent diagrams [15] of the original dataset $D(X)$ and its transformation $D(f_\theta(X))$:

$$f_\theta^* = \min_\theta W_2(D(X), D(f_\theta(X))) \tag{2}$$

We select the transformation f_θ from the class $\mathcal{F}$ of linear transformations, which makes our approach linear dimensionality reduction method. Although being conceptually simple, this class of methods enjoy interpretability, computational tractability, and explicit mapping for out-of-sample data points.

Problem formulation

Find the linear projection $P \in \mathbb{R}^{m \times n}$ to a n-dimensional linear space $Z = P^T X$ in the m-dimensional ambient space projection to which minimizes the loss function mentioned previously:

$$P^* = \min_P W_2(D(X), D(P^T X)) \tag{3}$$

2.1. Optimization problem

Following the approach to solving linear dimensionality reduction problems proposed in the the seminal paper [12] we pose the optimization problem of topology-preserving linear dimensionality reduction with differentiable Wasserstein loss [18] as optimization on Stiefel manifold $St(n, m)$. This makes our optimization problem Riemannian optimization problem, which is unconstrained optimization problem over the constrained feasible set with well established theory [14, 1] and solvers available [8, 39].

$$\begin{aligned} &\min W_2(D(X), D(P^T X)), \\ &\text{s.t. } P \in St(n, m). \end{aligned} \tag{4}$$

2.2. Solution

Finding the solution requires giving precise answers to the following questions: (1) what are persistent diagrams $D(X)$ and how to compute them? (2) how to find optimal linear transformation P^* and how exactly it is parametrized? (3) how to solve the problem of learning with Wasserstein loss efficiently? We address the aforementioned questions in the two following sections.

3. Methods

3.1. Topological inference and computational topology

For a finite sample from a probability distribution persistent homology allows to robustly infer and quantify the topology of the data. Surprisingly enough, while the topology is inferenced using the discrete structures – simplicial complexes, the resulting persistent homology mapping is differentiable under mild assumptions [26].

3.1.1 Persistent homology

Given a topological space X and a function f defined on it, the persistent homology [3, 16] quantifies the topology of a pair (X, f), more concretely the changes in topology of the sequence $\{X_t\}_{t \in T}$ of sublevel sets $X_t = f^{-1}(-\infty, t] = \{x \in X \mid f(x) \leq t\}$ of the function f.

For a sequence of sublevel sets $\{X_t\}_{t \in T}$, called a *filtration* of a set X w.r.t. the function f, persistent homology sets in correspondence a collection of k-th persistence diagrams $\{D_f^k(X)\}$, where $k = 0, 1, \ldots, k_{max}$ for some k_{max}. A *k-th persistence diagram* is a multiset of pairs $\{(b_i, d_i)\}_{i \in I}$ of birth b_i and death d_i of topological features of order k, such as connected components and k-dimensional holes, bounded by $(k-1)$-dimensional cycles. In the following we will omit k, with the order of the persistent homology mapping and the persistent diagram to be inferred from the context. A topological feature is called

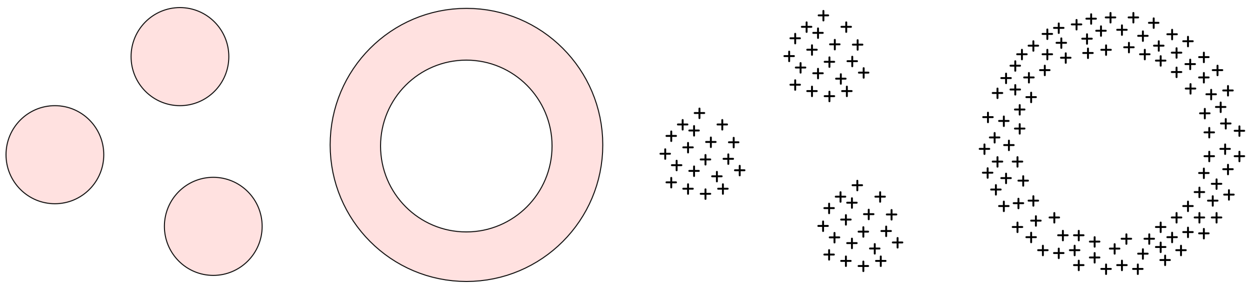

Figure 1. Topology of a continuous sets (clusters, annulus) and their finite samples. The continuous sets topology is non-trivial, having 3 connected components for the clusters and 1 connected component and 1 hole for annulus. For the finite samples from these sets the number of connected components equals to the number of points, and the all the higher-order topological features are non-existent.

persistent if it exists for a long interval of the filtration parameter t.

3.1.2 Persistent homology of finite metric spaces

Finite metric space topology is uninteresting by itself, with the number of connected components equals to the number of points, and the all the higher-order topological features are non-existent (see the Figure 1). By assuming the data is just a finite sample from a continuous set, one can inference the topology of the underlying data manifold by considering the collection of ε-thickenings of the space [6]. In practice ε-thickenings are modeled by discrete geometric simplicial complexes having data points X as the vertex set, showed to have the same topology as the union of balls [7, 15] (or close approximation) in the ε-thickening, such as *Čech* (or *Vietoris-Rips* complex) respectively.

An *(abstract) simplicial complex* K on a finite set X (called *vertex set*) is the family of subsets $\sigma = \{v_0, \dots, v_k\}$ of X closed under inclusion. That is for every $\sigma \in K$ any of its subset $\tau \subseteq \sigma$ is also in K. A subset σ having $k+1$ elements is called *k-simplex* and has *dimension k*. In a pair of simplices $(\tau \subseteq \sigma) \in K$, τ is a *face* of σ and σ is a *coface* of τ.

Given a finite metric space X and an ε-ball $B_\varepsilon(x)$ at point x a *Čech complex* of X is as a family of k-simplices corresponding to intersections of k ε-balls $\mathrm{Cech}_\varepsilon(X) = \{\sigma \in X \mid \bigcap_{x \in \sigma} B_\varepsilon(x) \neq 0\}$. A *Vietoris-Rips complex* of X at radius ε is defined $VR_\varepsilon(X) = \{\sigma \in X \mid d(x, x') \leq 2\varepsilon,\ \forall (x, x') \in \sigma\}$, that is $k+1$ points form a k-simplex if they are all pairwise 2ε-distant.

3.1.3 Differentiability of persistent homology

A persistent homology can be viewed as a mapping which takes a topological space X and a function f to the k-th persistent diagram $D_f(X)$

$$PH : (X, f) \to D_f(X). \tag{5}$$

Consider a topological space $\mathcal{X}$ with a continuous function $f : \mathcal{X} \to \mathbb{R}$ defined on it. During the filtration (i.e. changing ε from 0 to ∞) topological properties of the sublevel sets of the function f will change only at finite number of critical points of f. A *persistence pairing* $\Pi_f(\mathcal{X}) = \{(b_i, d_i)\}_{i \in I}$ is the set of pairs of critical points of the function corresponding to the birth and death of topological features. A *persistence diagram* is the set of function values at critical points $D_f(\mathcal{X}) = \{(f(b_i), f(d_i)) \in \mathbb{R}^2 \mid (b_i, d_i) \in \Pi_f(\mathcal{X})\}_{i \in I}$.

Now consider the finite sample X of the topological space $\mathcal{X}$ with the same continuous function f defined on it. The topology of the space is inferred by taking a sequence of ε-thickenings of the space and tracking the changes of topological properties of the sublevel sets of the function f, modeled by a geometric simplicial complex $K(X)$. In this setting, a *persistence pairing* $\Pi_f(K(X)) = \{(\tau_{b_i}, \sigma_{d_i})\}_{i \in I}$, such that $\dim(\sigma) - \dim(\tau) = 1$ is the set of pairs of simplices of $K(X)$, corresponding to the critical points of the function f. Then, a *persistence diagram* is the set of function values at simplices corresponding to the critical points $D_f(K(X)) = \{(f(\tau_{b_i}), f(\sigma_{d_i})) \in \mathbb{R}^2 \mid (b_i, d_i) \in \Pi_f(X)\}_{i \in I}$.

In the case of the Vietoris-Rips filtration of a finite metric space X the function value on a simplex $f(\sigma)$ is extended from the pairwise distances between the pairs of points (x_i, x_j) constituting a simplex σ

$$\begin{aligned} f(\sigma) &= \max_{(x_i, x_j) \in \sigma} f(x_i, x_j) = \max_{(x_i, x_j) \in \sigma} \|x_i - x_j\|_2, \\ f(x_i) &= 0. \end{aligned} \tag{6}$$

The *inverse function* π_f [20, 35] maps points in the diagram $D_f(X)$ corresponding to the critical values of the function f to the vertices of the critical simplices giving birth and death to topological features, thus allowing to differentiate the points in the diagram w.r.t. the data points X [9]. In the case of the Vietoris-Rips filtration this function is given by the persistent homology algorithm [44] and can be efficiently computed [29, 4].

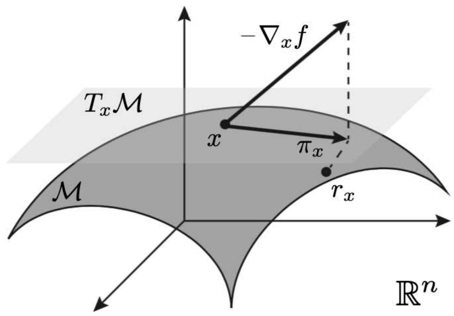

Figure 2. Given a Riemannian manifold $\mathcal{M} \in \mathbb{R}^n$ and a scalar function f, for the each iteration the Euclidean gradient $\nabla_x f$ is projected π_x to the tangent space $T_x\mathcal{M}$ where the gradient step is performed, followed by the retraction r_x to the manifold.

3.2. Riemannian optimization

Riemannian optimization [14, 1, 22] generalizes optimization algorithms to the Riemannian manifolds [25] other than $\mathbb{R}^n$. Applications which can benefit of such approach include dimensionality reduction [12, 42], matrix factorization and completion, optimization with orthogonality constraints [14], computing statistics of manifold-valued data [30], and learning on manifold-valued data [23] or with manifold-valued outputs [21].

Let $\mathcal{M}$ be a Riemannian manifold, and $f : \mathcal{M} \to \mathbb{R}$ is a scalar real valued function defined on $\mathcal{M}$, then a Riemannain optimization problem

$$\min_{x \in \mathcal{M}} f(x) \tag{7}$$

could be solved iteratively by gradient descent along the geodesics:

Data: A manifold $\mathcal{M}$, a scalar field f on $\mathcal{M}$, a projection $\pi_x : \mathbb{R}^n \to T_x\mathcal{M}$, a retraction $r_x : T_x\mathcal{M} \to \mathcal{M}$, an initial iterate $x_0 \in \mathcal{M}$, step size $\alpha \in \mathbb{R}_+$
Result: Sequence of iterates $\{x_k\}$

for $k = 1, 2, \ldots$ **do**
  $x_{k+1} = (r_{x_k} \circ \alpha\pi_{x_k})(-\nabla_x f(x_k))$;
return $\{x_k\}$

Algorithm 1: Gradient descent on a Riemannian manifold [1]

3.2.1 Optimization on the Stiefel manifold

As we previously mentioned we pose the algorithm's optimization problem as optimization on the Stiefel manifold, parameterizing the optimal linear subspace w.r.t. the loss function. The *Stiefel manifold* [1] is the set of all n-dimensional orthonormal frames in m-dimensional space, represented by $m \times n$ matrices

$$St(n, m) = \{\mathbf{X} \in \mathbb{R}^{m\times n} : \mathbf{X}^T\mathbf{X} = \mathbf{I}\}, \tag{8}$$

where $\mathbf{I} \in \mathbb{R}^{n\times n}$.

In the case of the Steifel manifold the projection and retraction maps are given in terms of matrix decompositions. Specifically, the projection can be based on SVD-decomposition [1], let $\mathbf{X} = \mathbf{USV}^T$, then $\pi_X : \mathbb{R}^{m\times n} \to T_X St(n, m)$ is $\pi_X = \mathbf{UV}$.

The retraction mapping based on QR-decomposition [1] is given by $r_{\mathbf{X}} := qf(\mathbf{X} + \mathbf{V})$, where $qf(\mathbf{A})$ denotes the $\mathbf{Q}$ factor of the decomposition of $\mathbf{A} \in \mathbb{R}^{m\times n}$ as $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q}$ belongs to $St(n, m)$ and $\mathbf{R}$ is an upper triangular $m \times n$ matrix with strictly positive diagonal elements.

3.3. Computational optimal transport

Optimal transport considers comparing the measures over the domain X. Given a ground metric $d : X \times X \to \mathbb{R}$ optimal transport equips the space of measures $\mathcal{P}(X)$ with a metric referred to as *Wasserstein distance*, which for any $\mu, \nu \in \mathcal{P}(X)$ and $p \geq 1$ is defined as [40, 27]

$$W_p^p(\mu, \nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times X} d^p(x, y)d\pi(x, y) \tag{9}$$

where W_p^p denotes the p-th power of W_p and $\Pi(\mu, \nu)$ is the set of probability measures on the product space $X \times X$ whose marginals coincide with μ and ν; namely

$$\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(X \times X) \mid P_1\#\pi = \mu, P_2\#\pi = \nu\}, \tag{10}$$

with $P_i(x_1, x_2) = x_i$ the projection operators for $i = 1, 2$ and $P_i\#\pi$ the push-forward of π [40].

Consider the discrete measures $\mu, \nu \in \mathcal{P}(X)$ that can be written as linear combinations $\mu = \sum_{i=1}^n a_i\delta_{x_i}$ and $\nu = \sum_{j=1}^m b_j\delta_{y_j}$ of Dirac's deltas centred at a finite number n and m of points $(x_i)_{i=1}^n$ and $(y_j)_{j=1}^m$. In order for μ and ν to be probabilities, the vector weights $\mathbf{a} = (a_1, \ldots, a_n)^T \in \Delta_n$ and $\mathbf{b} = (b_1, \ldots, b_m)^T \in \Delta_m$ must belong respectively to the n and m-dimensional simplex, defined as

$$\Delta_n = \{\mathbf{p} \in \mathbb{R}_+^n \mid \mathbf{p}^T\mathbf{1}_n = 1\}, \tag{11}$$

where R_+^n is the set of vectors $\mathbf{p} \in \mathbb{R}^n$ with non-negative entries and $\mathbf{1}_n \in \mathbb{R}^n$ denotes the vector of all ones, so that $\mathbf{p}^T\mathbf{1}_n = \sum_{i=1}^n p_i$ for any $\mathbf{p} \in \mathbb{R}^n$.

The Wasserstein distance between the two discrete measures μ and ν with corresponding weight vectors $\mathbf{a}$ and $\mathbf{b}$ corresponds to the solution of network flow problem

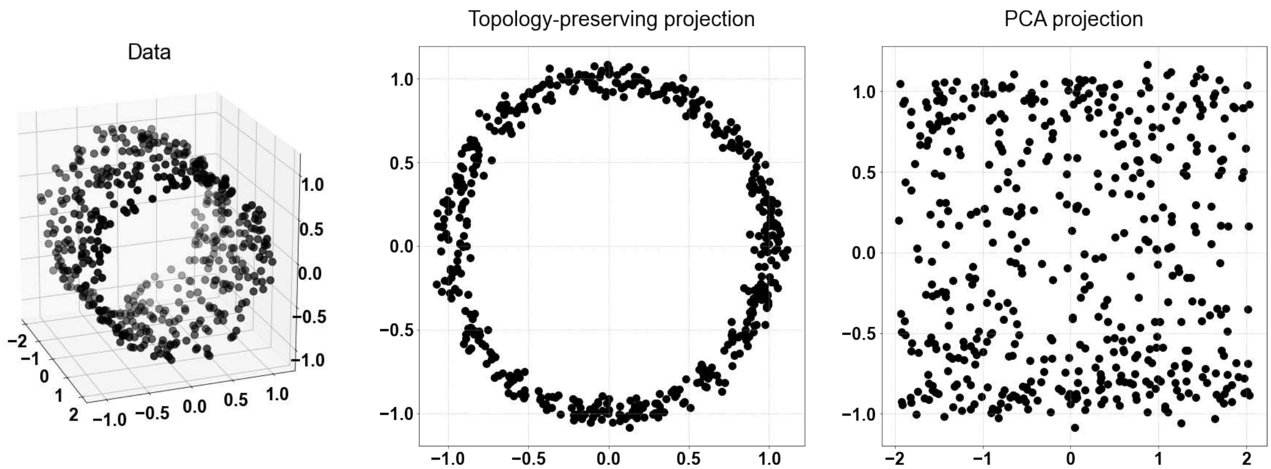

Figure 3. Example on model data of noisy cylinder $\mathbb{S}^1 \times [-2, 2]$ for which our algorithm and PCA give two different low-dimensional representations, preserving the cycle in the former case and recovering the maximum variance in the latter.

$$W_p^p(\mu, \nu) = \min_{\mathbf{T} \in \Pi(\mathbf{a}, \mathbf{b})} \langle \mathbf{T}, \mathbf{M} \rangle_F, \quad (12)$$

where $\mathbf{M} \in \mathbb{R}^{n \times m}$ is the cost matrix with entries $\mathbf{M}_{ij} = d(x_i, y_j)^p$, $\langle \mathbf{T}, \mathbf{M} \rangle_F$ is the elementwise Frobenius inner product, and $\Pi(\mathbf{a}, \mathbf{b})$ denotes the transportation polytope

$$\Pi(\mathbf{a}, \mathbf{b}) = \{\mathbf{T} \in \mathbb{R}_+^{n \times m} : \mathbf{T}\mathbf{1}_m = \mathbf{a}, \mathbf{T}^T \mathbf{1}_n = \mathbf{b}\}. \quad (13)$$

3.3.1 Optimal transport for persistence diagrams

Consider persistence diagrams $D(X)$ and $D(Z)$ of cardinality n and m respectively. The cost matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ is augmented [15, 32] with matrices $\Delta_{D(X)} \in \mathbb{R}^{n \times 1}$ and $\Delta_{D(Y)} \in \mathbb{R}^{1 \times m}$ where each entry is the persistence topological feature $p \in D(\cdot)$ to effectively balance measures, allowing to match points in the diagrams with a diagonal Δ, meaning the topological feature with zero persistence.

$$\tilde{\mathbf{M}} = \left(\begin{array}{c|c} \mathbf{M} & \Delta_{D(X)} \\ \hline \Delta_{D(Y)} & 0 \end{array} \right) \quad (14)$$

3.3.2 Gradient of approximated Wasserstein distance

While it is possible to compute a subgradient of the Wasserstein distance, the complexity of its evaluation is quadratic with respect to the dimension of the output space [18, 27]. Thus, computationally Wasserstein distance is approximated either by the entropic regularization [13, 2] or as the integral of 1D measures [11] derived from the original one. Computation of both approximations consist of matrix operations for which closed form of the gradient exists or automatic gradient [28] can be efficiently evaluated.

Sinkhorn distance

The Sinkhorn algorithm gives the solution to the optimal transport problem with the entropic regularization, and gives gradient in closed form [18], automatic differentiation is also available.

The Sinkhorn distance, given the entropy of the transport plan matrix $h(\mathbf{T})$ and the regularization parameter λ is defined

$$S_\lambda(\mathbf{a}, \mathbf{b}) = \min_{\mathbf{T} \in \Pi(\mathbf{a}, \mathbf{b})} \langle \mathbf{T}, \mathbf{M} \rangle_F - \frac{1}{\lambda} h(\mathbf{T}), \quad (15)$$

$$\text{where } h(\mathbf{T}) = -\sum_{i,j=1}^{n,m} \mathbf{T}_{ij} (\log \mathbf{T}_{ij} - 1)$$

The optimization can be solved efficiently via Sinkhorn's matrix scaling algorithm [13] and its variants [2].

4. Algorithm

With all the technicalities described in the previous section we conclude the persistent-homology based projection pursuit algorithm.

Given a dataset $X \in \mathbb{R}^m$, the Vietoris-Rips simplicial complex $K_f(X)$ with the filtration function $f(\sigma) = \max_{(i,j) \in \sigma} \|x_i - x_j\|_2$ construct the following mappings and their gradients:

Orthogonal projection

Orthogonal projection $Z = P^T X$, $P \in St(n, m)$ and its Riemannian gradient $\frac{\partial \mathbf{Z}}{\partial \mathbf{P}}$ according to the Algorithm 1 using projection π and retraction r mappings for the Stiefel manifold.

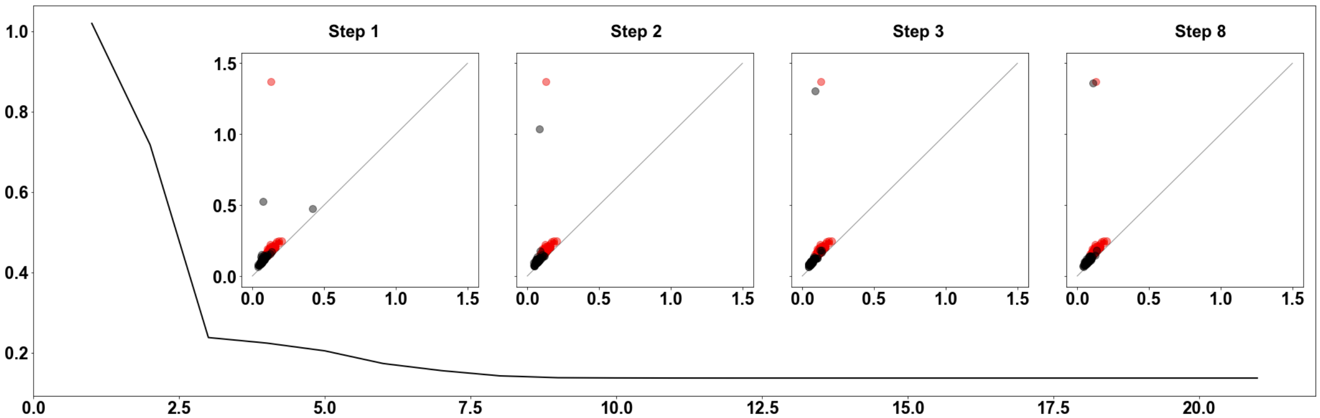

Figure 4. Loss function and persistence diagrams of data $D(X)$ (red, fixed) and low-dimensional projection $D(Z)$ (black) at k-th steps of gradient descent.

Persistent homology mapping

Compute the persistent homology mapping $PH : (X, f) \to D_f(X)$ by constructing the filtration of the Vietoris-Rips complex $K_f(Z)$, computing its boundary matrix $\mathbf{B}$, applying the persistence algorithm [44] to obtain a reduced boundary matrix $\mathbf{R}$ and extracting the persistent diagram $D_f(X)$ and the persistent pairing $\Pi_f(X)$ out of it. To compute the gradient $\frac{\partial D(Z)}{\partial Z}$ for each (b_i, d_i) in $\Pi_f(X)$ apply the inverse function π_f to find two pairs of critical vertices $(x_j^+, x_k^+) = \arg\max_{(x_j,x_k)\in\tau_{b_i}} \|x_j - x_k\|_2$, and $(x_j^-, x_k^-) = \arg\max_{(x_j,x_k)\in\sigma_{d_i}} \|x_j - x_k\|_2$ belonging to the simplices giving birth and death to the topological features and extreme w.r.t. the function f. Evaluate the gradients $\partial b_i/\partial x_j^+$, $\partial b_i/\partial x_k^+$, $\partial d_i/\partial x_j^-$, and $\partial d_i/\partial x_k^-$ given in the case of the Vietoris-Rips filtration by $\partial_{x_j} f(\sigma) = \frac{x_j - x_k}{\|x_j - x_k\|_2}$, and $\partial_{x_k} f(\sigma) = -\frac{x_j - x_k}{\|x_j - x_k\|_2}$.

Wasserstein distance

Compare two diagrams of the data $D(X)$ (which is kept fixed, thus computed only once) and its projection $D(Z)$ computed at each step of gradient descent using Wasserstein distance $d_{W_2} : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ loss L, particulary its entropic approximation – Sinkhorn distance d_E [13, 2] and its gradient $\frac{\partial L}{\partial D(Z)}$ computed by the automatic differentiation.

Total gradient

Finally, given $L = d_{W_2}(D(X), D(Z))$, $D_Z = PH(Z)$, $Z = P^T X$, the total gradient $\nabla_P L = \frac{\partial L}{\partial P}$ is given by the chain rule

$$\frac{\partial L}{\partial P} = \frac{\partial L}{\partial D(Z)} \frac{dD(Z)}{dZ} \frac{\partial Z}{\partial P}. \tag{16}$$

5. Evaluation

We give the proof of concept of the proposed algorithm on the model data sampled from a cylinder $\mathbb{S}^1 \times [-2, 2]$ with added Gaussian noise with zero mean and $\sigma^2 = 0.05$.

We used Python's `pymanopt` [39] package as the solver, `topologylayer` [9] package for the computation of the persistent homology mapping and its' gradient, and our own implementation of the Sinkhorn distance.

In the Figure 3 we show the example of the model data for which our algorithm and PCA give two different low dimensional representations, preserving the cycle in the former case and recovering the maximum variance in the latter. For the data expressing periodic behavior, for example the sliding window embeddings of periodic time series [33] preserving the global topology given by cycles in data could be of more importance then finding the maximum variance.

The graph of the convergence of the total loss function optimized by gradient descent along with the persistence diagrams of data $D(X)$ and low-dimensional projection $D(Z)$ for the consecutive steps of the optimization algorithm are shown in the Figure 4.

6. Conclusions

We have showed proof of concept of linear dimensionality reduction seeking to preserve the global structure of data. Our approach could be seen as generalization of the classic projection pursuit method designed to preserve clusters in data. Clusters are topological invariants of data of order 0, and the higher-order invariants include cycles bounding holes or voids. Preserving the higher-order may be of interest for data expressing periodic behavior.

The efficient numerical implementation scaling to hundreds of thousands of points in high-dimensional spaces applied to real data is left for the future work.

Acknowledgements

Part of this research was performed while the author was visiting the Institute for Pure and Applied Mathematics (IPAM), which is supported by the National Science Foundation (Grant No. DMS-1440415).

References

[1] Pierre-Antoine Absil, Robert E. Mahony, and Rodolphe Sepulchre. Optimization algorithms on matrix manifolds. In *Optimization Algorithms on Matrix Manifolds*, 2007. 2, 4

[2] Jason Altschuler, Jonathan Weed, and Philippe Rigollet. Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In *Advances in Neural Information Processing Systems*, pages 1964–1974, 2017. 5, 6

[3] Sergey Barannikov. The framed morse complex and its invariants. 1994. 2

[4] Ulrich Bauer. Ripser: efficient computation of vietoris-rips persistence barcodes. *arXiv preprint arXiv:1908.02518*, 2019. 3

[5] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15:1373–1396, 2002. 1

[6] Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and topological inference*, volume 57. Cambridge University Press, 2018. 3

[7] Karol Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fundamenta Mathematicae*, 35(1):217–234, 1948. 3

[8] Nicolas Boumal, Bamdev Mishra, Pierre-Antoine Absil, and Rodolphe Sepulchre. Manopt, a matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15:1455–1459, 2014. 2

[9] Rickard Brüel-Gabrielsson, Bradley J Nelson, Anjan Dwaraknath, Primoz Skraba, Leonidas J Guibas, and Gunnar Carlsson. A topology layer for machine learning. *arXiv preprint arXiv:1905.12200*, 2019. 3, 6

[10] Francesco Camastra and Antonino Staiano. Intrinsic dimension estimation: Advances and open problems. *Information Sciences*, 328:26–41, 2016. 2

[11] Mathieu Carriere, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 664–673. JMLR. org, 2017. 5

[12] John P. Cunningham and Zoubin Ghahramani. Linear dimensionality reduction: survey, insights, and generalizations. *Journal of Machine Learning Research*, 16:2859–2900, 2015. 2, 4

[13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013. 5, 6

[14] Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Analysis Applications*, 20:303–353, 1998. 2, 4

[15] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. American Mathematical Soc., 2010. 2, 3, 5

[16] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Proceedings 41st annual symposium on foundations of computer science*, pages 454–463. IEEE, 2000. 2

[17] Jerome H Friedman and John W Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on computers*, 100(9):881–890, 1974. 1

[18] Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya-Polo, and Tomaso A. Poggio. Learning with a wasserstein loss. In *NIPS*, 2015. 1, 2, 5

[19] Rickard Brüel Gabrielsson, Vignesh Ganapathi-Subramanian, Primoz Skraba, and Leonidas J Guibas. Topology-aware surface reconstruction for point clouds. *arXiv preprint arXiv:1811.12543*, 2018. 2

[20] Marcio Gameiro, Yasuaki Hiraoka, and Ippei Obayashi. Continuation of point clouds via persistence diagrams. *Physica D: Nonlinear Phenomena*, 334:118–132, 2016. 2, 3

[21] Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *Advances in neural information processing systems*, pages 5345–5355, 2018. 4

[22] Tingran Gao, Lek-Heng Lim, and Ke Ye. Numerical differential geometry. *XXI Householder Symposium on Numerical Linear Algebra*, 2020. 4

[23] Zhiwu Huang and Luc Van Gool. A riemannian network for spd matrix learning. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017. 4

[24] Peter J Huber. Projection pursuit. *The annals of Statistics*, pages 435–475, 1985. 1

[25] John M Lee. Smooth manifolds. In *Introduction to Smooth Manifolds*, pages 1–31. Springer, 2013. 4

[26] Jacob Leygonie, Steve Oudot, and Ulrike Tillmann. A framework for differential calculus on persistence barcodes. *arXiv preprint arXiv:1910.00960*, 2019. 2

[27] Giulia Luise, Alessandro Rudi, Massimiliano Pontil, and Carlo Ciliberto. Differential properties of sinkhorn approximation for learning with wasserstein distance. In *NeurIPS*, 2018. 4, 5

[28] Dougal Maclaurin, David Duvenaud, and Ryan P Adams. Autograd: Effortless gradients in numpy. In *ICML 2015 AutoML Workshop*, volume 238, 2015. 5

[29] Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The gudhi library: Simplicial complexes and persistent homology. In *International Congress on Mathematical Software*, pages 167–174. Springer, 2014. 3

[30] Nina Miolane, Johan Mathe, Claire Donnat, Mikael Jorda, and Xavier Pennec. geomstats: a python package for riemannian geometry in machine learning. *arXiv preprint arXiv:1805.08308*, 2018. 4

[31] Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. Topological autoencoders. *arXiv preprint arXiv:1906.00722*, 2019. 1

[32] Steve Y Oudot. *Persistence theory: from quiver representations to data analysis*, volume 209. American Mathematical Soc., 2015. 5

[33] Jose Perea and Chris Traile. Sliding windows and persistence. *The Journal of the Acoustical Society of America*, 141(5):3585–3585, 2017. 1, 6

[34] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. 2

[35] Adrien Poulenard, Primoz Skraba, and Maks Ovsjanikov. Topological function optimization for continuous shape matching. In *Computer Graphics Forum*, volume 37, pages 13–25. Wiley Online Library, 2018. 2, 3

[36] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290 5500:2323–6, 2000. 1

[37] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015. 2

[38] Joshua B. Tenenbaum, Vin De Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290 5500:2319–23, 2000. 1

[39] James Townsend, Niklas Koep, and Sebastian Weichwald. Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. *Journal of Machine Learning Research*, 17(137):1–5, 2016. 2, 6

[40] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008. 2, 4

[41] Lin Yan, Yaodong Zhao, Paul Rosen, Carlos Scheidegger, and Bei Wang. Homology-preserving dimensionality reduction via manifold landmarking and tearing. *arXiv preprint arXiv:1806.08460*, 2018. 1

[42] Ke Ye, Ken Sze-Wai Wong, and Lek-Heng Lim. Optimization on flag manifolds. *arXiv preprint arXiv:1907.00949*, 2019. 4

[43] Zhenyue Zhang and Hongyuan Zha. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. In *IDEAL*, 2002. 1

[44] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005. 3, 6