Multidimensional persistence

A natural question is whether or not there exists a multi-dimensional generalization of persistence. The starting point for this generalization is a structure which filters the space in multiple dimensions simultaneously, a *multifiltration*. Our goal is to use such a structure to identify persistent features by examining the entire multifiltration. Such a generalization has been thought to have enormous potential in terms of applications []. For example, one of the often noted drawbacks of persistence is its instability with respect to strong outliers. That is, although persistence diagrams are stable with respect to their input and thus are robust to noise, the presence of outliers can artificially introduce topological features such as loops, obscuring the persistence of significant topological features [**?**]. Indeed, this obfuscation can hinder the ability of 1D persistence to uncover "significant" topological features, notably one of the most important applications of persistence. In theory, this formalism enables the practitioner to analyze even noisy data sets directly without resorting to sophisticated data preprocessing techniques which may or may not preserve the underlying topology of the data.

Unfortunately, invariants related to multidimensional persistence have proven challenging to use in practice due to the high algorithmic complexity of their corresponding computation. A number of computational strategies have been proposed to counter this, such as homology preserving multi-filtration simplification or parallelization schemes using shared-memory computation []. Another strategy is to focus specifically on the case of 2-D persistence. Utilizing the equivalence between the rank and fibered barcode invariants, Lesnick and Wright [] developed an elegant way of computing the former via a reparameterization of the latter using standard point-line duality. This clever reparameterization reduces the computation of the rank invariant to sequence of 1-D barcode computations at "template points" lying within the 2-cells of particular planar subdivision of the half-plane $[0, \infty) \times \mathbb{R}$.

The Rank Invariant

An *invariant* of n-parameter persistence modules is a function from the collection of persistence modules to some set S with the property that $f(M) = f(M')$ when $M \cong M'$. An invariant is said to be *complete* if $M \cong M'$ whenever $f(M) = f(M')$. Thus, incomplete invariants are invariants which may have the same value on two non-isomorphic modules. *The rank invariant* proposed by Carlsson et al [**?**] is one such invariant that is complete when $n = 1$ but is otherwise incomplete when $n \geq 2$. It is defined as follows:

Definition 1 (Rank Invariant). *For $n \geq 1$, the* rank invariant *of a $\mathbb{R}^n$-indexed persistence module M over $\mathcal{H}^n \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is the function:*

$$\mathrm{rank}(M) : \quad \mathcal{H}^n \to \mathbb{N}$$
$$(a, b) \mapsto \mathrm{rank}\, M(a, b)$$

for $a \leq b \in \mathcal{H}^n$. When $n = 1$, the rank invariant is complete, thus $\mathrm{rank}(M)$ and the persistence barcodes $\mathcal{B}(M)$ determine each other.

Thus, in one dimension, persistent homology provides a complete invariant that fully captures the lifetime of topological features in a given filtration: 1-D persistent homology modules decompose in an essentially unique way into indecomposable summands [**?**]. In contrast, the theory of multidimensional persistence shows that no complete discrete invariant exists—the structure of the target for the invariant depends on the structure of the underlying field, and although multiparameter persistence modules also decompose in a similar way as the 1-D case, the set of isomorphism classes is known to be extremely complicated [**?**, **?**]. Nonetheless, there are invariants related to multi-parameter persistence that can be quite useful in practice, even though they are incomplete. For example, even though the rank invariant does not encode the isomorphism type of the underlying module M when $n > 1$, it does captures important "first-order" information about the modules structure.

Along this vein of research, Lesnick and Wright [**?**] identify three simple invariants associated with a multidimensional persistence module M: the *dimension function of M*, the *multigraded Betti numbers of M*, and the *fibered barcode of M*. The fibered barcode, denoted as $\mathcal{B}(M)$, is defined as follows:

Definition 2 (Fibered barcode). *The fibered barcode $\mathcal{B}(M)$ of a bipersistence module M is the map which sends each line L with non-negative slope to the barcode $\mathcal{B}_L(M)$:*

$$L \mapsto \mathcal{B}_L(M)$$

Thus, the fibered barcode is simply a 2-parameter family of barcodes given by the collection of 1-D affine slices of M.

Each line L induces a 1-parameter restriction on M. The fibered barcode enjoys a number of stability properties readily expressed using interleaving distances, see []. More important to this work, the fibered barcode $\mathcal{B}(M)$ and the rank invariant $\mathrm{rank}(M)$ determine each other [].

Fibered barcode reparameterization

Let $\overline{\mathcal{L}}$ denote the collection of all lines in $\mathbb{R}^2$ with non-negative slope, $\mathcal{L} \subset \overline{\mathcal{L}}$ the collection of all lines in $\mathbb{R}^2$ with non-negative, finite slope, and $\mathcal{L}^\circ$ the collection of all affine lines in $\mathbb{R}^2$ with positive, finite slope. There is a standard point-line duality that gives a parameterization of $\mathcal{L}$ with the half-plane $[0, \infty) \times \mathbb{R}$ that is convenient to work with here. Define the *line* and *point* dual transforms $\mathcal{D}_\ell$ and $\mathcal{D}_p$, respectively, as follows:

$$
\begin{aligned}
\mathcal{D}_\ell : \mathcal{L} \to [0, \infty) \times \mathbb{R} \qquad & \mathcal{D}_p : [0, \infty) \times \mathbb{R} \to \mathcal{L} \\
y = ax + b \mapsto (a, -b) \qquad & (c, d) \mapsto y = cx - d
\end{aligned}
\tag{1}
$$

The transforms $\mathcal{D}_\ell$ and $\mathcal{D}_p$ are *dual* to each other in the sense that for any point $a \in [0, \infty) \times \mathbb{R}$ and any line $L \in \mathcal{L}$, $a \in L$ if and only if $D_\ell(L) \in D_p(a)$. Now, for some fixed line L, define the *push map* $\mathrm{push}_L(a) : \mathbb{R}^2 \to L \cup \infty$ as:

$$
\mathrm{push}_L(a) \mapsto \min\{v \in L \mid a \leq v\}
\tag{2}
$$

The push map satisfies a number of useful properties. Namely:

1. For $r < s \in \mathbb{R}^2$, $\mathrm{push}_L(r) \leq \mathrm{push}_L(s)$

2. For each $a \in \mathbb{R}^2$, $\mathrm{push}_L(a)$ is continuous on $\mathcal{L}^\circ$

3. For $L \in \mathcal{L}^\circ$ and $S \subset \mathbb{R}^2$, push_L induces a totally ordered partition S_L on S

Property (1) asserts that for any $L \in \overline{\mathcal{L}}$, the partial order on $\mathbb{R}^2$ restricts to a total order on L. Properties (2) and (3) qualify the following definition:

Definition 3 (Critical Lines). *For some fixed $S \subset \mathbb{R}^2$, a line $L \in L^\circ$ is defined to be* regular *if there is an open ball $B \in L^\circ$ containing L such that $S_L = S_{L'}$ for all $L' \in B$. Otherwise, the line L is defined as* critical.

The set of critical lines $\mathrm{crit}(M)$ with respect to some fixed set $S \subset \mathbb{R}^2$ fully characterizes a certain planar subdivision of the half plane $[0, \infty) \times \mathbb{R}$. Explicitly, define $\mathcal{A}^1(M)$ to the 1-skeleton of the line arrangement defined by:

$$
\mathcal{A}^1(M) = \{\mathcal{D}_\ell(\mathrm{crit}(M)) \cup (\{0\} \times \mathbb{R})\}
\tag{3}
$$

Expanding the 1-skeleton to a 2-D cell complex $\mathcal{A}(M)$ fully partitions the upper half plane. A consequence of equation 3 that if the duals of two lines $L, L' \in \mathcal{L}$ are contained in the same 2-cell in $\mathcal{A}(M)$, then $S_L = S_{L'}$, i.e. the partitions induced by push_L are equivalent. Indeed, the total order on S_L is the pullback of the total order on L with respect to the push map. Since $\mathcal{A}(M)$ partitions the entire half-plane, the dual to every line $L \in \mathcal{L}$ is contained within $\mathcal{A}(M)$—the desired reparameterization.

Let $S = \mathrm{supp}\,\xi_0(M) \cup \mathrm{supp}\,\xi_1(M)$, where the functions $\xi_0(M), \xi_1(M)$ are 0th and 1st bigraded Betti numbers of M, respectively. The main mathematical result from [**?**] is a characterization of the barcodes $\mathcal{B}_L(M)$, for any $L \in \mathcal{L}$, in terms of a set of *barcode templates* $\mathcal{T}$ computed at every 2-cell in $\mathcal{A}(M)$. More formally, for any line $L \in \overline{\mathcal{L}}$ and e any 2-cell in $\mathcal{A}(M)$ whose closure contains the dual of L under point-line duality, the 1-parameter restriction of the persistence module M induced by L is given by:

$$
\mathcal{B}_L(M) = \{[\,\mathrm{push}_L(a),\, \mathrm{push}_L(b)\,) \mid (a, b) \in \mathcal{T}^e, \mathrm{push}_L(a) < \mathrm{push}_L(b)\}
\tag{4}
$$

Minor additional conditions are needed for handling completely horizontal and vertical lines. The importance of this theorem for this effort lies in the fact that the fibered barcodes are completely defined from the precomputed barcode templates $\mathcal{T}$: once $\mathcal{T}$ has been computed and augmented onto $\mathcal{A}(M)$, $\mathcal{B}(M)$ is completely characterized, and the barcodes $\mathcal{B}_L(M)$ of any L-induced 1-D filtration can be efficiently computed via a point-location query on $\mathcal{A}(M)$ and a $O(|\mathcal{B}_L(M)|)$ application of the push map.
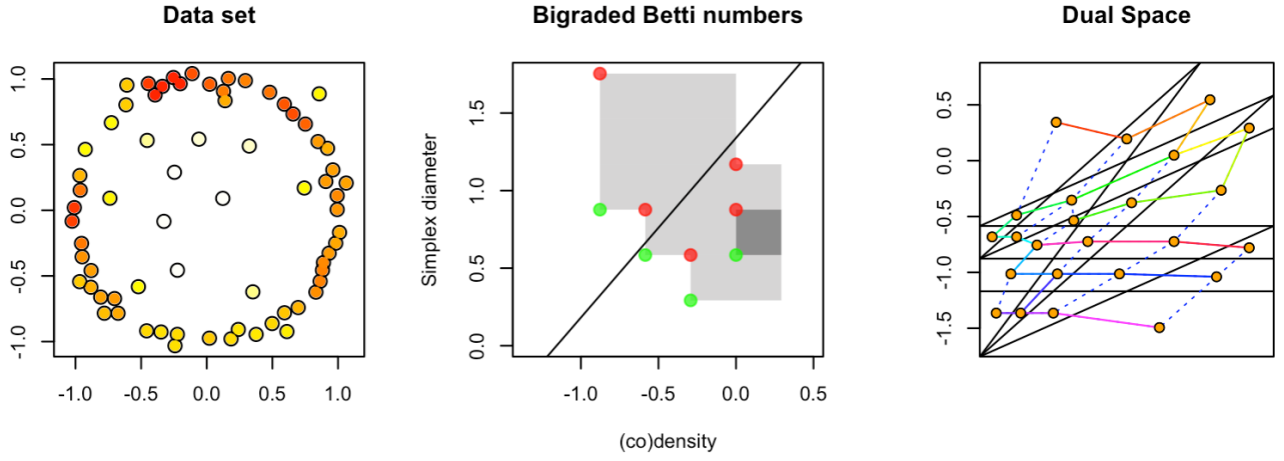
Figure 1: Bipersistence example on an 8×8 coarsened grid. On left, the input data, colored by density. In the middle, the bigraded Betti numbers $\xi_0(M)$ and $\xi_1(M)$ (green and red, respectively), the Hilbert function (gray), and a line L emphasizing the persistence of features with high density. On the right, the line arrangement $\mathcal{A}(M)$ lying in the dual space $D_p(\alpha)$ derived from the $\xi(M)$. The barycenters of the 2-cells in $\mathcal{A}(M)$ depicting where the barcodes templates $\mathcal{T}$ are computed are drawn as orange points, and the dual graph G of $\mathcal{A}(M)$ is drawn using dotted blue lines. The eulerian path through G is shown by the solid colored lines.