13 Primer in Persistent Homology II

We continue the introduction of persistent homology. One additional topic is the extension of persistence using relative homology to get measurements not only of the classes that are born and die but also of the essential classes that never die. The other topic is the stability of the diagram, which is essential in any effort to study the statistics of persistence diagrams.

Height function example. In almost every application of persistent homology, the filtration is given in terms of some function, and we do not draw the indices where births and deaths happen in the diagram, but rather the function values of these events. A prototypical example is a Morse function on a manifold, for which the sublevel set changes its homology only when we pass a critical value. Let \( M \) be a compact manifold, and \( f: M \to \mathbb{R} \) a Morse function; see Figure 20 for an example. Recall that \( f \) has only finitely many critical points and therefore only finitely many critical values. Furthermore, the sublevel sets have only finite rank homology groups.

![Figure 20: The height function on the torus-with-a-nose on the left, the Betti numbers in the middle, and the persistence intervals on the right.](image)

The infinite intervals represent the essential homology (that of \( M \)), while the finite intervals characterize the function and say nothing about the manifold. In Figure 20, there are four infinite intervals corresponding to the four generators of the essential homology of the torus, which has \( \beta_0 = 1, \beta_1 = 2, \beta_2 = 1 \).

Relative homology. It is sometimes useful to extend persistence to measure the essential homology. We do this using relative homology, which is defined for a pair of topological spaces, \( M_0 \subseteq M \). We begin with the relative chain groups:

\[
\cdots \xrightarrow{\partial_{p+1}} \mathbb{C}_p(M) \xrightarrow{\partial_p} \mathbb{C}_{p-1}(M_0) \xrightarrow{\partial_{p-1}} \cdots \tag{36}
\]

This amounts to ignoring differences in \( M_0 \). The boundary operators are still defined. Then

\[
Z_p(M, M_0) = \ker \partial_p, \quad B_p(M, M_0) = \text{im} \partial_{p+1}, \quad H_p(M, M_0) = Z_p/B_p \tag{37, 38, 39}
\]

Height function example, continued. Write \( M_a = f^{-1}(-\infty, a] \) and \( M^a = f^{-1}[a, \infty) \), calling them sublevel sets and superlevel sets of the function \( f \). For the superlevel sets, we have

\[
\cdots \xrightarrow{H_p(M, M^b)} H_p(M, M^a) \xrightarrow{H_p(M, M^a)} \cdots \tag{40}
\]

for \( a < b \). Figure 21 shows the Betti numbers of the relative homology groups of the pairs, and the persistence intervals.

![Figure 21: The down-phase of the height function during which we connect the relative homology groups.](image)

Extended persistence diagram. To get the extended persistence diagram, we glue the two sequences of homology groups together, calling the first half the up-phase and the second half the down-phase. Since the sequence starts with the trivial group and ends with the trivial group, every class that is born also dies. A class may be born and die going up, or it may be born going up and die coming down, or it may be born and die coming down. For each dimension, \( p \), we get a multi-set of points in the real plane, which we call the \( p \)-th persistence diagram of the function, denoted as \( \text{Dgm}_p(f) \). Very often, we will draw the
diagrams for different dimensions on top of each other, labeling the points with the dimension. In this case, we drop the dimension from the notation and denote the diagram by \( \text{Dgm}(f) \); see Figure 22.

![Figure 22: The extended persistence diagram of the height function on the torus-with-a-nose illustrated in Figures 20 and 21.](image)

The diagram distinguishes between the three cases by having the points in different domains; see Figure 22. Decomposing the center diamond into two triangles, we get four domains of the extended persistence diagram: the ordinary diagram (the lower triangle on the left), the horizontal diagram (the left center triangle), the vertical diagram (the right center triangle), and the relative diagram (the lower triangle on the right). The two middle triangles receive their names from the fact that the horizontal diagram contains classes that are visible within level sets of the function, while the vertical diagram contains classes that are not visible in level sets. We note that the diagram in Figure 22 is symmetric with respect to reflection across the vertical axis. This is not a coincidence. Indeed, Lefschetz duality, which applies to manifolds with boundary, implies that the extended persistence diagram of a function on a manifold (without boundary) is symmetric.

### Comparing diagrams

An important application of persistent homology is the possibility to compare two functions on the space by comparing the two persistence diagrams instead. Since the persistence diagram is a highly condensed representation, we should not expect that this could give a metric on the functions. Indeed, It is easy to design two different functions that have the same persistence diagram.

We compare two persistence diagrams using the bottleneck distance between them, which is the infimum, over all matchings between two diagrams, of the length of the longest edge in the matching. To cope with the possibility that the two diagrams do not have the same number of points, we allow for the use of extra points on the horizontal axis, to match with yet unmatched points, or to decrease the length of the longest edge. Formally, we assume that each persistence diagram contains each point on the horizontal axis an infinite number of times. Then

\[
d_B(F, G) = \inf_{\gamma} \sup_{A \in F} ||A - \gamma(A)||_1,
\]

where \( F \) and \( G \) are persistence diagrams, and the supremum is over all bijections \( \gamma: F \to G \). Recall that the coordinates of a point is the sum and the difference of birth- and death-values: \( A = (b + d, d - b) \) and \( X = (x + y, y - x) \). The \( L_1 \)-distance between the points is

\[
||A - X||_1 = |b + d - x - y| + |b - d - x + y| = 2 \max\{|b - x|, |d - y|\},
\]

which is twice the \( L_\infty \)-distance between the points \( (b, d) \) and \( (x, y) \). Note that the bottleneck distance is a stronger notion than the Hausdorff distance between the two multi-sets of points. Indeed, the Hausdorff distance does not imply a bijection that maps points not further than the measured distance.

### Stability

We use the bottleneck distance to formalize in which sense the persistence diagram is stable. Informally this means that when the function changes slightly then the diagram changes only slightly. Let \( f, g: \mathbb{X} \to \mathbb{R} \) be functions on a common topological space. We measure their difference as the maximum difference in function value:

\[
||f - g||_\infty = \sup_{x \in \mathbb{X}} |f(x) - g(x)|.
\]

Recall that the corresponding diagrams are \( F = \text{Dgm}(f) \) and \( G = \text{Dgm}(g) \). The diagrams are defined as long as the functions are tame, by which we mean that the homology groups of the sublevel and superlevel sets have all finite ranks, and there are only a finite number of values at which the homology groups change non-isomorphically.

**Stability Theorem.** Let \( f, g: \mathbb{X} \to \mathbb{R} \) be tame functions. Then \( d_B(\text{Dgm}_p(f), \text{Dgm}_p(g)) \leq ||f - g||_\infty \), for every dimension \( p \).

We emphasize that stability holds is great generality, for Morse functions on manifolds, but also for functions on non-manifold spaces. The original proof of this result can be found in \(^7\). It has since been generalized to settings that are even more general than stated above.