Inverse problems on persistence diagrams

Ippei Obayashi

Jan. 7, 2021

Abstract

In this talk, I will discuss a kind of inverse problem from a persistence diagram to a pointcloud. Since the solution is not unique, we need an additional constraint to remove the ambiguity. In our research we consider the problem to find a pointcloud whose k-th persistence diagram is our desired diagram and nearest to the given pointcloud. The key mathematical idea is piecewise differentiability of the map from pointclouds to persistence diagrams. This research is a joint work with M. Gamerio (Universidade de São Paulo, Brazil) and H. Hiraoka (Kyoto University, Japan) [1].

1 Introduction

Persistent homology (PH) is homology theory on filtrations, and it can summarize the shape of data quantitatively. PH is one of the most important tool of topological data analysis, and it is applied to various data analysis studies such as molecular phylogenetics [2], materials science [3], and structural biology [4].

Figure 1 shows the outline of PH. To characterize the shape of the input pointcloud from the viewpoint of topology, we put discs (for 2d) or balls (for 3d) with the same radius on all points (Fig. 1(a)). To capture the topological information of various scales, we consider the process of increasing balls as the arrow in Fig. 1(a) indicates. We analyze the appearances and disappearances of homology generators such as holes and cavities in the process. The theory of PH enables us to make a set of pairs (strictly saying, a multiset of pairs) of appearance and disappearance. The theory also ensures the uniqueness of the set of pairs and gives the algorithm to compute the pairs.

The radii of the appearance of homology generators are called *birth times*, and the radii of disappearance is called *death times*, and the pair of a birth time and a death time is called *birth-death pairs*. The multiset of all birth-death pairs is called a *persistence diagram (PD)* (Fig. 1(b)). For each birth-death pair, (death time) − (birth time) is called a *lifetime*.

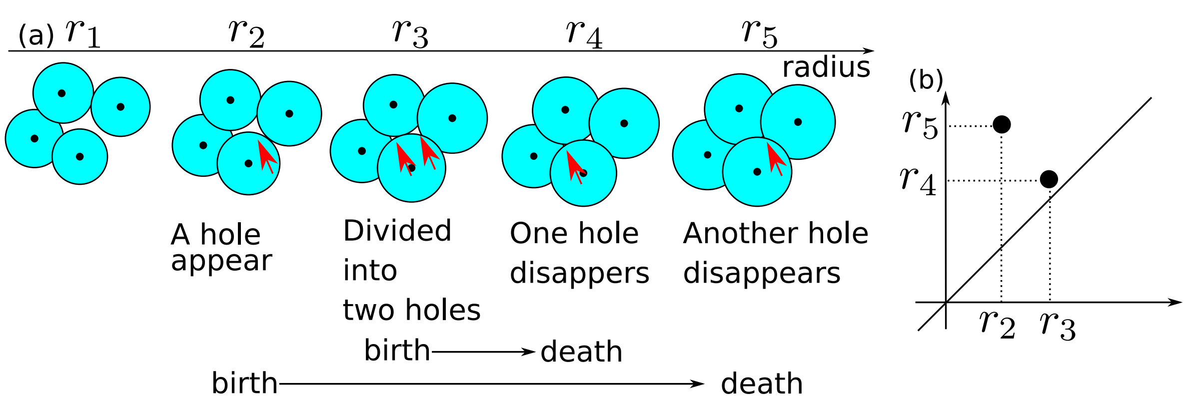

Figure 1: Outline of persistent homology

2 Inverse problem on PD

In some applications of PH to materials science [3], we applied PH to the molecular dynamical simulation data and we found that PH can capture some typical geometric structures of materials in atomic configurations and the important change the structures by the simulation conditions. The facts says that PDs correlate to the physical properties of the materials and it suggest the possibility to design materials from the viewpoint of PH. The first step to realize the PH-based materials design is to find the way to construct a pointcloud whose k-th persistence diagram is the desired diagram.

In fact the problem is ill-posed, it means that the solution is not unique. Here we consider the the additional constraint that the pointcloud should be as close as possible to the given pointcloud. The constraint comes from the original research question of materials science. We cannot realize any atomic configuration in nature, therefore it is natural to consider how to realize better atomic configuration with smallest displacement. Of course it is difficult to make such atomic configuration in the real world, but our PH-designed structure suggests an insight for new material design.

Finally we formalize the inverse problem from a PD to a pointcloud as follows.

Problem 1. *For a given PD, D_k, and a pointcloud, P_0, find a pointcloud P satisfying the following conditions:*

$$\textit{minimize } d(P, P_0), \ \textit{subject to } D_k(P) = D_k,$$

where d is a distance between pointcloud and $D_k(P)$ is the k-th persistence diagram of P.

Here we consider ℓ^2 norm as the distance by fixing the number of points N and the order of points. Finally the problem can be formalized as follows:

Problem 2. *For a given PD D_k and a pointcloud P_0, find a pointcloud P satisfying the following conditions:*

$$\textit{minimize } \|P - P_0\|_2, \ \textit{subject to } P \in \mathbb{R}^{3N} \textit{ and } D_k(P) = D_k. \tag{1}$$

Numerically solve the problem, we use the Newton-Raphson method by pseudo-inverse. The method solves the problem if D_k is a differentiable map from $\mathbb{R}^{3N}$ to $\mathbb{R}^M$. The first difficulty to apply the Newton-Raphson method is the fact that the number of pairs in $D_k(P)$ is not constant. We can avoid the difficulty by the stability theorem of PH. We can show the following fact.

Fact 1. *Let $D_k(P; L)$ be $\{(b, d) \in D_k(P) \mid d - b > L\}$, and $\tilde{P}$ be a pointcloud. We assume that any birth-death pair $(b, d) \in D_K(P; L)$ does not satisfy $L - \epsilon < b - d < L + \epsilon$. Then the number of pairs in $D_k(P; L)$ is the same as $D_k(\tilde{P}; L)$ if $\|P - \tilde{P}\|_\infty < \epsilon$, where $\|\cdot\|_\infty$ is the maximum norm on $\mathbb{R}^{3N}$. Moreover the map $P \mapsto D_k(P; L)$ is continuous in this neighborhood.*

This fact means that we can define the map with good properties by ignoring birth-death pairs with small lifetime from a persistence diagram.

The second difficulty is $D_k(P; L)$ is not differentiable. However we can show the following proposition.

Proposition 1. *If P satisfies the general position condition, $D_k(P; L)$ is differentiable in its small neighborhood. Moreover we can explicitly write the map by using the coordinates of points P. Each element of the map is a rational function.*

Finally we write the iteration step of the Newton-Raphson method as follows:

$$P^{(j+1)} = P^{(j)} - (dD_k(P^{(j)}; L))^\dagger D_k(P^{(j)}; L), \tag{2}$$

where $dD_k(P; L)$ is the total derivative of $D(P; L)$ given by Proposition 1 and $A^\dagger$ means the Moore–Penrose pseudo-inverse of matrix A.

3 Discussions and Conclusions

The key mathematical fact is the piecewise differentiability of the map $P \mapsto D_k(P; L)$. The differentiability is useful for not only Newton-Raphson method, but other optimization methods. We can apply gradient decent to persistence diagram. One example is the following minimization problem.

Problem 3. *For a given PD D_k and a pointcloud P_0, find a pointcloud P satisfying the following conditions:*

$$\textit{minimize } \|P - P_0\|_2^2 + \lambda \|D_k(P; L) - D_k\|_2^2, \tag{3}$$

where λ is a weight parameter.

This minimization problem gives the similar solution of Problem 1. In some sense we can regard $\|D_k(P; L) - D_k\|_2^2$ as "topological cost function".

Some studies after our study utilize the piecewise differentiability of persistence diagrams and the concept of topological cost function. Hu et al. [5] applied the idea to image processing. We can define persistence diagram on any dimensional bitmap data and the map from bitmaps to persistence diagrams is also piecewise differentiable if we ignore pairs with small lifetime. In the paper the authors construct a neural network for image segmentation with this topological cost function. The cost function penalize the topologically bad result. Integrating $D_k(P; L)$ to the neural network is not so difficult since piecewise differentiability is sufficient for backward propagation, fundamental neural network learning mechanism.

We also have an idea that we can utilize the topological cost function to material design. For the design, we need various physical constraints. To integrate such constraints with PH, we consider the following "total cost function" and minimize it.

$$\begin{aligned} &\text{(energetic cost function comes from physical property)} \\ &+\text{(material manufacturing cost function)} \\ &+\text{(topological cost function)} \\ &+ \cdots . \end{aligned} \tag{4}$$

This is just an idea, but I believe that integrating PH to various optimization problems has interesting applications.

References

[1] M. Gameiro, Y. Hiraoka, and I. Obayashi. Continuation of point clouds via persistence diagrams. Physica D: Nonlinear Phenomena, **334**(1), 118–132 (2016)

[2] J. M. Chan, G. Carlsson, and R. Rabadan. Topology of viral evolution. PNAS **110** (46), 18566–18571 (2013)

[3] A. Hirata, T. Wada, I. Obayashi and Y. Hiraoka. Structural changes during glass formation extracted by computational homology with machine learning. Communications Materials **1**, 98 (2020)

[4] T. Ichinomiya, I. Obayashi, and Y. Hiraoka. Protein folding analysis using features obtained by persistent homology. Biophysical Journal **118**, 2926–2937 (2020)

[5] X. Hu, F. Li, D. Samaras, and C. Chen. Topology-Preserving Deep Image Segmentation. Advances in Neural Information Processing Systems **32**, 5657–5668 (2019)