VI.4 Extended Persistence

In this section, we discuss an extension of persistence that is motivated by the problem of fitting shapes to each other. This arises when we solve a puzzle but also in protein docking, which is the attempt to predict protein interactions computationally.

Elevation. Let $\mathbb{M}$ be a smoothly embedded 2-manifold in $\mathbb{R}^3$. Given a direction $u \in \mathbb{S}^2$, the *height function* in this direction, $f = f_u : \mathbb{M} \to \mathbb{R}$, is defined by $f(x) = \langle x, u \rangle$. We usually draw u vertically going up and think of the height as the signed distance from a horizontal base plane, as in Figure VI.13. Given

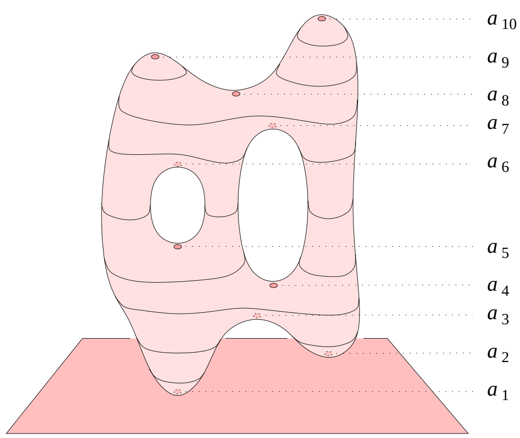

Figure VI.13: A smoothly embedded 2-manifold with level sets and critical points of the vertical height function marked.

a threshold $a \in \mathbb{R}$, we recall that the sublevel set consists of all points with height a or less, $\mathbb{M}_a = f^{-1}(-\infty, a]$. As mentioned in the previous section, the sublevel sets are nested and define persistence through the corresponding linear sequence of homology groups. For a generic smooth surface, the homological critical values of the height function are the height values of isolated points. We will be more precise about this in the next few sections where these points are referred to as the critical points. For now it will suffice to note that there are three different types of critical points, minima starting components, saddles merging components or completing loops, and maxima filling holes. Assuming the critical points have distinct heights, we can interpret the points in the persistence diagrams of f as pairs of critical points. We have minimum-saddle pairs in the dimension 0 diagram and saddle-maximum pairs in the dimension 1 diagram.

To define *elevation* on the embedded 2-manifold, $E : \mathbb{M} \to \mathbb{R}$, we consider all possible directions $u \in \mathbb{S}^2$. A point $x \in \mathbb{M}$ is critical for the height function in direction $u = \pm\mathbf{n}_x$, where $\mathbf{n}_x$ is the unit normal at x. If x is paired with another critical point y we define the elevation of x and y as their absolute height difference, $E(x) = E(y) = |f_u(x) - f_u(y)|$, where $u = \pm\mathbf{n}_x = \pm\mathbf{n}_y$. Since x is critical twice, for $u = \pm\mathbf{n}_x$, we need to make sure that the pairing is the same in both directions, else we get contradictory assignments of elevation. We also need all critical points to be paired, else we get white areas in which elevation remains undefined. The latter is the reason we extend persistence and the former is a constraint we need to observe in this extension.

Extended filtration. Let $a_1 < a_2 < \ldots < a_n$ be the homological critical values of the height function $f : \mathbb{M} \to \mathbb{R}$. At interleaved values $b_0 < b_1 < \ldots < b_n$ we get sublevel sets $\mathbb{M}_{b_i} = f^{-1}(-\infty, b_i]$ which are 2-manifolds with boundary. Symmetrically, we define *superlevel sets* $\mathbb{M}^{b_i} = [b_i, \infty)$ which are also 2-manifolds with boundary. We construct a sequence of homology groups going up and a sequence of relative homology groups coming back down,

$$\begin{array}{rcllcl} 0 & = & \mathsf{H}_p(\mathbb{M}_{b_0}) & \to \ldots \to & \mathsf{H}_p(\mathbb{M}_{b_n}) & \\ & = & \mathsf{H}_p(\mathbb{M}, \mathbb{M}^{b_n}) & \to \ldots \to & \mathsf{H}_p(\mathbb{M}, \mathbb{M}^{b_0}) & = \quad 0. \end{array}$$

The homomorphisms are induced by inclusion. We recall that for modulo 2 arithmetic the homology groups are isomorphic to the cohomology groups. Furthermore, Lefschetz duality implies $\mathsf{H}^p(\mathbb{M}_b) \simeq \mathsf{H}_{d-p}(\mathbb{M}, \mathbb{M}^b)$. This shows that the construction is intrinsically symmetric although not necessarily within the same dimension. Since we go from the trivial group to the trivial group, everything that gets born eventually dies. As a consequence, all critical points will be paired.

Tracing what gets born and killed in the relative homology groups is a bit less intuitive than for the absolute homology groups going up. However, we can translate the events between the absolute homology of $\mathbb{M}^b$ and the relative homology of the pair $(\mathbb{M}, \mathbb{M}^b)$. Coming down the threshold decreases so the superlevel set grows. Recall that a homology class in the superlevel set is essential if it lives all the way down to b_0.

Case 1. A dimension p homology class of $\mathbb{M}^b$ gets killed at the same time a dimension $p + 1$ relative homology class of $(\mathbb{M}, \mathbb{M}^b)$ gets killed.

Case 2. An inessential dimension p homology class of $\mathbb{M}^b$ gets born at the same time a dimension $p+1$ relative homology class of $(\mathbb{M}, \mathbb{M}^b)$ gets born.

Case 3. An essential dimension p homology class of $\mathbb{M}^b$ gets born at the same time a dimension p relative homology class of $(\mathbb{M}, \mathbb{M}^b)$ gets killed.

Example. We illustrate the extended filtration and the translation rules for the height function of the genus-2 torus of Figure VI.13. Going up, a_1 and a_2 give birth to classes in H_0, a_4, a_5, a_6, a_7, a_8 give birth to classes in H_1, and a_{10} gives birth to a class in H_2. All classes live until the end of the ascending pass, except for the dimension 0 class of a_2 which is killed by a_3 and the dimension 1 class of a_8 which is killed by a_9. These are the only two off-diagonal points in the persistence diagram as we used to know it. Coming down, a_{10} kills the class in H_0 and a_9 gives birth to a class in H_1 that is killed by a_8. Furthermore, a_7, a_6, a_5, a_4 kill the classes in H_1, a_3 gives birth to a class in H_2 that is killed by a_2, and finally a_1 kills the class in H_2. To summarize, the pairs of critical values defining the points in the diagrams are $(a_1, a_{10}), (a_2, a_3)$ in dimension 0, $(a_4, a_7), (a_5, a_6), (a_6, a_5), (a_7, a_4), (a_8, a_9), (a_9, a_8)$ in dimension 1, and $(a_{10}, a_1), (a_3, a_2)$ in dimension 2. We show the diagrams in Figure VI.14 using different symbols for points born and killed going up, born going up and killed coming down, and born and killed coming down. They make up the

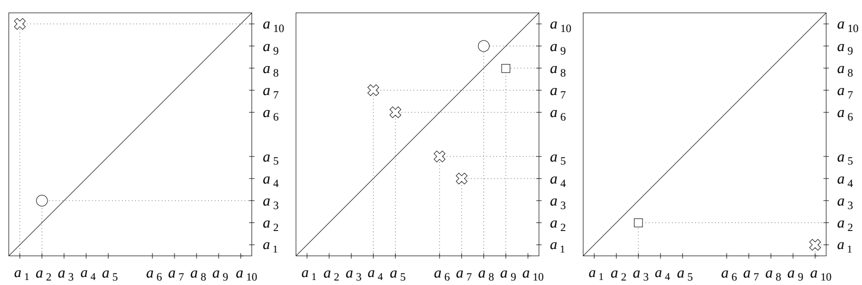

Figure VI.14: From left to right: the dimension 0, dimension 1, and dimension 2 persistence diagrams of the height function in Figure VI.13.

ordinary, the *extended*, and the *relative sub-diagrams*, which we denote as Ord, Ext, and Rel with the dimension in the index and the function in parenthesis, as before. Note that the points of the ordinary sub-diagram lie above and those of the relative sub-diagram lie below the diagonal. The points of the extended sub-diagram can lie on either side.

Duality and symmetry. The symmetries we observe in Figure VI.14 are not coincidental. They arise as consequences of the Lefschetz duality between absolute and relative homology groups of complementary dimensions, $\mathsf{H}_p(\mathbb{M}_b) \simeq \mathsf{H}_{d-p}(\mathbb{M}, \mathbb{M}^b)$. This translates into a duality result for persistence diagrams which we state without proof. We use a superscript 'T' to indicate reflection across the main diagonal.

DUALITY THEOREM. A continuous function f on a d-manifold has persistence diagrams that are reflections of each other as follows,

$$\begin{aligned} \mathrm{Ord}_p(f) &= \mathrm{Rel}^T_{d-p}(f); \\ \mathrm{Ext}_p(f) &= \mathrm{Ext}^T_{d-p}(f); \\ \mathrm{Rel}_p(f) &= \mathrm{Ord}^T_{d-p}(f). \end{aligned}$$

Equivalently, the full dimension p persistence diagram is the reflection of the full dimension $d-p$ persistence diagram, $\mathrm{Dgm}_p(f) = \mathrm{Dgm}^T_{d-p}(f)$. We have $d = 2$ for the example in Figures VI.13 and VI.14 and we indeed have diagrams that are reflections of each other as described. For $2p = d$ the extended sub-diagram is the reflection of itself and therefore symmetric across the main diagonal.

Recall that the definition of elevation requires the pairing of critical points be the same for opposing height functions. We can use duality to prove that they are indeed the same. More specifically, we have the following structural result again expressed in terms of sub-diagrams of the persistence diagrams and given without proof. We use the superscript 'R' to indicate reflecting along the minor diagonal.

SYMMETRY THEOREM. Continuous functions f and $-f$ on a d-manifold have persistence diagrams that are reflections of each other as follows,

$$\begin{aligned} \mathrm{Ord}_p(f) &= \mathrm{Ord}^R_{d-p-1}(-f); \\ \mathrm{Ext}_p(f) &= \mathrm{Ext}^R_{d-p}(-f); \\ \mathrm{Rel}_p(f) &= \mathrm{Rel}^R_{d-p-1}(-f). \end{aligned}$$

The reflection across the minor diagonal expresses the symmetry we require. A point (a, b) maps to $(-b, -a)$, reversing birth and death and changing the sign of each coordinate.

Lower and upper stars. To describe how we compute extended persistence, let K be a triangulation of a d-manifold $\mathbb{M}$. We assume the height function is defined at the vertices. We also assume that the height values are distinct so we can index the vertices such that $f(v_1) < f(v_2 < \ldots < f(v_n)$. Let $f : K \to \mathbb{R}$ be the continuous function on $\|K\|$ obtained by piecewise linear extension of the values at the vertices. Writing $a_i = f(v_i)$ and introducing interleaved values $b_0 < b_1 < \ldots < b_n$ we can define sublevel sets and superlevel sets as before. The set of points $x \in \|K\|$ with $f(x) \leq b_i$ is homeomorphic to $\mathbb{M}_{b_i}$ and thus a manifold with boundary. Similarly, the set of points with $f(x) \geq b_i$ is homeomorphic to $\mathbb{M}^{b_i}$ and a manifold with boundary. We can retract the partially used simplices and get homotopy equivalent subcomplexes of K. Specifically, let K_i be the full subcomplex defined by the first i vertices in the ordering along f and let K^i be the full subcomplex defined by the last $n - i$ vertices. The two subcomplexes of K are disjoint although together they cover all n vertices. The only simplices not in either subcomplex are the ones that connect the first i with the last $n - i$ vertices. Recall that the star of a vertex v_i consists of all simplices $\sigma \in K$ that have v_i as a vertex. The *lower star* is the subset of simplices for which v_i is the highest vertex and the *upper star* is the subset for which v_i is the lowest vertex,

$$\begin{aligned} \mathrm{St}_- v_i &= \{\sigma \in \mathrm{St}\, v_i \mid v \in \sigma \Rightarrow f(v) \leq f(v_i)\}, \\ \mathrm{St}^+ v_i &= \{\sigma \in \mathrm{St}\, v_i \mid v \in \sigma \Rightarrow f(v) \geq f(v_i)\}. \end{aligned}$$

Since every simplex has a unique highest vertex, the lower stars partition K. Similarly, the upper stars partition K. With this notation, $K_0 = \emptyset$ and $K_i = K_{i-1} \cup \mathrm{St}_- v_i$ for $1 \leq i \leq n$. Equivalently, K_i is the union of the first i lower stars. Symmetrically, $K^n = \emptyset$, $K^i = K^{i+1} \cup \mathrm{St}^+ v_{i+1}$, and K^i is the union of the last $n - i$ upper stars.

Computation. By construction, the K_i have the same homotopy type as the sublevel sets and the K^i have the same homotopy types as the superlevel sets of $\mathbb{M}$. We can therefore compute persistence by adding the simplices accordingly. Let A be the boundary matrix for the ascending pass, storing the simplices in blocks that correspond to the lower stars of v_1 to v_n, in this order. Within each block, we store the simplices in the order of non-decreasing dimension and break remaining ties arbitrarily. All simplices in the same block are assigned the same value, namely the height of the vertex defining the lower star. If two simplices in the same block are paired they therefore define a point on the diagonal of the appropriate persistence diagram. In other words, the homology class dies as soon as it is born and therefore has zero persistence. Only pairs between blocks carry any significance.

Let B be the boundary matrix for the descending pass, storing the simplices in blocks that correspond to the upper stars of v_n to v_1, in this order. Using A and B we form a bigger matrix by adding the zero matrix at the lower left and the permutation matrix that translates between A and B at the upper right, as in Figure VI.15. We can think of the result as the boundary matrix

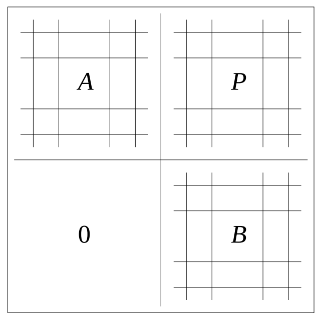

Figure VI.15: Schematic picture of the block structure representing the construction of K going up and the subsequence destruction coming down.

of a new complex, namely the cone over K. We pick a new, dummy vertex v_0 and for each i-simplex σ add the $(i+1)$-simplex $\sigma \cup \{v_0\}$. Adding the cone removes any non-trivial homology. This explains why reducing the big matrix works. As we move from left to right we first construct K forming pairs by reducing A. At the halfway point the only unpaired simplices are the ones that gave birth to the essential homology classes. As we continue we cone off K step by step, eventually removing all non-trivial homology. In the end, the ordinary, extended, and relative sub-diagrams are given by the lowest ones in the upper-left, upper-right, and lower-right quadrants of the reduced matrix.

Bibliographic notes. The extension of persistence described in this section is due to Cohen-Steiner et al. [2]. It makes essential use of Poincaré and Lefschetz duality to obtain the desired symmetry properties for manifolds. The construction applies equally well to general topological spaces but without guarantee of duality and symmetry. The main motivation for the extension is the elevation function introduced in [1] to help in the prediction of interactions between known protein structures.

[1] P. K. Agarwal, H. Edelsbrunner, J. Harer and Y. Wang. Extreme elevation on a 2-manifold. *Discrete Comput. Geom.* **36** (2006), 553–572.

[2] D. Cohen-Steiner, H. Edelsbrunner and J. Harer. Extended persistence using Poincaré and Lefschetz duality. Manuscript, Dept. Comput. Sci., Duke Univ., Durham, North Carolina, 2006.