SIGNAL FILTERING AND PERSISTENT HOMOLOGY: AN ILLUSTRATIVE EXAMPLE

Mijail Guillemard and Armin Iske

Department of Mathematics, University of Hamburg,
{guillemard, iske}@math.uni-hamburg.de

ABSTRACT

During the last few years, recent applications of differential geometry and algebraic topology have provided powerful tools for the analysis of point cloud datasets $X = \{x_i\}_{i=1}^m \subset \mathbb{R}^n$. In particular, recent methods for nonlinear dimensionality reduction were inspired by fundamental concepts in differential geometry. In parallel developments, applied topology has delivered new methods for computing homological information of a point cloud data X. In this context, an important task is to understand the interaction of these novel tools with well-established signal analysis methods such as wavelets, Fourier transforms, etc. In this paper, we present illustrative examples describing topological effects when applying convolution filters to signals x_i in a dataset X. We use persistent homology as a main tool for measuring topological properties.

Keywords— Wavelets, STFT, modulation maps, persistent homology, nonlinear dimensionality reduction.

1. INTRODUCTION

In the last few years we have seen an important and fast development of new application tools for data analysis using concepts from algebraic topology and differential geometry. An important motivation for these developments is the increasing availability of inexpensive computer power allowing abstract concepts, traditionally from pure mathematics, to be used in modern application fields. Additionally, the increasing complexity of new engineering problems demands novel analysis tools and provides challenging tasks to the theoretical environment. Recent methods for nonlinear dimensionality reduction were inspired by fundamental concepts in differential geometry (e.g. isomap, local tangent space alignment, Riemannian normal coordinates, etc. [1, 5]) In parallel developments, applied algebraic topology has delivered robust algorithms for computing topological properties of a point cloud dataset $X = \{x_i\}_{i=1}^m$. In this context, a natural question is to understand the interaction of these novel tools with well-established signal analysis methods such as wavelet functions, short term Fourier transforms (STFT), etc.

In this paper, we illustrate some of these interactions with toy examples describing the interplay between filter operations in signal processing and topological features of a point cloud data. In this context, the notion of modulation maps plays a key role by constructing a set of signals as a point cloud data representing a low-dimensional object embedded in a high-dimensional Euclidean space.

The outline of this paper is as follows. In Section 2, we sketch our framework and some basic ideas for combining dimensionality reduction methods and signal transformations. Here, we also explain the concept of modulation maps, and we illustrate one of its key properties: the construction of a new type of challenging datasets $X = \{x_i\}_{i=1}^m$ for modern dimensionality reduction methods. In Section 3, we briefly recall elementary concepts of simplicial and persistent homology as the main tool for measuring topological features of a point cloud data $X = \{x_i\}_{i=1}^m$. Finally, in Section 4 we describe a toy example of a point cloud data X_f representing a signal f, and the interaction between filtering transformations on f, and homology measurements of X_f.

2. DIMENSIONALITY REDUCTION AND SIGNAL TRANSFORMS

In dimensionality reduction, we study a *point cloud data* defined as a finite family of vectors $X = \{x_i\}_{i=1}^m \subset \mathbb{R}^n$ located in an n-dimensional Euclidean space. The fundamental assumption is that X is sampled from $\mathcal{M}$, a (low-dimensional) *space* (manifold or topological space, e.g., CW-complex, simplicial complex) embedded in $\mathbb{R}^n$. We have therefore, $X \subset \mathcal{M} \subset \mathbb{R}^n$ with $p := \dim(\mathcal{M}) \ll n$. An additional key concept is the consideration of a *ideal model* representing $\mathcal{M}$, and denoted by Ω, embedded in a low-dimensional space $\mathbb{R}^d$ (with $d \ll n$), together with a homeomorphism (or ideally an isometry) $\mathcal{A} : \mathbb{R}^d \supset \Omega \to \mathcal{M} \subset \mathbb{R}^n$. The space Ω represents an ideal representation of $\mathcal{M}$ that could be used for analysis procedures in a low-dimensional environment. For instance, in the case of $\mathcal{M}$ being the well-know Swiss roll dataset, the space Ω is a rectangle. However, in practice, we can only try to approximate Ω with a dimensionality reduction map $P : \mathbb{R}^n \supset \mathcal{M} \to \Omega' \subset \mathbb{R}^d$, where Ω' is an homeomorphic copy of Ω.

Now we discuss the interactions of dimensionality reduction tools with signal transformations. A basic characteristic of short term Fourier analysis is the high dimensionality of the Euclidean space where the time-frequency data is embedded. In this context, for many applications a preprocessing step using dimensionality reduction methods can potentially improve the quality of the data analysis. More precisely, we consider a bandlimited signal $f \in L^2(\mathbb{R})$ and a segmentation of its domain in such a way that small consecutive signal patches are analyzed, as routinely performed in STFT or wavelet analysis. For instance, the set of signal patches X_f can be defined as a dataset of vectors in $\mathbb{R}^n$, derived by drawing n samples from a signal f:

$$X_f = \{x_i\}_{i=1}^m \subset \mathbb{R}^n, \quad x_i = (f(t_{k(i-1)+j}))_{j=0}^{n-1} \in \mathbb{R}^n,$$

for $k \in \mathbb{N}$ a fixed hop-size. Here, the regular sampling grid $\{t_\ell\}_{\ell=0}^{km-k+n-1} \subset \mathbb{R}$ is constructed when considering the Nyquist-Shannon theorem for f. This situation can naturally be related to the dimensionality reduction framework by considering X_f to be a subset of $\mathcal{M}$, a (low-dimensional) space, embedded in the high-dimensional Euclidean space $\mathbb{R}^n$. Therefore, we have $X_f \subset \mathcal{M} \subset \mathbb{R}^n$ with $p := \dim(\mathcal{M}) \ll n$. We recall that there is a well-known framework for studying properties of sets X_f in the context of nonlinear time series and dynamical systems (see e.g. [4]). But in our situation, we are additionally considering a close interaction with signal processing transforms T, together with specialized dimensionality reduction techniques P.

The construction of time-frequency data can be described as the application of a map $T : \mathcal{M} \supset X_f \to T(X_f) \subset \mathcal{M}_T$, where $\mathcal{M}_T := T(\mathcal{M})$, and $T(x_i)$ is the signal transformation of x_i (Fourier transform, wavelet, etc.). The following diagram shows the basic situation:

$$\begin{tikzcd}
\mathbb{R}^d \supset \Omega \arrow[r, "\mathcal{A}"] & \mathcal{M} \supset X_f \subset \mathbb{R}^n \arrow[d, "T"] \\
\mathbb{R}^d \supset \Omega' & \mathcal{M}_T \supset T(X_f) \subset \mathbb{R}^n \arrow[l, "P"']
\end{tikzcd}$$

Modulation Maps

We are interested in function analysis strategies that combine signal transformations (wavelet, Fourier etc.) with modern tools of applied topology and dimensionality reduction. In this context, we propose the notion of *modulation maps*, which summarizes the standard modulation concept in signal processing, using a geometrical and topological language. The fundamental objective of a modulation map is to construct spaces $\mathcal{M}$ using generating functions $\{\phi_k\}$ and a parametrization space Ω:

Definition (Modulation Maps [2,3])**.** Let $\{\phi_k\}_{k=1}^d \subset \mathcal{H}$ be a set of vectors in an Euclidean space $\mathcal{H}$, and $\{s_k : \Omega \to \mathcal{C}_\mathcal{H}(\mathcal{H})\}_{k=1}^d$ a family of smooth maps from a space Ω to $\mathcal{C}_\mathcal{H}(\mathcal{H})$ (the continuous functions from $\mathcal{H}$ into $\mathcal{H}$). We say that $\mathcal{M} \subset \mathcal{H}$ is a $\{\phi_k\}_{k=1}^d$*-modulated space* if

$$\mathcal{M} = \left\{ \sum_{k=1}^d s_k(\alpha)\phi_k, \ \alpha \in \Omega \right\}.$$

In this case, the map $\mathcal{A} : \Omega \to \mathcal{M}$, $\alpha \mapsto \sum_{k=1}^d s_k(\alpha)\phi_k$, is denoted *modulation map*.

An explicit example of this concept is given by a *frequency modulation map*, which considers $\phi(t) = \sin(t)$ and a modulation with the coordinates of points in Ω:

Example. Consider the map $\mathcal{A} : \Omega \to \mathbb{R}^n$ for a space $\Omega \subset \mathbb{R}^3$, with $\{t_i\}_{i=1}^n \subset [0, 1]$ and

$$\mathcal{A}_\alpha(t_i) = \sum_{j=0}^2 \sin((\alpha_c^j + \gamma\alpha_j)t_i), \quad \alpha = (\alpha_0, \alpha_1, \alpha_2) \in \Omega.$$

We construct each $\mathcal{A}_\alpha$ as a signal with n-samples, such that their Fourier representation has three prominent frequency bands centered at the values $\{\alpha_c^j\}_{j=0}^2$ (see Figure 1). We construct the parameter space such that $\Omega \subset [-1, 1]^3$, and we denote the value γ as the *bandwidth parameter*. When selecting a finite sampling $Y = \{y_i\}_{i=1}^m \subset \Omega$, we obtain a dataset $X = \{x_i\}_{i=1}^m = \mathcal{A}(Y)$, with $x_i = \mathcal{A}_{y_i}$. A key property of the point cloud data $X = \mathcal{A}(Y)$ is that they provide challenging examples for dimensionality reduction methods. More precisely, consider these datasets $X = \mathcal{A}(Y)$ to be described in terms of the bandwidth parameter γ. We remark that it is possible to prove that for some ranges of γ, $\mathcal{A}$ is an homeomorphism (diffeomorphism) into its image (see [2]). Additionally, it can be experimentally verified that for small values of γ, standard dimensionality reduction methods are able to correctly approximate Ω using X. But when increasing the parameter γ, the datasets $X = \mathcal{A}(Y)$ turn out to be challenging structures for many dimensionality reduction methods.

As an illustration, consider a function f such that each element of $X_f = \{x_i\}_{i=1}^m$ is of the type $x_i = \mathcal{A}_{y_i}$, for a modulation map $\mathcal{A}$ with parameter space Ω. For instance, in Figures 1, 2 and 3, we use the torus $\Omega = \mathbb{T}^2$, and we compare the resulting dimensionality reduction projections $P(X_f)$ for P being PCA and isomap.

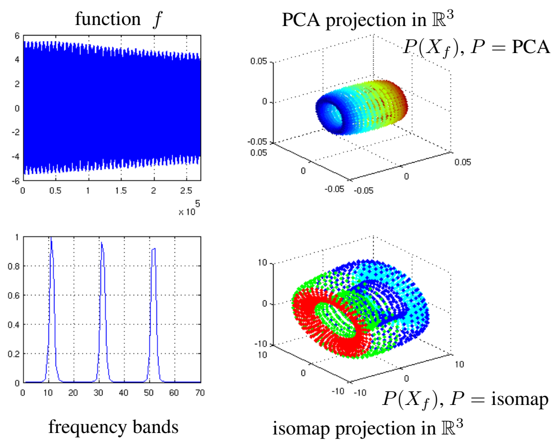

Fig. 1. Small γ: PCA and isomap recover $\Omega = \mathbb{T}^2$.

The example in Figure 1 illustrates how for a small bandwidth parameter γ, both PCA and isomap are able to recover Ω. But when increasing the bandwidth parameter γ, PCA is no longer able to recover the space Ω (Figure 2). Subsequently, isomap also breaks down when using even higher values of γ (Figure 3).

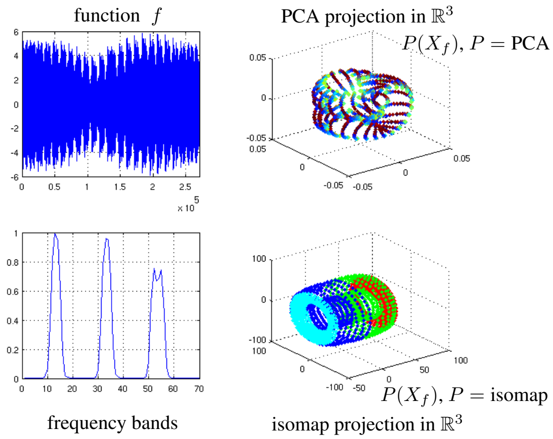

Fig. 2. Middle γ: PCA fails, but isomap recovers $\Omega = \mathbb{T}^2$.

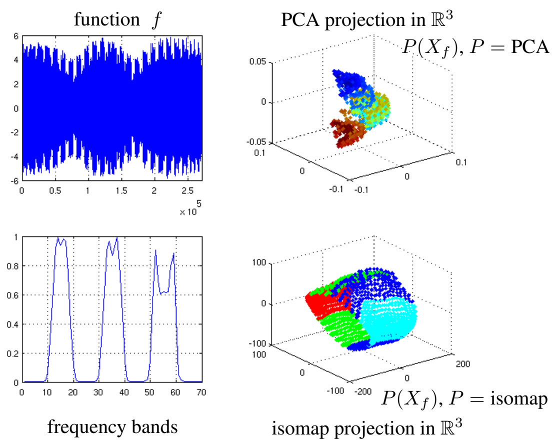

Fig. 3. Large γ: Both PCA and isomap fail to recover $\Omega = \mathbb{T}^2$.

3. SIMPLICIAL AND PERSISTENT HOMOLOGY

We now briefly explain basic concepts on persistent homology as an important tool for conceptually analyze our previous experiments. We first recall basic information on simplicial homology as a basic homology theory used for constructing algebraic data from topological spaces. A basic component in this context is a (finite) *abstract simplicial complex* which is a nonempty family of subsets K of a vertex set $V = \{v_i\}_{i=1}^m$ such that if $v \in V$ then $\{v\} \in K$, and if $\alpha \in K, \beta \subseteq \alpha$, then $\beta \in K$. The elements of K are denominated *faces*, and their *dimension* is defined as their cardinality minus one. Faces of dimension zero and one are called vertices and edges, respectively. A *simplicial map* between simplicial complexes is a function respecting their structural content by mapping faces in one structure to faces in the other. These concepts represent combinatorial structures capturing the topological properties of many geometrical constructions. Given an abstract simplicial complex K, an explicit topology is defined by considering a *geometric realization* or *polyhedron*, denoted by $|K|$, which is constructed by mapping faces to generalized versions of triangles or tetrahedrons in Euclidean spaces.

A basic analysis tool of a simplicial complex K, is the construction of algebraic structures for computing topological invariants, which are properties of $|K|$ that do not change under homeomorphisms. From an algorithmic point of view, we compute topological invariants of K by translating its combinatorial structure in the language of linear algebra. For this task, a basic scenario is to consider the following three steps. First, we construct the *module of k-chains* C_k, defined as the formal combinations of k-dimensional faces with coefficients in a ring. We then consider linear maps between the group of k-chains by constructing the *boundary operators* ∂_k, defined as the linear transformation which maps a face $\sigma = [p_0, \cdots p_n] \in C_n$ into C_{n-1} by $\partial_n \sigma = \sum_{k=0}^{n} (-1)^k [p_0, \cdots, p_{k-1}, p_{k+1}, \cdots p_n]$. As a third step, we construct the *homology groups* defined as the quotient $H_k := \ker(\partial_k)/\mathrm{im}(\partial_{k+1})$. Finally, the k-dimensional holes are defined as the rank of the homology groups, $\beta_k = \mathrm{rank}(H_k)$ (these are the Betti numbers). For instance, in the case of a sphere, we have zero one-dimensional holes, and one two-dimensional hole. In the case of a torus, there are two one-dimensional holes, and one two-dimensional hole.

Persistent Homology

A major problem when using the previous framework for studying a dataset $X = \{x_i\}_{i=1}^m$ is the fact that we do not have a simplicial complex structure at hand. Persistent homology (see [1]) provides a strategy for constructing topological information of a point cloud data. The fundamental idea is to construct a family of simplicial complexes by considering the spaces $\mathbb{X}_\epsilon = \cup_{i=1}^m B(x_i, \epsilon)$, where a ball $B(x_i, \epsilon)$ of radius $\epsilon > 0$ is centered around each point of the dataset X. Various well-known structures (e.g. Vietoris Rips complexes) are available for studying homological information of $\mathbb{X}_\epsilon$.

There are two crucial remarks for implementing these ideas in an efficient computational framework. On the one hand, despite the fact that we are considering a continuous parameter $\epsilon > 0$, it can be verified that for a given dataset X, there are actually only a finite number of non-homeomorphic simplicial complexes $K_1 \subset K_2 \subset \cdots \subset K_r$ derived from $\{\mathbb{X}_\epsilon, \epsilon > 0\}$. On the other hand, another crucial property is that there are efficient computational procedures for calculating homological information of the complete family $K_1 \subset K_2 \subset \cdots \subset K_r$ (see [1] for details).

The output of the persistent homology algorithm are diagrams representing the evolution, with respect to the parameter $\epsilon > 0$, of the topological features of X. For instance, in Figure 4, the two lower right plots are two persistent diagrams, where the red dots represent 1- and 2-dimensional holes, respectively. The dots located far from the diagonal represent stable features, while dots close to the diagonal are unstable and noise-like components. The dataset $P(X_f)$ of Figure 4 contains, in the second persistent diagram, a single dot, far away of the diagonal: this corresponds to a stable single two-dimensional hole. The first diagram contains dots only close the diagonal, and thus $P(X_f)$ has no one-dimensional holes. Therefore, the set from which $P(X_f)$ is sampled, is homeomorphic to a sphere.

4. FILTERING AND PERSISTENT HOMOLOGY

We now present an illustrative example of a filtering procedure and its interaction with topological measurements of a dataset X_f. We consider $f = (1-\alpha)g + \alpha h$, $\alpha \in [0,1]$ to be a sum of two functions g and h, where the datasets X_g and X_h are sampled from spaces homeomorphic to a sphere $\mathbb{S}^2$ and a torus $\mathbb{T}^2$, respectively.

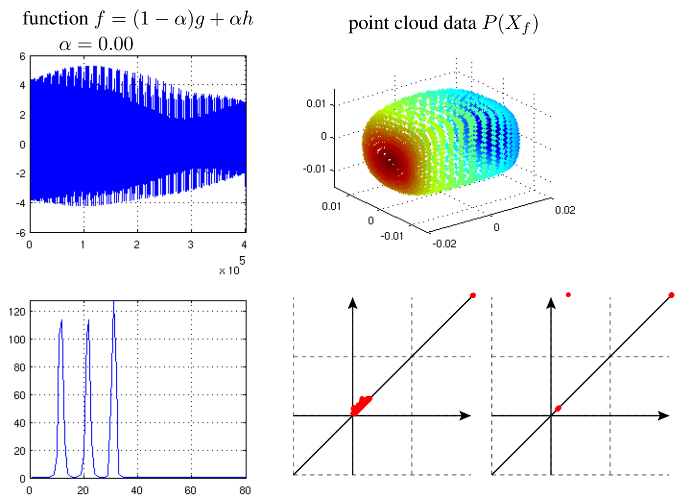

Fig. 4. $f = g$, and $X_f \subset \mathcal{M}$ with $\mathcal{M}$ homeomorphic to $\mathbb{S}^2$.

We construct X_g and X_h, as described in our examples on modulation maps in Section 2. For instance, in Figure 4, each element x of the point cloud data X_f is a signal whose main frequency content is located in three frequency bands depicted in the lower left diagram of Figure 4. We additionally design each element $x \in X_g$ and $y \in X_h$, such that their frequency content do not overlap. For example, in Figure 5, the lower diagram shows the six different frequency bands for the signal $x + y$: the first three bands corresponding to a typical element $x \in X_g$, and the other bands correspond to elements $y \in X_h$.

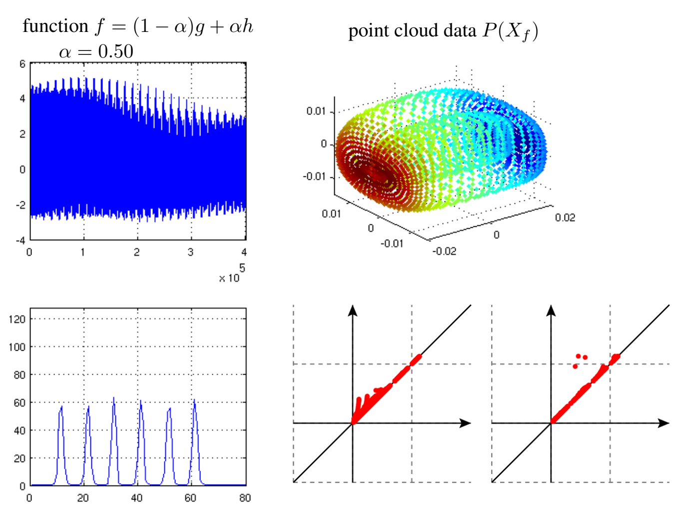

Fig. 5. $f = (g+h)/2$, and X_f as an intermediate structure.

For these examples, the variations of the parameter α corresponds to a filtering process, where we selectively remove (or add) the component g (or h) from the signal f. The topological effects can be seen by studying the persistent homology diagrams of X_f. For each Figure 4, 5, and 6, we have diagrams representing the first and second homology level. With this information we have an estimation for the number of one- and two-dimensional holes in X_f.

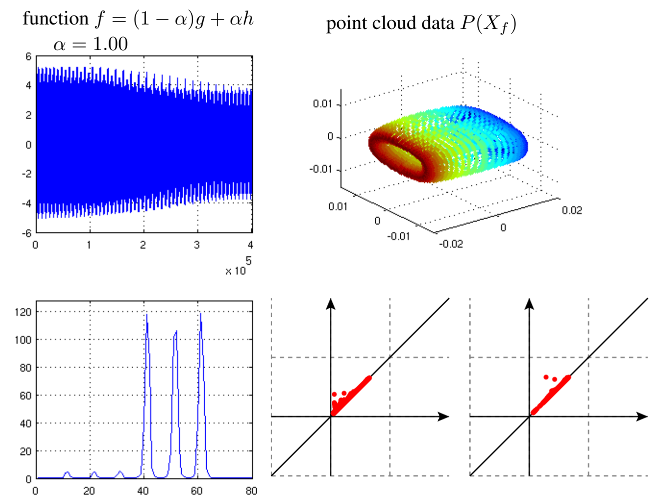

Fig. 6. $f = h$, and $X_f \subset \mathcal{M}$ with $\mathcal{M}$ homeomorphic to $\mathbb{T}^2$.

In the case of Figure 4, the persistent diagram for X_f shows a clear stable two-dimensional hole, and only noise-like one-dimensional holes. As previously mentioned, this corresponds to a spherical structure for X_f. Figure 6 shows two closely related, one-dimensional holes, and additionally two two-dimensional holes, which (approximately) corresponds to a torus structure. The persistent homology diagrams for the intermediate structure $X_{(g+h)/2}$ are depicted in Figure 5, where several two-dimensional holes are present.

5. REFERENCES

[1] G. Carlsson: Topology and data. *American Mathematical Society* **46**(2), 2009, 255–308.

[2] M. Guillemard and A. Iske: Curvature analysis of frequency modulated manifolds in dimensionality reduction. *Calcolo*. DOI: 10.1007/s10092-010-0031-8.

[3] M. Guillemard and A. Iske: Analysis of high-dimensional signal data by manifold learning and convolutions. In: *Sampling Theory and Applications (SampTA'09)*, L. Fesquet and B. Torrésani (eds.), Marseille (France), May 2009, 287–290.

[4] H. Kantz and T. Schreiber: *Nonlinear Time Series Analysis*. Cambridge University Press, 2004.

[5] J.A. Lee and M. Verleysen: *Nonlinear Dimensionality Reduction*. Springer, New York, 2007.