A Theory of Sub-barcodes

Oliver A. Chubet, Kirk P. Gardner, Donald R. Sheehy*

Abstract

From the work of Bauer and Lesnick, it is known that there is no functor from the category of pointwise finite-dimensional persistence modules to the category of barcodes and *overlap matchings*. In this work, we introduce *sub-barcodes* and show that there is a functor from the category of *factorizations* of persistence module homomorphisms to a poset of barcodes ordered by the *sub-barcode relation*. Sub-barcodes and factorizations provide a looser alternative to bottleneck matchings and interleavings that can give strong guarantees in a number of settings that arise naturally in topological data analysis. The main use of sub-barcodes is to make strong claims about an unknown barcode in the absence of an interleaving. For example, given only upper and lower bounds $g \geq f \geq \ell$ of an unknown real-valued function f, a sub-barcode associated with f can be constructed from ℓ and g alone. We propose a theory of sub-barcodes and observe that the subobjects in the functor category $\mathbf{Fun}(\mathbf{Int}^{\mathrm{op}}, \mathbf{Mch})$ naturally correspond to sub-barcodes.

1 Introduction

A persistence module is an algebraic object with a complete invariant known as a *barcode* that can be efficiently extracted from data. The algebraic stability theorem [1] states that an interleaving between persistence modules implies that the corresponding barcodes are close in bottleneck distance, providing a computable basis for comparison. However, when there is no interleaving, or only a very loose one, the existing theory provides only weak guarantees. In this paper, we address the following question:

> *Given a barcode, what can be said about the corresponding persistence module without an interleaving?*

Due to the isometry established by the algebraic stability theorem and its converse [2], there is no answer to this question in terms of the bottleneck distance. However, there are still strong guarantees to be derived from *half* an interleaving, where a "half interleaving" is a *factorization* of persistence module homomorphisms. This perspective leads to a natural subobject relation that can replace the bottleneck distance in some theoretical guarantees. The resulting theory of sub-barcodes addresses a fundamental question in the theory of persistence:

> *How can we reliably extract part of a barcode if we only have part of the data?*

A barcode $\mathcal{B}$ is represented by a collection of intervals with multiplicity, and a sub-barcode $\mathcal{A}$ of $\mathcal{B}$ is formed by either discarding or taking a subinterval of each interval of $\mathcal{B}$ (Figure 1; see Sections 2.3 and 3 for formal definitions). Using the induced matchings of Bauer and Lesnick [3, 4], we show how factorizations of persistence module homomorphisms induce sub-barcode matchings. As a result, we
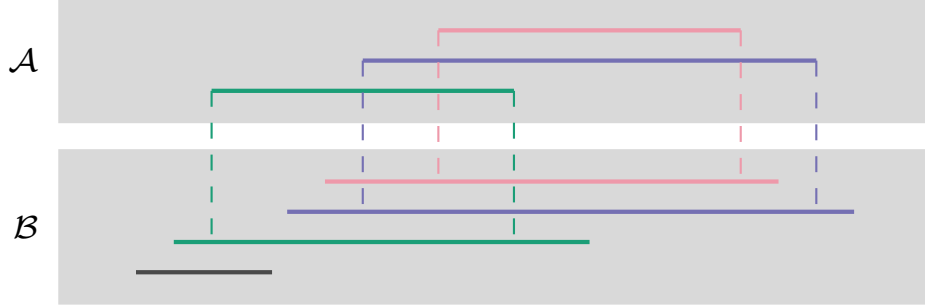

Figure 1: A sub-barcode matching.

can extract *part* of the barcode of a persistence module $\mathbb{U}$ using only a homomorphism $G : \mathbb{V} \to \mathbb{W}$ that factors through it. In particular, Theorem 3.3 implies that the barcode $\mathcal{B}_G$ of the image of G is a sub-barcode of $\mathcal{B}_{\mathbb{U}}$ if Diagram (1) commutes.

$$
\begin{array}{ccc}
\mathbb{V} & \xrightarrow{\;\;G\;\;} & \mathbb{W} \\
& \varphi_1 \searrow \quad \nearrow \varphi_2 & \\
& \mathbb{U} &
\end{array}
\tag{1}
$$

There are some natural cases in topological data analysis where this situation arises. We discuss two such cases in Section 4. Perhaps the simplest is the case when one has only upper and lower bounds on an unknown function f. Theorem 3.3 implies that the barcode associated with the inclusion of the upper bound into the lower bound is a sub-barcode of the barcode assocoiated with f. Figure 2 depicts this situation.
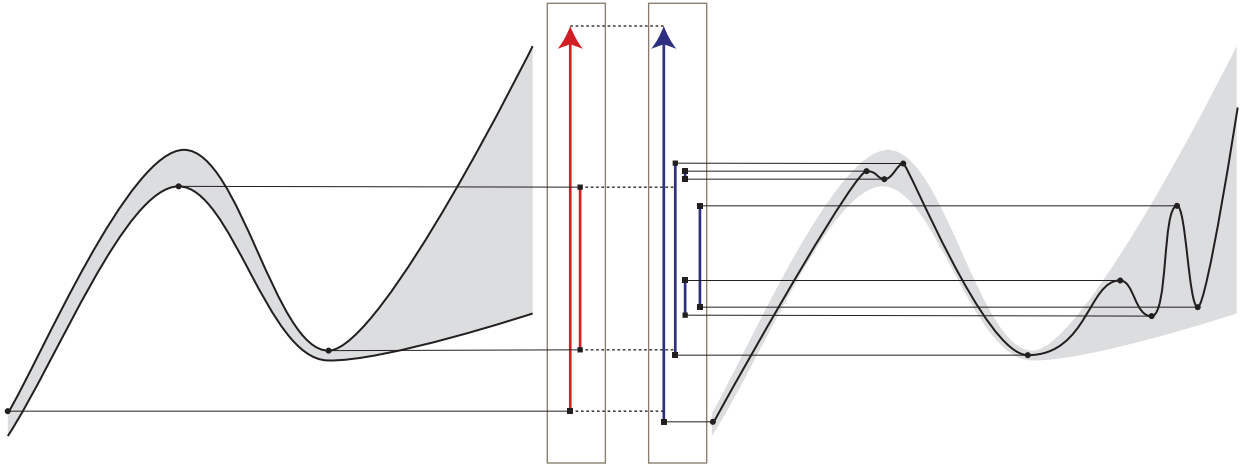

Figure 2: On the left, two functions are depicted, one is an upper bound and the other is a lower bound on an unknown function f. There is a corresponding barcode associated with the pair that matches minima in the upper bound to maxima in the lower bound. On the right is a candidate function f that lies between the upper and lower bounds and its barcode $\mathcal{B}_f$. The barcode of the inclusion of the upper and lower bounds is a sub-barcode of $\mathcal{B}_f$.

1.1 Related Work

Edelsbrunner et al. [5] first introduced an efficient algorithm for computing a persistence diagram (barcode) from a sequence of finite simplicial complexes in $\mathbb{R}^3$. This sequence, known as a *filtration*, can be easily computed from point-cloud data, and the resulting barcode captures the evolution of topological invariants known as the *betti numbers* β_k, each of which qualifies the k-dimensional connectivity of the underlying space. Zomorodian and Carlsson [6] later extended the theory to the persistent homology of simplicial complexes in $\mathbb{R}^d$ over any field.

Persistent homology has since seen a number of extensions and generalizations including extended persistence [7], Zigzag persistence [8], and persistent cohomology [9]. This is in addition to a number of foundational results in topological data analysis [10, 11, 12, 13] supported by significant improvements to the original persistence algorithm [14, 15, 16, 17, 18]. Most of these results, if not all, rely in some way on the the *stability theorem*.

Cohen-Steiner et al. [19] first showed that the bottleneck distance d_B between barcodes $\mathcal{B}_f, \mathcal{B}_g$ of real-valued functions $f, g : X \to \mathbb{R}$ is 1-Lipschitz with respect to the ℓ_∞-norm on the functions:

$$(stability) \qquad d_B\big(\mathcal{B}_f, \mathcal{B}_g\big) \leq \|f - g\|_\infty.$$

This result was quickly generalized by Chazal et al. [1] as the *algebraic stability theorem* which states that a δ-*interleaving* of pointwise-finite dimensional (p.f.d.) persistence modules implies a δ-bottleneck matching between barcodes, effectively generalizing the stability theorem from the ℓ_∞-norm on functions to the *interleaving distance* d_I on persistence modules $\mathbb{V}, \mathbb{W}$:

$$(algebraic\ stability) \qquad d_B\big(\mathcal{B}_\mathbb{V}, \mathcal{B}_\mathbb{W}\big) \leq d_I(\mathbb{V}, \mathbb{W}).$$

The converse algebraic stability theorem was then shown by Lesnick [2] to follow directly as a result of the structure theorem for p.f.d. [20] persistence modules, establishing a correspondence between the interleaving and bottleneck distances known as the *isometry theorem*:

$$(isometry) \qquad d_B(\mathcal{B}_\mathbb{V}, \mathcal{B}_\mathbb{W}) = d_I(\mathbb{V}, \mathbb{W}).$$

While persistence originated as a tool for topological data analysis (TDA), the categorification of persistence opens the door to more general applications beyond topology. Moreover, the perspective offered by categorification not only simplifies proofs and avoids redundancies, but also provides a bird's-eye-view from which gaps in the existing theory can be clearly identified. Bauer and Lesnick [3, 4] use this perspective to provide a novel constructive proof of the algebraic stability theorem. They show that the epi and mono morphisms in the category $\mathbf{vec}^\mathbf{R}$ of p.f.d. $\mathbf{R}$-persistence modules induce partial injective functions between their barcodes in the category $\mathbf{Mch}$ of *matchings* [21, 3, 22] (see Section 2.3). The algebraic stability theorem follows as a corollary of their induced matching theorem, which states that homomorphisms with ε-trivial (co)kernels induce matchings with ε-trivial (co)kernels (Theorem 6.1 [3], Theorem 1.4 [4]). In the follow-up [4], the authors restrict themselves to the category $\mathbf{Barc}$ of barcodes and overlap matchings (see Section 2.3) and show that this formulation is equivalent to the functor category $\mathbf{Mch}^\mathbf{R}$ that appears in the work of Edelsbrunner et al. [22] as *towers* of matchings.

Other categorical perspectives include the theses of Curry [23] and Lesnick [2], the formal categorification of persistence by Bubenik and Scott [24] and the follow-up work with de Silva [25], categorified reeb graphs by de Silva et al. [26], generalized persistence diagrams by Patel [27], and more recently a formal construction of homological algebra for persistence modules by Bubenik and Milićević [28].

The barcodes associated with the image of a persistence module homomorphism play a major role in this paper. The first algorithm to compute such a barcode from a nested pair of filtrations was due to Cohen-Steiner et al. [29]. The key idea has its antecedents in the work of Chazel and Lieutier [30]. Image persistence has found fruitful use in scalar field analysis [12] and clustering in manifolds [31]. More recently, Bauer and Schmahl developed an efficient algorithm for computing image persistence [32].

1.2 Overview

We begin with a review of the relevant background in Section 2. In Section 3 we define sub-barcode matchings and the sub-barcode relation. In Section 3.1, we show that there exists a functor from the category of factorizations of persistence modules to a poset of barcodes ordered by the sub-barcode relation. We then present the barcode functor in Section 3.2, and show that sub-barcode matchings correspond to subobjects in the resulting functor category. In Section 4, we discuss some practical applications of sub-barcodes to TDA.

2 Background

We begin with a review of the relevent categorical constructions to establish notation in Section 2.1 We direct the interested reader to Kashiwara and Schapira [33] and Mac Lane [34] for a full treatment. After formally defining persistence modules and their barcodes in Section 2.2, we consider barcodes more generally and review the induced matchings of Bauer and Lesnick [3, 4] in Section 2.3.

2.1 Categories

For any category $\mathbf{C}$ let 1 denote the identity and let $\circ$ denote composition. We write $b, c \in \mathbf{C}$ to denote objects $b, c \in \mathrm{Ob}(\mathbf{C})$ and $f : b \to c$ to denote arrows $f \in \mathrm{Hom}_{\mathbf{C}}(b, c)$ in $\mathbf{C}$, which may be referred to as morphisms $f \in \mathrm{Mor}(\mathbf{C})$. The opposite category $\mathbf{C^{op}}$ associated with $\mathbf{C}$ has objects $\mathrm{Ob}(\mathbf{C^{op}}) = \mathrm{Ob}(\mathbf{C})$ and arrows $\mathrm{Hom}_{\mathbf{C^{op}}}(c, b) = \mathrm{Hom}_{\mathbf{C}}(b, c)$ for all $b, c \in \mathbf{C}$. Two objects $b, c \in \mathbf{C}$ are related if there exists an arrow between them and connected if there exists a finite sequence of objects $b = x_0, \ldots, x_n = c$ in $\mathbf{C}$ such that, for all $i < n$, x_i and x_{i+1} are related. A category is small if its objects form a set, thin if there is at most one arrow between any two objects, and connected if every pair of objects is connected.

A partially ordered set (poset) $(P, \leq)$ is a small and thin category $\mathbf{P}$ with arrows denoted $b \leq c$ in which $b \leq c$ and $b \geq c$ implies $b = c$ for all $b, c \in \mathbf{P}$. A poset is a totally ordered set if every pair of objects are related. In the following, $\mathbf{P}$ will denote a partially ordered set, $\mathbf{T}$ will denote a totally ordered set, and $\mathbf{R}$ will denote the poset of real numbers $(\mathbb{R}, \leq)$.

Given a poset $\mathbf{P}$, a subset $S \subseteq \mathbf{P}$ is convex if for all $b, c \in S$, $b \leq t \leq c$ implies $t \in S$. A non-empty convex subset $I \subseteq \mathbf{P}$ is an interval in $\mathbf{P}$ if it is connected as a full subcategory of $\mathbf{P}$. Note that every convex subset of a totally ordered set is an interval. Let $\mathcal{I}_{\mathbf{P}}$ denote the *set* of intervals in $\mathbf{P}$, and let $\mathbf{Int_P}$ denote the poset $(\mathcal{I}_{\mathbf{P}}, \subseteq)$ of intervals in $\mathbf{P}$ ordered by inclusion.

The category $\mathbf{Set}$ has sets as objects and functions as arrows. The category $\mathbf{Rel}$ has sets A, B as objects and binary relations $R \subseteq A \times B$ as arrows. Let $\pi_1 R = \big\{ a \in A \mid (a, b) \in R \big\}$ and $\pi_2 R = \big\{ b \in B \mid (a, b) \in R \big\}$ denote the projections of R onto A and B, respectively. The composition of relations $R \subseteq A \times B$ and $S \subseteq B \times C$ in $\mathbf{Rel}$ is defined

$$S \circ R = \big\{ (a, c) \in A \times C \mid (a, b) \in R \text{ and } (b, c) \in S \text{ for some } b \in B \big\}.$$

The <u>transpose</u> or *reverse* $R^\mathsf{T} \subseteq B \times A$ of a relation $R \subseteq A \times B$ is defined

$$R^\mathsf{T} := \big\{(b, a) \in B \times A \mid (a, b) \in R\big\}.$$

Functor categories   A <u>functor</u> $F : \mathbf{B} \to \mathbf{C}$ between categories $\mathbf{B}$ and $\mathbf{C}$ is given by an *object map*

$$F(-) : \mathrm{Ob}(\mathbf{B}) \to \mathrm{Ob}(\mathbf{C})$$

and *arrow maps*

$$F[-] : \mathrm{Hom}_\mathbf{B}(b, c) \to \mathrm{Hom}_\mathbf{C}(F(b), F(c)),$$

for all $b, c \in \mathrm{Ob}(\mathbf{B})$ in which $F[1_b] = 1_{F(b)}$ for all $b \in \mathbf{B}$ and $F[g \circ f] = F[g] \circ F[f]$ for all composable pairs $f, g \in \mathrm{Mor}(\mathbf{B})$.

Given parallel functors $F, G : \mathbf{B} \to \mathbf{C}$, a <u>natural transformation</u> $\eta : F \Rightarrow G$ is given by a family of *components* $\eta_b : F(b) \to G(b)$ for each $b \in \mathbf{B}$ such that Diagram (2) commutes for all $f \in \mathrm{Mor}(\mathbf{B})$.

$$\begin{array}{ccc}
b & F(b) \xrightarrow{\ \eta_b\ } G(b) & \\
\downarrow{\scriptstyle f} & \Big\downarrow{\scriptstyle F[f]} \qquad \Big\downarrow{\scriptstyle G[f]} & \qquad(2)\\
c & F(c) \xrightarrow{\ \eta_c\ } G(c) &
\end{array}$$

A <u>functor category</u> $\mathbf{C}^\mathbf{B}$, which may be denoted $\mathbf{Fun}(\mathbf{B}, \mathbf{C})$ for notational purposes, has functors $F : \mathbf{B} \to \mathbf{C}$ as objects and natural transformations $F \Rightarrow G$ as arrows.

2.2   Persistence Modules

Working over a fixed field k, $\mathbf{Vec}$ is the category of vector spaces and linear maps and $\mathbf{vec}$ is the full subcategory of $\mathbf{Vec}$ restricted to finite-dimensional vector spaces. Given a poset $\mathbf{P}$, the functor category $\mathbf{Vec}^\mathbf{P}$ of <u>P-persistence modules</u> has functors $\mathbb{V} : \mathbf{P} \to \mathbf{Vec}$ as objects and natural transformations $F : \mathbb{V} \Rightarrow \mathbb{W}$ as arrows referred to as *homomorphisms*. A persistence module $\mathbb{V}$ is <u>pointwise finite-dimensional (p.f.d.)</u> if $\mathbb{V}(t)$ is finite-dimensional for all $t \in \mathbf{P}$. The full subcategory of $\mathbf{Vec}^\mathbf{P}$ restricted to p.f.d. persistence modules is $\mathbf{vec}^\mathbf{P}$.

Given an interval $I \in \mathcal{I}_\mathbf{P}$, an <u>interval module</u> k_I is a $\mathbf{P}$-persistence module defined

$$\mathsf{k}_I : \mathbf{P} \longrightarrow \mathbf{Vec}$$
$$(objects) \qquad\qquad t \longmapsto \mathsf{k} \text{ if } t \in I;\ 0 \text{ otherwise,}$$
$$(arrows) \qquad (s \leq t) \longmapsto 1_\mathsf{k} \text{ if } s, t \in I;\ 0 \text{ otherwise.}$$

The following definition subsumes the formulation of barcodes as multisets found in prior work.

Definition 2.1 (Barcode)**.** Given a poset $\mathbf{P}$, a <u>P-indexed barcode</u> is a function

$$\mathcal{B} : \underline{\mathcal{B}} \to \mathcal{I}_\mathbf{P},$$

where the set $\underline{\mathcal{B}}$ is the set of <u>bars</u> β that represent intervals $\mathcal{B}(\beta) \in \mathcal{I}_\mathbf{P}$.

Given a barcode $\mathcal{B}$, we can construct a persistence module as a direct sum of interval modules

$$\bigoplus_{\beta \in \underline{\mathcal{B}}} \mathsf{k}_{\mathcal{B}(\beta)}.$$

An interval decomposition of a persistence module $\mathbb{V} \in \mathbf{Vec}^\mathbf{P}$ is given by a barcode $\mathcal{B}_\mathbb{V} : \underline{\mathcal{B}_\mathbb{V}} \to \mathcal{I}_\mathbf{P}$ such that

$$\mathbb{V} \simeq \bigoplus_{\beta \in \underline{\mathcal{B}_\mathbb{V}}} \mathsf{k}_{\mathcal{B}_\mathbb{V}(\beta)}.$$

If it exists, then the interval decomposition (and therefore the barcode) of a persistence module indexed by a totally ordered set $\mathbf{T}$ is unique up to isomorphism. In particular, every p.f.d. $\mathbf{T}$-persistence module $\mathbb{V} \in \mathbf{vec}^\mathbf{T}$ has an interval decomposition [20], and therefore a barcode $\mathcal{B}_\mathbb{V}$. For any homomorphism $F : \mathbb{U} \to \mathbb{V}$ in $\mathbf{vec}^\mathbf{T}$ the image of F is a persistence module with a barcode that will be denoted $\mathcal{B}_F := \mathcal{B}_{\mathsf{im}\, F}$. In the following, we will refrain from restricting ourselves to a total order unless considering the barcode $\mathcal{B}_\mathbb{V} \in \mathbf{Bar}_\mathbf{T}$ of a p.f.d. persistence module $\mathbb{V} \in \mathbf{vec}^\mathbf{T}$.
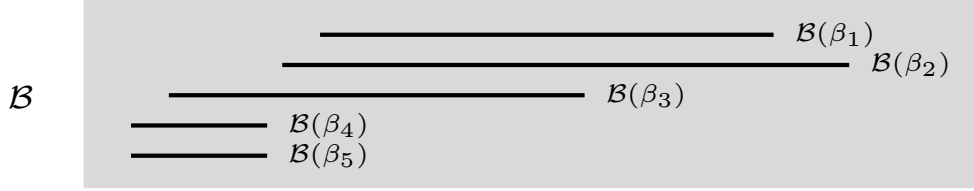

Figure 3: A $\mathbf{P}$-indexed barcode $\mathcal{B} : \underline{\mathcal{B}} \to \mathcal{I}_\mathbf{P}$ with $\underline{\mathcal{B}} = \{\beta_1, \ldots, \beta_5\}$.

2.3 Matchings and Induced Matchings

Given sets A and B, a matching $M : A \nrightarrow B$ is a relation $M \subseteq A \times B$ in which

(a) for all $a \in A$ there is at most one $b \in B$ with $(a, b) \in M$, and

(b) for all $b \in B$ there is at most one $a \in A$ with $(a, b) \in M$.

Let $\mathbf{Mch}$ denote the subcategory of $\mathbf{Rel}$ restricted to matchings. The mono (resp. epi) morphisms in $\mathbf{Mch}$ are matchings M with $\pi_1 M = A$ (resp. $\pi_2 M = B$). Matchings correspond to partial injective functions, and the mono (resp. epi) morphisms in $\mathbf{Mch}$ will therefore be referred to as injections (resp. coinjections).

A matching M of $\mathbf{P}$-indexed barcodes $\mathcal{A}$ and $\mathcal{B}$ is a matching of bars $(\alpha, \beta) \in M$ that *represent* intervals $\mathcal{A}(\alpha), \mathcal{B}(\beta) \in \mathcal{I}_\mathbf{P}$. Formally, a barcode matching is a matching $M : \underline{\mathcal{A}} \nrightarrow \underline{\mathcal{B}}$ in $\mathbf{Mch}$ such that $\mathcal{A}(\alpha) \cap \mathcal{B}(\beta) \neq \emptyset$ for all $(\alpha, \beta) \in M$. Let $\mathbf{Bar}_\mathbf{P}$ denote the category with $\mathbf{P}$-indexed barcodes as objects and arrows $M : \mathcal{A} \nrightarrow \mathcal{B}$ given by barcode matchings. Composition of matchings $M : \mathcal{A} \nrightarrow \mathcal{B}$ and $N : \mathcal{B} \nrightarrow \mathcal{C}$ in $\mathbf{Bar}_\mathbf{P}$ is defined

$$N \circ M = \big\{ (\alpha, \gamma) \in N \circ M \mid \mathcal{A}(\alpha) \cap \mathcal{C}(\gamma) \neq \emptyset \big\}.$$

The theory of induced matchings [3, 4] provides a canonical matching $\mathcal{B}_\mathbb{V} \nrightarrow \mathcal{B}_\mathbb{W}$ between the barcodes of persistence modules that is induced by a homomorphism $\mathbb{V} \to \mathbb{W}$. In particular, given a totally ordered set $\mathbf{T}$, the epi and mono morphisms in $\mathbf{vec}^\mathbf{T}$ induce canonical (co)injections between the corresponding barcodes in $\mathbf{Bar}_\mathbf{T}$ (Theorem 2.4). In Section 3.1, we use these matchings to define *induced sub-barcode matchings*.

Given a poset $\mathbf{P}$ and intervals $I, J \in \mathcal{I}_{\mathbf{P}}$, we say that J *bounds I below* if there exists some $a \in J$ with $a \leq t$ for all $t \in I$, and I *bounds J above* if there exists $d \in I$ with $t \leq d$ for all $t \in J$.

$$(J \ bounds \ I \ below) \qquad\qquad (I \ bounds \ J \ above)$$

I and J *coincide below* if J bounds I below and I bounds J below. I and J *coincide above* if I bounds J above and J bounds I above.

$$(I \ and \ J \ coincide \ below) \qquad\qquad (I \ and \ J \ coincide \ above)$$
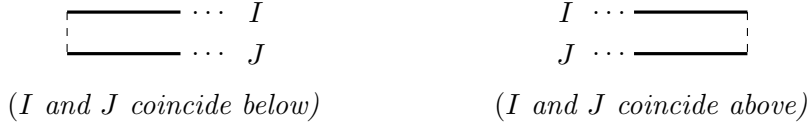
Definition 2.2 (Overlap Matching). <u>I overlaps J above</u> if $I \cap J \neq \emptyset$, J bounds I below, and I bounds J above.

$$(I \ overlaps \ J \ above)$$
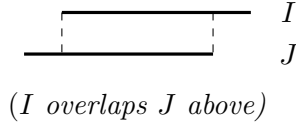
A matching $M : \mathcal{A} \nrightarrow \mathcal{B}$ in $\mathbf{Bar_P}$ is an <u>overlap matching</u> if $\mathcal{A}(\alpha)$ overlaps $\mathcal{B}(\beta)$ above for all $(\alpha, \beta) \in M$ (Figure 4).

Let $\mathbf{Barc_P}$ denote the subcategory of $\mathbf{Bar_P}$ restricted to overlap matchings. The following lemma implies that the composition of overlap matchings in $\mathbf{Bar_P}$ is an overlap matching.

Lemma 2.3. *If $M : \mathcal{A} \nrightarrow \mathcal{B}$ and $N : \mathcal{B} \nrightarrow \mathcal{C}$ are overlap matchings and $(\alpha, \gamma) \in N \circ M$, then there exists some $\beta \in \underline{\mathcal{B}}$ with $(\alpha, \beta) \in M$ and $(\beta, \gamma) \in N$ such that*

$$\mathcal{A}(\alpha) \cap \mathcal{C}(\gamma) \subseteq \mathcal{B}(\beta).$$

Proof. Assume $t \in \mathcal{A}(\alpha) \cap \mathcal{C}(\gamma)$. Because M is an overlap matching, $\mathcal{B}(\beta)$ bounds $\mathcal{A}(\alpha)$ below, so there exists some $b \in \mathcal{B}(\beta)$ such that $b \leq t$. Because N is an overlap matching, $\mathcal{B}(\beta)$ bounds $\mathcal{C}(\gamma)$ above, so there exists some $c \in \mathcal{B}(\beta)$ such that $t \leq c$. Because $\mathcal{B}(\beta)$ is an interval it is convex, so $b \leq t \leq c$ implies $t \in \mathcal{B}(\beta)$ for all $t \in \mathcal{A}(\alpha) \cap \mathcal{C}(\gamma)$. $\qquad \square$

The category $\mathbf{Barc}$ of barcodes and overlap matchings was first introduced by Bauer and Lesnick [4]. The following theorem is the critical result from their work that we will state in the larger category $\mathbf{Bar_T}$.

Theorem 2.4 (Theorem 4.2 [3], Theorem 1.1 [4]). *Let $\mathbf{T}$ be a totally ordered set. The mono and epi morphisms of $\mathbf{vec^T}$ induce canonical matchings in $\mathbf{Bar_T}$ that define functors*

(a) $\mathcal{J} : \mathrm{Mono}(\mathbf{vec^T}) \to \mathrm{Mono}(\mathbf{Bar_T})$ *taking monomorphisms $m : \mathbb{V} \hookrightarrow \mathbb{W}$ to injective overlap machings $\mathcal{J}[m] : \mathcal{B}_{\mathbb{V}} \hookrightarrow \mathcal{B}_{\mathbb{W}}$ in which, for all $(\alpha, \beta) \in \mathcal{J}[m]$,*

    (i) $\mathcal{B}_{\mathbb{V}}(\alpha)$ *and $\mathcal{B}_{\mathbb{W}}(\beta)$ coincide above and*

    (ii) $\mathcal{B}_{\mathbb{V}}(\alpha) \subseteq \mathcal{B}_{\mathbb{W}}(\beta)$;

Figure 4: An overlap matching in $\mathbf{Barc_P}$.

(b) $\mathcal{Q} : \mathrm{Epi}(\mathbf{vec^T}) \to \mathrm{Epi}(\mathbf{Bar_T})$ *taking epimorphisms* $e : \mathbb{V} \twoheadrightarrow \mathbb{W}$ *to coinjective overlap matchings* $\mathcal{Q}[e] : \mathcal{B}_{\mathbb{V}} \twoheadrightarrow \mathcal{B}_{\mathbb{W}}$ *in which, for all* $(\alpha, \beta) \in \mathcal{Q}[e]$,

(i) $\mathcal{B}_{\mathbb{V}}(\alpha)$ *and* $\mathcal{B}_{\mathbb{W}}(\beta)$ *coincide below and*

(ii) $\mathcal{B}_{\mathbb{V}}(\alpha) \supseteq \mathcal{B}_{\mathbb{W}}(\beta)$.
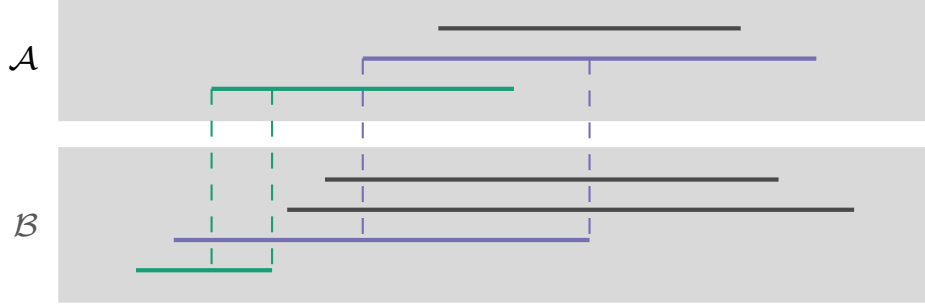
Notation 2.1. For any $F : \mathbb{V} \to \mathbb{W}$ in $\mathbf{vec^T}$ with an epi-mono factorization

$$\mathbb{V} \xrightarrow{q_F} \mathrm{im}\, F \xrightarrow{j_F} \mathbb{W},$$

let $\mathcal{Q}_F = \mathcal{Q}[q_F]$ and $\mathcal{J}_F = \mathcal{J}[j_F]$.

Remark 2.2. The matchings induced by mono and epi morphisms are given by *canonical injections* [3] that rely only on the *existence* of a mono or epi morphism between two persistence modules. That is, given parallel mono (resp. epi) morphisms $F, G : \mathbb{V} \rightrightarrows \mathbb{W}$ in $\mathbf{vec^T}$, $\mathcal{J}[F] = \mathcal{J}[G]$ (resp. $\mathcal{Q}[F] = \mathcal{Q}[G]$).
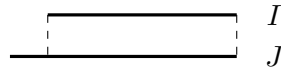

Figure 5: A matching given by $\mathcal{J}$ is an overlap matching of intervals $I \subseteq J$ that coincide above.

In the next section, we introduce sub-barcode matchings $M : \mathcal{A} \twoheadrightarrow \mathcal{B}$ in which $\mathcal{A}(\alpha) \subseteq \mathcal{B}(\beta)$ for all $(\alpha, \beta) \in M$. Although overlap matchings in $\mathbf{Bar_T}$ given by $\mathcal{J}$ (Figure 5) and the *reverse* of overlap matchings given by $\mathcal{Q}$ are sub-barcode matchings, not all sub-barcode matchings are overlap matchings (Figure 6). This is the main motivation for working in the larger category $\mathbf{Bar_T}$ containing both overlap and sub-barcode matchings.
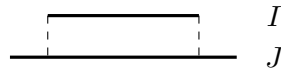
Figure 6: A sub-barcode matching is an injective matching of intervals $I \subseteq J$.

3 Sub-barcodes and Barcode Functors

Definition 3.1 (Sub-barcode Matchings and the Sub-barcode Relation)**.** Given $\mathcal{A}, \mathcal{B} \in \mathbf{Bar_P}$, a matching $M : \mathcal{A} \twoheadrightarrow \mathcal{B}$ in $\mathbf{Bar_P}$ is a sub-barcode matching if $\mathcal{A}(\alpha) \subseteq \mathcal{B}(\beta)$ for all $(\alpha, \beta) \in M$. If there

exists an *injective* sub-barcode matching from $\mathcal{A}$ to $\mathcal{B}$, we say that $\mathcal{A}$ is a sub-barcode of $\mathcal{B}$, and denote $\mathcal{A} \sqsubseteq \mathcal{B}$.

Let $\mathbf{SubBar_P}$ denote the subcategory of $\mathbf{Bar_P}$ restricted to sub-barcode matchings. $\mathbf{SBar_P}$ is the poset of isomorphism classes of barcodes in $\mathbf{SubBar_P}$ ordered by the sub-barcode relation.

In Section 3.1 we show how the induced matchings of Bauer and Lesnick [4] can be used to define a functor from the category $\mathbf{Fac}(\mathbf{vec^T})$ of *factorizations* of persistence modules to the poset $\mathbf{SBar_T}$. We then introduce the *barcode functor* $\mathrm{bar}_\mathcal{B} : \mathbf{In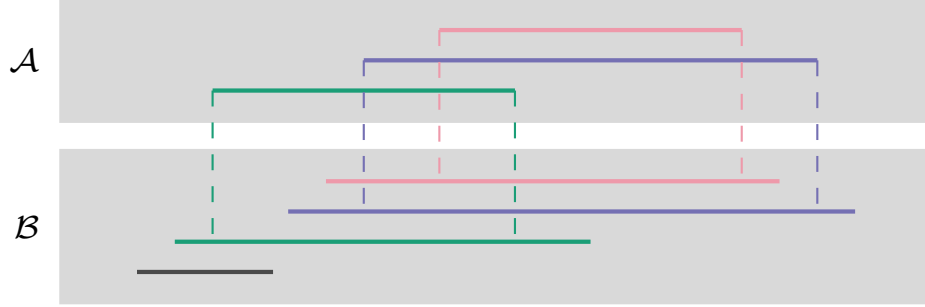t_P^{op}} \to \mathbf{Mch}$ associated with a barcode $\mathcal{B}$, and show that sub-barcodes correspond to subobjects in the functor category $\mathbf{Fun}(\mathbf{Int_P^{op}}, \mathbf{Mch})$ in Section 3.2.

3.1 Factorizations and Induced Sub-barcode Matchings

Though not explicitly named, factorizations are used throughout the existing theory of persistence. For example, letting $\mathbb{V}(\epsilon)$ denote a persistence module $\mathbb{V} \in \mathbf{Vec^R}$ shifted by a constant $\epsilon \in \mathbb{R}$, a 2ϵ-interleaving is traditionally given by a pair of commuting diagrams

$$
\begin{array}{cc}
\begin{array}{c}
\mathbb{V}(-\epsilon) \xrightarrow{\ \mathbb{V}\varepsilon\ } \mathbb{V}(\epsilon) \\
{\scriptstyle \Phi(-\epsilon)} \searrow \quad \nearrow {\scriptstyle \Psi} \\
\mathbb{U}
\end{array}
& (3)
\end{array}
\qquad
\begin{array}{cc}
\begin{array}{c}
\mathbb{V} \\
{\scriptstyle \Psi(-\epsilon)} \nearrow \quad \searrow {\scriptstyle \Phi} \\
\mathbb{U}(-\epsilon) \xrightarrow[\ \mathbb{U}\varepsilon\ ]{} \mathbb{U}(\epsilon)
\end{array}
& (4)
\end{array}
$$

depicting factorizations of homomorphisms $\mathbb{V}\varepsilon : \mathbb{V}(-\epsilon) \to \mathbb{V}(\epsilon)$ and $\mathbb{U}\varepsilon : \mathbb{U}(-\epsilon) \to \mathbb{U}(\epsilon)$ by a pair

$$\big(\Phi : \mathbb{V} \to \mathbb{U}(\epsilon),\ \Psi : \mathbb{U} \to \mathbb{V}(\epsilon)\big).$$

In this section, we will show that the commutativity of Diagrams (3) and (4) imply sub-barcode relations $\mathcal{B}_{\mathbb{V}\varepsilon} \sqsubseteq \mathcal{B}_{\mathbb{U}}$ and $\mathcal{B}_{\mathbb{V}} \sqsupseteq \mathcal{B}_{\mathbb{U}\varepsilon}$.

Definition 3.2 (Category of Factorizations (Exercise IX.6.3 [34], Definition 2.1.15 [33])). Given a category $\mathbf{C}$, the category of factorizations is a category $\mathbf{Fac}(\mathbf{C})$ whose objects are the morphisms of $\mathbf{C}$ and arrows $\varphi : g \rightleftharpoons f$ are the *factorizations in* $\mathbf{C}$, which are commutative diagrams:

$$
\begin{array}{ccc}
a & \xrightarrow{\ \varphi_1\ } & b \\
{\scriptstyle g} \downarrow & \rightleftharpoons & \downarrow {\scriptstyle f} \\
d & \xleftarrow[\ \varphi_2\ ]{} & c
\end{array}
\qquad \text{where}\ \ \varphi := (\varphi_1, \varphi_2) : g \rightleftharpoons f. \qquad (5)
$$

As an example, the category **Fac(P)** of factorizations of a poset **P** is a category with ordered pairs $(b \leq c)$ as objects that can be interpreted as intervals. Moreover, **Fac(P)** is a poset with factorizations $J \rightleftharpoons I$ corresponding to inclusions of intervals $J \supseteq I$.

$$
\begin{array}{c}
\overset{b}{\rule{0pt}{0pt}} \rule{120pt}{0.5pt} \overset{c}{\rule{0pt}{0pt}} \quad I \\
\underset{a}{\rule{0pt}{0pt}} \rule{160pt}{0.5pt} \underset{d}{\rule{0pt}{0pt}} \quad J
\end{array}
$$

The objects in the category $\mathbf{Fac(vec^T)}$ are homomorphisms $F : \mathbb{U} \to \mathbb{V}$, $G : \mathbb{T} \to \mathbb{W}$ and the arrows are factorizations $\varphi : G \rightleftharpoons F$ given by commuting diagrams

$$
\begin{array}{ccc}
& \mathbb{U} \xrightarrow{\ F\ } \mathbb{V} & \\
{}^{\varphi_1}\nearrow & & \searrow{}^{\varphi_2} \\
\mathbb{T} \xrightarrow{\hspace{3em} G \hspace{3em}} & & \mathbb{W}.
\end{array} \tag{6}
$$

The existence of a factorization $G \rightleftharpoons F$ implies a sub-barcode relation $\mathcal{B}_G \sqsubseteq \mathcal{B}_F$. This fact is stated formally in Theorem 3.3. The proof will follow directly from the construction of induced sub-barcode matchings (Definition 3.4, Lemma 3.5).

Theorem 3.3. *Given a totally ordered set* $\mathbf{T}$*, the map* $F \mapsto \mathcal{B}_F$ *defines a functor*

$$
\mathbf{Fac(vec^T)} \to \mathbf{SBar_T}.
$$

To simplify notation, we will provide shorthand for the various restrictions involved in the epi-mono factorizations associated with a factorization (Diagram 7).

Notation 3.1. Given a factorization $\varphi : G \rightleftharpoons F$ in $\mathbf{Fac(vec^T)}$, let $F_\varphi : \operatorname{im} \varphi_1 \to \operatorname{im} F$ and $\varphi_F : \operatorname{im} F \to \operatorname{im} \varphi_2$ denote the restrictions

$$
F_\varphi := F|_{\operatorname{im} \varphi_1} \quad \text{and} \quad \varphi_F := \varphi_2|_{\operatorname{im} F}.
$$

Because φ is a factorization of G through F we have that $G = \varphi_2 \circ F \circ \varphi_1$, thus

$$
\operatorname{im} G = \operatorname{im} (\varphi_2 \circ F \circ \varphi_1) = \operatorname{im} \varphi_2|_{\operatorname{im} F \circ \varphi_1} = \operatorname{im} \varphi_2|_{\operatorname{im} F_\varphi}.
$$

Let $\varphi_* : \operatorname{im} F_\varphi \to \operatorname{im} \varphi_F$ denote the restriction

$$
\varphi_* := \varphi_2|_{\operatorname{im} F_\varphi}
$$

so that $\operatorname{im} \varphi_* = \operatorname{im} G$.

The following diagram of epi-mono factorizations is given by the commutativity of Diagram (6):

$$\tag{7}$$

10

Definition 3.4 (Induced Sub-barcode Matchings)**.** Given a factorization $\varphi : G \rightleftharpoons F$ in $\mathbf{Fac}(\mathbf{vec}^{\mathsf{T}})$, the <u>induced sub-barcode matching</u> $\mathcal{M}_\varphi : \mathcal{B}_G \hookrightarrow \mathcal{B}_F$ of φ is an injective matching in $\mathbf{Bar_T}$ defined as the composition

$$\mathcal{M}_\varphi := \mathcal{J}_{F_\varphi} \circ \mathcal{Q}_{\varphi_*}^{\mathsf{T}};$$

the <u>induced *super*-barcode matching</u> $\mathcal{E}_\varphi : \mathcal{B}_F \twoheadrightarrow \mathcal{B}_G$ of φ is a coinjective matching in $\mathbf{Bar_T}$ defined as the composition

$$\mathcal{E}_\varphi := \mathcal{J}_{\varphi_*}^{\mathsf{T}} \circ \mathcal{Q}_{\varphi F}.$$

$$
\begin{array}{ccccc}
 & \xrightarrow{\mathcal{J}_{F_\varphi}} \mathcal{B}_F & & \mathcal{B}_F \xrightarrow{\;\;\mathcal{Q}_{\varphi F}\;\;} & \\
\mathcal{B}_{F_\varphi} & \Big\uparrow \mathcal{M}_\varphi & \mathcal{E}_\varphi \Big\downarrow & & \mathcal{B}_{\varphi F} \\
 & \xrightarrow{\mathcal{Q}_{\varphi_*}^{\mathsf{T}}} \mathcal{B}_G. & & \mathcal{B}_G. \xleftarrow{\;\;\mathcal{J}_{\varphi_*}^{\mathsf{T}}\;\;} &
\end{array}
\qquad (8)
$$

Note that, in general, $\mathcal{M}_\varphi \neq \mathcal{E}_\varphi^{\mathsf{T}}$. That is, the theory of induced matchings gives two different canonical matchings associated with a factorization. This is related to the fact that the composition of matchings given by the functors $\mathcal{J}$ and $\mathcal{Q}$ is not functorial.

Theorem 3.3 is a direct consequence of the following lemma.

Lemma 3.5. $\mathcal{M}_\varphi$ and $\mathcal{E}_\varphi^{\mathsf{T}}$ are injective sub-barcode matchings.

Proof. Because $\mathcal{M}_\varphi$ is a composition of injective matchings it is injective as well. So it suffices to show that $\mathcal{B}_G(\alpha) \subseteq \mathcal{B}_F(\beta)$ for all matched pairs $(\alpha, \beta) \in \mathcal{M}_\varphi$. If $(\alpha, \beta) \in \mathcal{M}_\varphi$ then, by the definition of compositon in $\mathbf{Bar_T}$, there exists a bar x of $\mathcal{B}_{F_\varphi}$ and pairs $(x, \alpha) \in \mathcal{Q}_{\varphi_*}$ and $(x, \beta) \in \mathcal{J}_{F_\varphi}$. By Theorem 2.4(b)(ii), $\mathcal{B}_G(\alpha) \subseteq \mathcal{B}_{F_\varphi}(x)$, and by Theorem 2.4(a)(ii) $\mathcal{B}_{F_\varphi}(x) \subseteq \mathcal{B}_F(\beta)$. So

$$\mathcal{B}_G(\alpha) \subseteq \mathcal{B}_{F_\varphi}(x) \subseteq \mathcal{B}_F(\beta),$$

and we conclude that $\mathcal{M}_\varphi$ is a sub-barcode matching.

The proof that $\mathcal{E}_\varphi^{\mathsf{T}}$ is a sub-barcode matching follows in the same way. $\qquad\square$

3.2 Barcode Functors

In previous work [4], it is shown that the category $\mathbf{Barc}$ of $\mathbf{R}$-indexed barcodes and *overlap matchings* is isomorphic to the category $\mathbf{Mch}^{\mathbf{R}}$ of functors $\mathbf{R} \to \mathbf{Mch}$ [22]. Similar to persistence modules, objects $\mathsf{b} \in \mathbf{Mch}^{\mathbf{R}}$ have arrow maps $\mathsf{b}\,[s \leq t]$ that serve to "glue" bars together by a parameter $t \in \mathbf{R}$ (Figure 7). The authors then show that, if $\mathsf{b}_{\mathbb{V}} : \mathbf{R} \to \mathbf{Mch}$ corresponds to the barcode of a persistence module $\mathbb{V}$, then the number of intervals in $\mathsf{im}\,\mathsf{b}\,[s \leq t]$ is equal to the rank of $\mathbb{V}\,[s \leq t]$. Because $\mathbf{Barc} \simeq \mathbf{Mch}^{\mathbf{R}}$, the natural transformations between two such barcodes corresponds to an overlap matching between them.

In this section we propose an alternative definition of $\mathbf{P}$-indexed barcodes as contravariant functors from the poset $\mathbf{Int_P}$ of intervals ordered by inclusion to $\mathbf{Mch}$. That is, we consider a larger functor category of intervals in a poset rather than the poset itself. This generalizes $\mathbf{Mch}^{\mathbf{P}}$ to a category in which natural transformations not only capture overlap matchings, but also sub-barcode matchings. In particular, Theorem 3.7 states that $\mathcal{A} \sqsubseteq \mathcal{B}$ is a sub-barcode matching if and only if there exists a monomorphism in the category $\mathbf{Fun}(\mathbf{Int_P^{op}}, \mathbf{Mch})$ of barcode functors defined formally as follows.
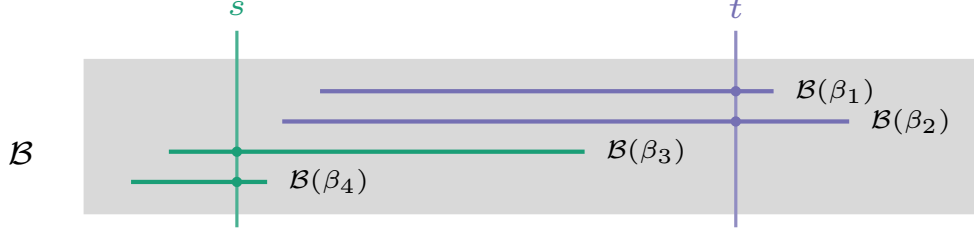
Figure 7: A functor $\mathsf{b}_{\mathcal{B}} \in \mathbf{Mch}^{\mathbf{P}}$ in which $\mathsf{b}_{\mathcal{B}}(s) = \{\beta_3, \beta_4\}$ and $\mathsf{b}_{\mathcal{B}}(t) = \{\beta_1, \beta_2\}$.
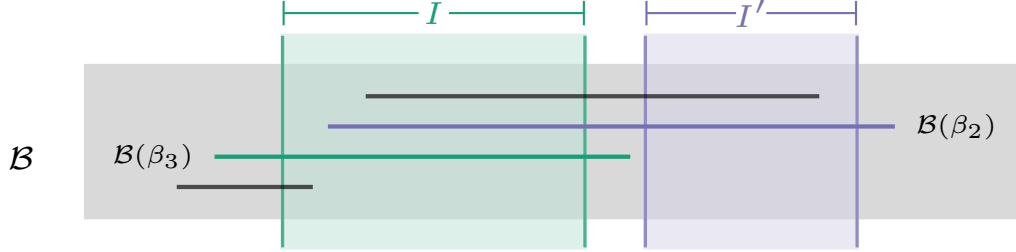
Figure 8: A barcode functor $\mathsf{bar}_{\mathcal{B}}$ and intervals I and I' where $\mathsf{bar}_{\mathcal{B}}(I) = \{\beta_3\}$ and $\mathsf{bar}_{\mathcal{B}}(I') = \{\beta_2\}$.

Given a barcode $\mathcal{B} \in \mathbf{Bar_P}$ and an interval $I \in \mathbf{Int_P}$ let $\underline{\mathcal{B}}_I \subseteq \underline{\mathcal{B}}$ denote the set of bars of $\mathcal{B}$ associated with intervals containing I:

$$\underline{\mathcal{B}}_I = \{\beta \in \underline{\mathcal{B}} \mid I \subseteq \mathcal{B}(\beta)\},$$

and let $\mathcal{B}_I := \mathcal{B}|_{\underline{\mathcal{B}}_I} : \underline{\mathcal{B}}_I \to \mathcal{I}_{\mathbf{P}}$ denote the restriction of $\mathcal{B}$ to I. If $I \subseteq J$ and $\beta \in \mathcal{B}(J)$ then $\beta \in \mathcal{B}(I)$. It follows that $\underline{\mathcal{B}}_I \supseteq \underline{\mathcal{B}}_J$ for all $I \subseteq J$ in $\mathbf{Int_P}$.

Definition 3.6 (Barcode Functor). Given $\mathcal{B} \in \mathbf{Bar_P}$, the associated <u>barcode functor</u> is a contravariant functor in the category $\mathbf{BarFun_P} := \mathbf{Fun}(\mathbf{Int_P^{op}}, \mathbf{Mch})$ defined

$$\mathsf{bar}_{\mathcal{B}} : \mathbf{Int_P^{op}} \longrightarrow \mathbf{Mch}$$
$$(objects) \qquad\qquad I \longmapsto \underline{\mathcal{B}}_I$$
$$(arrows) \qquad\qquad (I \subseteq J) \longmapsto (\underline{\mathcal{B}}_I \supseteq \underline{\mathcal{B}}_J).$$

Given a matching $M : \mathcal{A} \nrightarrow \mathcal{B}$ in $\mathbf{Bar_P}$, let $M_I \subseteq M$ denote the restriction of M to bars representing intervals containing I:

$$M_I = M \cap (\underline{\mathcal{A}}_I \times \underline{\mathcal{B}}_I).$$

Because $M_I \subseteq M$ and M is a matching, $M_I : \mathsf{bar}_{\mathcal{A}}(I) \nrightarrow \mathsf{bar}_{\mathcal{B}}(I)$ is a matching as well. Let $\mathsf{mch}_M : \mathsf{bar}_{\mathcal{A}} \Rightarrow \mathsf{bar}_{\mathcal{B}}$ denote the natural transformation with components M_I given by the commutativity of Diagram (9) for all $I \subseteq J$.

$$
\begin{array}{ccc}
I & \mathsf{bar}_{\mathcal{A}}(I) \xrightarrow{\;\;M_I\;\;} \mathsf{bar}_{\mathcal{B}}(I) \\
\big\downarrow{\subseteq} & \big\uparrow{\subseteq} \qquad\qquad \big\uparrow{\subseteq} \\
J & \mathsf{bar}_{\mathcal{A}}(J) \xrightarrow{\;\;M_J\;\;} \mathsf{bar}_{\mathcal{B}}(J)
\end{array}
\qquad (9)
$$

12

Theorem 3.7. *Given a poset* $\mathbf{P}$ *and barcodes* $\mathcal{A}, \mathcal{B} \in \mathbf{Bar_P}$, $\mathcal{A} \sqsubseteq \mathcal{B}$ *if and only if there exists a monomorphism* $\mathsf{M} : \mathsf{bar}_{\mathcal{A}} \hookrightarrow \mathsf{bar}_{\mathcal{B}}$ *in* $\mathbf{BarFun_P}$.

Proof. Because the arrow maps of barcode functors are inclusions, M is a monomorphism in $\mathbf{BarFun_P}$ if and only if its components M_I are injective matchings for all $I \in \mathbf{Int_P}$.

($\implies$) Suppose $\mathcal{A} \sqsubseteq \mathcal{B}$. Then there exists an injective sub-barcode matching $M : \mathcal{A} \leftrightarrow \mathcal{B}$ in $\mathbf{SubBar_P}$ with an associated natural transformation $\mathsf{mch}_M : \mathsf{bar}_{\mathcal{A}} \to \mathsf{bar}_{\mathcal{B}}$ in $\mathbf{BarFun_P}$. To show mch_M is a monomorphism, it suffices to show that the components $M_I : \mathsf{bar}_{\mathcal{A}}(I) \rightarrow \mathsf{bar}_{\mathcal{B}}(I)$ are injective matchings.

Let $I \in \mathbf{Int_P}$ be any interval. Let α be any bar in $\mathsf{bar}_{\mathcal{A}}(I)$. Because M is a sub-barcode matching, there exists a pair $(\alpha, \beta) \in M$ with $\mathcal{A}(\alpha) \subseteq \mathcal{B}(\beta)$. So $I \subseteq \mathcal{A}(\alpha) \subseteq \mathcal{B}(\beta)$ implies $\beta \in \mathsf{bar}_{\mathcal{B}}(I)$, thus $(\alpha, \beta) \in M_I$. Because we have such a β for any $\alpha \in \mathsf{bar}_{\mathcal{A}}(I)$, M_I is injective as desired.

($\impliedby$) Assume there exists a monomorphism $\mathsf{M} : \mathsf{bar}_{\mathcal{A}} \hookrightarrow \mathsf{bar}_{\mathcal{B}}$. Because $\underline{\mathcal{A}} = \bigcup_{I \in \mathbf{Int_P}} \mathsf{bar}_{\mathcal{A}}(I)$ and $\underline{\mathcal{B}} = \bigcup_{I \in \mathbf{Int_P}} \mathsf{bar}_{\mathcal{B}}(I)$, and because M_I is injective for all $I \in \mathbf{Int_P}$, $M = \bigcup_{I \in \mathbf{Int_P}} \mathsf{M}_I$ is an injective matching $M : \underline{\mathcal{A}} \leftrightarrow \underline{\mathcal{B}}$ in $\mathbf{Mch}$. So it suffices to show that M is a sub-barcode matching.

Let $\alpha \in \underline{\mathcal{A}}$. Because $\alpha \in \mathsf{bar}_{\mathcal{A}}(\mathcal{A}(\alpha))$ and $\mathsf{M}_{\mathcal{A}(\alpha)} : \mathsf{bar}_{\mathcal{A}}(\mathcal{A}(\alpha)) \hookrightarrow \mathsf{bar}_{\mathcal{B}}(\mathcal{A}(\alpha))$ is injective, there exists some $\beta \in \mathsf{bar}_{\mathcal{B}}(\mathcal{A}(\alpha))$ such that $(\alpha, \beta) \in \mathsf{M}_{\mathcal{A}(\alpha)}$, so $\mathcal{A}(\alpha) \subseteq \mathcal{B}(\beta)$. Because there exists such a pair $(\alpha, \beta) \in M$ for any $\alpha \in \underline{\mathcal{A}}$, we may conclude that M is a sub-barcode matching as desired.

$\square$

Theorem 3.7 implies that a barcode $\mathcal{A}$ is a sub-barcode of $\mathcal{B}$ if and only if $\mathsf{bar}_{\mathcal{A}}$ is a subobject of $\mathsf{bar}_{\mathcal{B}}$ in $\mathbf{BarFun_P}$. We note that not every object in this category should be considered a barcode functor, and hypothesize that the relevant subcategory of barcode functors is given by restricting to functors $\mathbf{Int_P^{op}} \to \mathbf{Bar_P}$ with projective limits in $\mathbf{Bar_P}$.

Importantly, an isomorphism $\mathcal{A} \simeq B$ in $\mathbf{Bar_P}$ only requires an isomorphism $\underline{\mathcal{A}} \simeq \underline{\mathcal{B}}$ in $\mathbf{Mch}$, and therefore does not correspond to an equivalence of barcodes. On the other hand, isomorphisms of barcodes in $\mathbf{Barc_P}$ and $\mathbf{SubBar_P}$ correspond to equivalences of barcodes, requiring both an isomorphism $s : \underline{\mathcal{A}} \xrightarrow{\sim} \underline{\mathcal{B}}$ and $\mathcal{A} = \mathcal{B} \circ s$. In fact, there are four obvious subcategories of $\mathbf{Bar_P}$ with this property that correspond to the four quadrants of the persistence plane (Figure 9). We will conclude this section by showing that there exist functors from these four subcategories to $\mathbf{BarFun_P}$.

Theorem 3.8. *There exist functors* $\mathbf{SubBar_P} \to \mathbf{BarFun_P}$ *and* $\mathbf{Barc_P} \to \mathbf{BarFun_P}$ *that take barcodes* $\mathcal{B}$ *to barcode functors* $\mathsf{bar}_{\mathcal{B}}$ *and matchings* M *to natural transformations* mch_M.

Proof. First we show that these maps take identity morphisms of $\mathbf{Bar_P}$ to identity morphisms in $\mathbf{BarFun_P}$. For all $I \in \mathbf{Int_P}$ and $\mathcal{B} \in \mathbf{Bar_P}$, we have

$$(\mathsf{mch}_{1_{\mathcal{B}}})_I = (1_{\mathcal{B}})_I = 1_{\mathcal{B}} \cap (\underline{\mathcal{B}}_I \times \underline{\mathcal{B}}_I) = 1_{\mathcal{B}_I} = (1_{\mathsf{bar}_{\mathcal{B}}})_I,$$

so $\mathsf{mch}_{1_{\mathcal{B}}} = 1_{\mathsf{bar}_{\mathcal{B}}}$.

We will now show that the composition of matchings in $\mathbf{SubBar_P}$ and $\mathbf{Barc_P}$ correspond to compositions of natural transformations in $\mathbf{BarFun_P}$. Let $M : \mathcal{A} \leftrightarrow \mathcal{B}$ and $N : \mathcal{B} \leftrightarrow \mathcal{C}$ be matchings
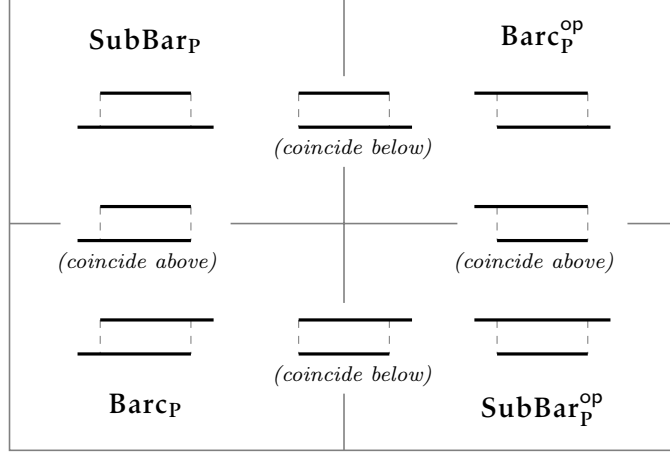
Figure 9: The four subcategories of $\mathbf{Bar_P}$ corresponding to matchings made in each quadrant of the persistence plane.

in $\mathbf{Bar_P}$. We will show that $\mathsf{mch}_{N \circ M} = \mathsf{mch}_N \circ \mathsf{mch}_M$ by showing that $(N \circ M)_I = N_I \circ M_I$ for all $I \in \mathbf{Int_P}$. Because

$$
\begin{aligned}
N_I \circ M_I &= (N \cap (\underline{\mathcal{B}}_I \times \underline{\mathcal{C}}_I)) \circ (M \cap (\underline{\mathcal{A}}_I \times \underline{\mathcal{B}}_I)) \\
&\subseteq (N \circ M) \cap (\underline{\mathcal{A}}_I \times \underline{\mathcal{C}}_I) \\
&= (N \circ M)_I
\end{aligned}
$$

it remains to show that $(N \circ M)_I \subseteq N_I \circ M_I$.

Let $(\alpha, \gamma) \in (N \circ M)_I$. So there exists some $\beta \in \underline{\mathcal{B}}$ such that $(\alpha, \beta) \in M$ and $(\beta, \gamma) \in N$. If $\beta \in \mathsf{bar}_{\mathcal{B}}(I)$ then $(\alpha, \gamma) \in N_I \circ M_I$, so it suffices to show that $\beta \in \mathsf{bar}_{\mathcal{B}}(I)$ when M and N are sub-barcode or overlap matchings.

If M is a sub-barcode matching then $(\alpha, \beta) \in M$ implies $\mathcal{A}(\alpha) \subseteq \mathcal{B}(\beta)$. Because $\alpha \in \mathsf{bar}_{\mathcal{A}}(I)$, $I \subseteq \mathcal{A}(\alpha) \subseteq \mathcal{B}(\beta)$, so $\beta \in \mathsf{bar}_{\mathcal{B}}(I)$. If M and N are overlap matchings then $\mathcal{A}(\alpha) \cap \mathcal{C}(\gamma) \subseteq \mathcal{B}(\beta)$ by Lemma 2.3. Because $\alpha \in \mathsf{bar}_{\mathcal{A}}(I)$ and $\gamma \in \mathsf{bar}_{\mathcal{C}}(I)$, $I \subseteq \mathcal{A}(\alpha) \cap \mathcal{C}(\gamma) \subseteq \mathcal{B}(\beta)$, so $\beta \in \mathsf{bar}_{\mathcal{B}}(I)$ as desired. $\qquad\square$

It is a straightforward exercise to show that there exist functors $\mathbf{SubBar_P^{op}} \to \mathbf{BarFun_P}$ and $\mathbf{Barc_P^{op}} \to \mathbf{BarFun_P}$ in the same way.

4   Sub-barcodes in TDA

The standard TDA pipeline spans several different categories, from functions to filtrations to persistence modules to barcodes (Diagram (10)). Ideally, in each of these steps, the transition would be functorial, making it clear how relationships between objects at one stage carry on to relationships later in the pipeline. Generally, it is the last step which fails to be functorial, resulting in the distinction between *hard* and *soft* stability [25]. However, the transition to barcodes is only Lipschitz with respect to metrics on the categories; in the poset $\mathbf{SBar}$ we set aside these metric considerations.

$$
X \xrightarrow{\ f\ } \mathbf{P} \xrightarrow{\ F\ } \mathbf{Top} \xrightarrow{\ \mathsf{H}\ } \mathbf{Vec} \dashrightarrow \mathbf{Mch}. \tag{10}
$$

In this section, we present an application of sub-barcodes to a concrete problem in topological data analysis. Given only an upper and lower bound $g \geq f \geq \ell$ of an unknown function f, the goal is to learn as much as possible about the barcode associated with f. We will begin by providing a clear framework for stating the problem. The result will then follow directly from the results of the previous section.

Given a poset $\mathbf{P}$ and a topological space X regarded as a discrete category, the category $\mathbf{P}^X$ of functions $X \to \mathbf{P}$ forms a poset in which $g \geq f$ if $g(t) \geq f(t)$ for all $t \in \mathbf{P}$. A filtration over $\mathbf{P}$ is a functor $F : \mathbf{P} \to \mathbf{Top}$ given by a topological space $F(t)$ for each $t \in \mathbf{P}$ and a continuous map $F[s \leq t] : F(s) \to F(t)$ for all $s \leq t$.

The most common way to construct a filtration $F \in \mathbf{Top}^{\mathbf{P}}$ is by taking sublevels of a function $f \in \mathbf{P}^X$:

$$F : \mathbf{P} \longrightarrow \mathbf{Top}$$
$$(objects) \qquad t \longmapsto \{x \in X \mid f(x) \leq t\}$$
$$(arrows) \qquad (s \leq t) \longmapsto (F(s) \subseteq F(t)).$$

Given sublevel filtrations $F, G \in \mathbf{Top}^{\mathbf{P}}$ of functions $g \geq f$ in $\mathbf{P}^X$, let $G \hookrightarrow F$ denote the natural inclusion map in $\mathbf{Top}^{\mathbf{P}}$ with components $G(t) \subseteq F(t)$ for all $t \in \mathbf{P}$. Note the contravariance introduced; larger functions give smaller sublevel sets.

Definition 4.1 (Sublevel Functor).

$$\mathsf{Sub} : \mathbf{P}^X \longrightarrow \mathbf{Top}^{\mathbf{P}}$$
$$(objects) \qquad f \longmapsto F$$
$$(arrows) \qquad (g \geq f) \longmapsto (G \hookrightarrow F).$$

Let $\mathsf{H} : \mathbf{Top} \to \mathbf{Vec}$ denote the homology functor H_k for some dimension k. The homology of a filtration $F : \mathbf{P} \to \mathbf{Top}$ gives a persistence module $\mathsf{H}(F) : \mathbf{P} \to \mathbf{Vec}$ in $\mathbf{Vec}^{\mathbf{P}}$. If the filtration is over a totally ordered set $\mathbf{T}$, then there is an associated barcode $\mathcal{B}_{\mathsf{H}(F)}$.

Notation 4.1. Given $g \geq f \geq \ell : X \to \mathbf{T}$ let $\mathcal{B}_F = \mathcal{B}_{\mathsf{H} \circ \mathsf{Sub}(f)}$ denote the barcode of $\mathsf{H}(F) = \mathsf{H} \circ \mathsf{Sub}(f)$, and let $\mathcal{B}_{G \hookrightarrow L} = \mathcal{B}_{\mathsf{im}\,\mathsf{H} \circ \mathsf{Sub}[g \geq \ell]}$ denote the barcode of $\mathsf{H}(G \hookrightarrow F) = \mathsf{H} \circ \mathsf{Sub}(g \geq \ell)$.

The following corollary of Theorem 3.3 follows directly from the fact that $g \geq f \geq \ell$ implies

$$G \hookrightarrow F \hookrightarrow L,$$

and asserts that we can compute a sub-barcode of f knowing only an upper and lower bound.

Corollary 4.2. *Let X be a topological space. Given functions $g \geq f \geq \ell : X \to \mathbf{T}$,*

$$\mathcal{B}_{G \hookrightarrow L} \sqsubseteq \mathcal{B}_F.$$

Naturally, if the bounds are loose everywhere, then the sub-barcode may be empty. However, when the bounds are sufficiently close in a neighborhood of significant topological features, one expects to see the relevant features represented in the sub-barcode. Importantly, the barcode $\mathcal{B}_{G \hookrightarrow L}$ can be computed using the image persistence algorithm of Cohen-Steiner et al. [29] (see also [32]), providing a natural computational variant of Corollary 4.2.

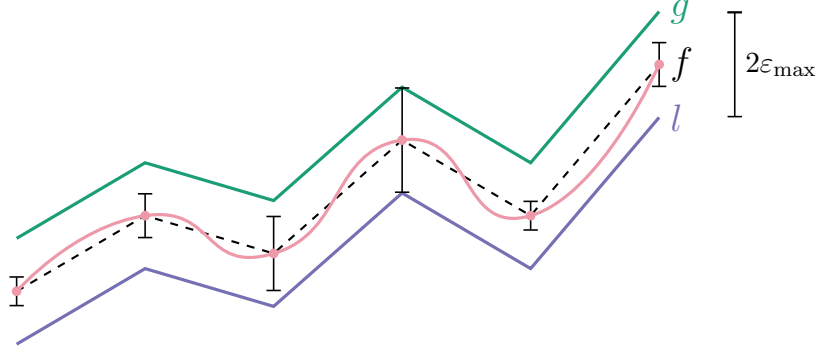We will conclude this section with two applications of Corollary 4.2.

Figure 10: An unknown function f and error bounds given for each $v \in V$ with upper and lower bounds g and ℓ given by the maximum error $\varepsilon_{\max}$.

Example 4.2 (Data on a simplicial complex with error bounds)**.** Let K be a finite abstract simplicial complex embedded in $\mathbb{R}^d$ with vertex set V, and let $\underline{K} \subset \mathbb{R}^d$ denote the image of the embedding. Let $f : \underline{K} \to \mathbb{R}$ be a function that is linear on the simplices of K so that f is completely specified by the restriction $f_V := f|_V : V \to \mathbb{R}$ of f to V. *Our goal is to compute or approximate the persistent homology of the sublevel filtration F of f.*

Let $K_f : \mathbf{R} \to \mathbf{Top}$ denote the filtration on K with components

$$K_f(t) = \big\{ \sigma \in K \mid \max_{v \in \sigma} f_V(v) \le t \big\}.$$

If f_V is known, then the barcode of F can be computed as the persistent homology of K_f. Otherwise, we may expect a bound ε_v on the error for each $v \in V$ with some guarantee. In this case, we propose the following question:

Question 4.3. Given $f' : V \to \mathbb{R}$ and an error bound ε_v with $|f'(v) - f(v)| \le \varepsilon_v$ for each $v \in V$,

what can we guarantee about $\mathcal{B}_F$?

There are two very different answers to this question, one from the perspective of stability, and the other from the perspective of sub-barcodes. We first give the standard approach via stability.

(Stability) Letting $\varepsilon_{\max} = \max_{v \in V} \varepsilon_v$ assume that f' has been extended to all of $\underline{K}$ so that

$$\|f - f'\|_\infty \le \varepsilon_{\max}.$$

It follows from the original stability theorem [19] that the bottleneck distance between $\mathcal{B}_{F'}$ and $\mathcal{B}_F$ will be within $\varepsilon_{\max}$; that is,

$$\mathsf{d}_{\mathrm{B}}\big(\mathcal{B}_{F'}, \mathcal{B}_F\big) \le \varepsilon_{\max}.$$

However, this bound depends on the worst case error. If there is even a single vertex v for which ε_v is large, then the guarantee will be quite weak.

(Sub-barcodes) We can use the error bounds ε_v associated with each vertex $v \in V$ to construct upper and lower bounds based not on the worst case error, but on the error at each vertex.

Formally, let $g \geq \ell : \underline{K} \to \mathbb{R}$ denote the piecewise linear functions with $g \geq f \geq \ell$ defined for each $v \in V$ as

$$g(v) := f'(v) + \varepsilon_v$$
$$\ell(v) := f'(v) - \varepsilon_v.$$

Using the image persistence algorithm [29] we can compute $\mathcal{B}_{G \hookrightarrow L}$ with the guarantee that

$$\mathcal{B}_{G \hookrightarrow L} \sqsubseteq \mathcal{B}_F$$

by Corollary 4.2.

Example 4.4 (Testing Topological Hypotheses)**.** The fundamental pattern in science is the falsification of hypotheses. It also happens to be a fundamental pattern in algebraic topology. Indeed, the main use of topological invariants is to distinguish between spaces, where a difference of invariants falsifies the hypothesis that the spaces are homeomorphic. Sub-barcodes provide a straightforward and computationally feasible way to test topological hypotheses.

Barcodes are topological invariants for filtrations. That is, if two filtrations are naturally isomorphic (as functors), then they have the same barcode. Given a barcode $\mathcal{A}$, the property of having $\mathcal{A}$ as a sub-barcode is also a topological invariant. Corollary 4.2 in particular, and Theorem 3.3 more generally, provide a way to compute a sub-barcode using only an upper and a lower bound, providing a testable hypothesis.

In data analysis, a natural isomorphism between filtrations often arises when the domain of a function has undergone some unknown transformation, usually as a result of measurement. For example, given an approximation of an established ground truth in a space X that has been measured in a space Y, we might expect the measurement to be a homeomorphism $h : X \to Y$ so that the barcode of f is equal to that of $f \circ h$. That is, we can answer questions about f without considering how it was measured (i.e., embedded) in Y. If different data sets were collected in different ways then their barcodes can be compared without having to "align" the data sets. This is the value of a topological invariant.

Formally, given $u : X \to \mathbb{R}$ along with upper and lower bounds $g \geq f \geq \ell : X \to \mathbb{R}$ of an unknown function f, we often encounter the following hypothesis.

Hypothesis 4.5. There exists a reparameterization $r : X \to X$ such that $f \circ r = u$.

If $\mathcal{B}_{G \hookrightarrow L} \not\sqsubseteq \mathcal{B}_U$, then no such reparameterization exists, and we have falsified the hypothesis.

5   Conclusion

While the stability of interleavings and bottleneck matchings provide a theoretical basis for the use of approximated barcodes in conventional data analysis, there are questions that can be answered in the absence of an interleaving. The goal of this work is not only to present sub-barcodes as a useful tool for answering some of these questions, but also to direct the application of persistence as a tool for answering questions about unknown data when conventional analysis cannot. The sub-barcode perspective shifts focus from questions about topological proximity to questions about topological obstructions.

There are natural computational questions that arise. For one interested in using sub-barcodes in data analysis the most pressing question is how to test if one barcode is a sub-barcode of another. In recent work Chubet showed that it is possible to find a maximum sub-barcode matching in the $O(n \log n)$ time for barcodes with n bars. The algorithm has the same complexity if the intervals

have arbitrary finite multiplicities. Moreover, for barcodes that do not satisfy a sub-barcode relation, one can compute the minimum shift needed to obtain a sub-barcode. The resulting sub-barcode distance is computable in expected $O(n \log^2 n)$ time, and is at most the bottleneck distance. Note, this means it is faster to test sub-barcodes than to compute bottleneck matchings.

In future work we will expand the theory of barcode functors and show how sub-barcodes and factorizations appear implicitly in prior work, offering a new perspective on barcodes, the persistence measure, and interleaving. In particular, we will apply the theory of generalized interleavings [25, 35] to δ-*smoothed persistence modules* $\mathbb{V}\delta : \mathbb{V}\delta_1 \to \mathbb{V}\delta_2$ given by pre-composition with a natural transformation of monotone functions:

$$
\mathbf{S} \xRightarrow[\delta_2]{\delta_1}{}_{\delta} \mathbf{T} \xrightarrow{\mathbb{V}} \mathbf{Vec} \rightsquigarrow \mathbf{S} \xRightarrow[\mathbb{V}\delta_2]{\mathbb{V}\delta_1}{}_{\mathbb{V}\delta} \mathbf{Vec}.
$$

The corresponding δ-smoothed barcode $\mathcal{B}_{\mathbb{V}}^{\delta}$ is given by pre-composition with the barcode functor $\mathsf{bar}_{\mathbb{V}}$ of $\mathcal{B}_{\mathbb{V}}$:

$$
\mathbf{Int}_{\mathbf{S}}^{\mathsf{op}} \xrightarrow{\delta} \mathbf{Int}_{\mathbf{T}}^{\mathsf{op}} \xrightarrow{\mathsf{bar}_{\mathbb{V}}} \mathbf{Mch} \rightsquigarrow \mathbf{Int}_{\mathbf{Q}}^{\mathsf{op}} \xrightarrow{\mathsf{bar}_{\mathbb{V}} \delta} \mathbf{Mch}.
$$

As an immediate corollary to Theorem 3.3, the existence of a δ-*factorization* $\varphi : \mathbb{V}\delta \rightleftharpoons 1_{\mathbb{U}}$ implies $\mathcal{B}_{\mathbb{V}}^{\delta} \sqsubseteq \mathcal{B}_{\mathbb{U}}$ when δ is given by a monotone function $\mathbf{Int}_{\mathbf{S}} \to \mathbf{Int}_{\mathbf{T}}$.

References

[1] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the Twenty-fifth Annual Symposium on Computational Geometry*, SCG '09, pages 237–246, New York, NY, USA, 2009. ACM.

[2] Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15(3):613–650, 2015.

[3] Ulrich Bauer and Michael Lesnick. Induced matchings and the algebraic stability of persistence barcodes. *Journal of Computational Geometry*, page Vol. 6 No. 2 (2015): Special issue of Selected Papers from SoCG 2014, 2015.

[4] Ulrich Bauer and Michael Lesnick. Persistence diagrams as diagrams: A categorification of the stability theorem. In *Topological Data Analysis*, pages 67–96. Springer, 2020.

[5] Edelsbrunner, Letscher, and Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4):511–533, Nov 2002.

[6] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, Feb 2005.

[7] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Extending persistence using poincaré and lefschetz duality. *Foundations of Computational Mathematics*, 9(1):79–103, 2009.

[8] Gunnar Carlsson and Vin De Silva. Zigzag persistence. *Foundations of computational mathematics*, 10(4):367–405, 2010.

[9] Vin de Silva, Dmitriy Morozov, and Mikael Vejdemo-Johansson. Dualities in persistent (co)homology. *Inverse Problems*, 27(12):124003, nov 2011.

[10] Vin de Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7:339–358, 2007.

[11] Frédéric Chazal and Steve Yann Oudot. Towards persistence-based reconstruction in euclidean spaces. In *Proceedings of the Twenty-fourth Annual Symposium on Computational Geometry*, SCG '08, pages 232–241, New York, NY, USA, 2008. ACM.

[12] Frédéric Chazal, Leonidas J. Guibas, Steve Y. Oudot, and Primoz Skraba. Scalar field analysis over point cloud data. *Discrete & Computational Geometry*, 46(4):743–775, 2011.

[13] Mickaël Buchet, Frédéric Chazal, Tamal K. Dey, Fengtao Fan, Steve Y. Oudot, and Yusu Wang. Topological analysis of scalar fields with outliers. In *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 827–841. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015.

[14] Gunnar Carlsson, Vin De Silva, and Dmitriy Morozov. Zigzag persistent homology and real-valued functions. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 247–256, 2009.

[15] Chao Chen and Michael Kerber. Persistent homology computation with a twist. In *Proceedings 27th European workshop on computational geometry*, volume 11, pages 197–200, 2011.

[16] Nicholas J Cavanna, Mahmoodreza Jahanseir, and Donald R Sheehy. A geometric perspective on sparse filtrations. *arXiv preprint arXiv:1506.03797*, 2015.

[17] Gregory Henselman and Robert Ghrist. Matroid filtrations and computational persistent homology. *arXiv preprint arXiv:1606.00199*, 2016.

[18] Ulrich Bauer. Ripser: efficient computation of vietoris–rips persistence barcodes. *Journal of Applied and Computational Topology*, 5(3):391–423, 2021.

[19] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, Jan 2007.

[20] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and its Applications*, 14(05):1550066, 2015.

[21] Frederic Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules, 2013.

[22] Herbert Edelsbrunner, Grzegorz Jabłoński, and Marian Mrozek. The persistent homology of a self-map. *Foundations of Computational Mathematics*, 15(5):1213–1244, 2015.

[23] Justin Michael Curry. *Sheaves, cosheaves and applications*. University of Pennsylvania, 2014.

[24] Peter Bubenik and Jonathan A Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51(3):600–627, 2014.

[25] Peter Bubenik, Vin De Silva, and Jonathan Scott. Metrics for generalized persistence modules. *Foundations of Computational Mathematics*, 15(6):1501–1531, 2015.

[26] Vin De Silva, Elizabeth Munch, and Amit Patel. Categorified reeb graphs. *Discrete & Computational Geometry*, 55(4):854–906, 2016.

[27] Amit Patel. Generalized persistence diagrams. *Journal of Applied and Computational Topology*, 1(3):397–419, 2018.

[28] Peter Bubenik and Nikola Milićević. Homological algebra for persistence modules. *Foundations of Computational Mathematics*, 21(5):1233–1278, 2021.

[29] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. Persistent homology for kernels, images, and cokernels. In *SODA: ACM-SIAM Symposium on Discrete Algorithms*, 2009.

[30] Frédéric Chazal and André Lieutier. Weak feature size and persistent homology: Computing homology of solids in R^n from noisy data samples. In *Proceedings of the 21st ACM Symposium on Computational Geometry*, pages 255–262, 2005.

[31] Frédéric Chazal Chazal, Leonidas J. Guibas, Steve Y. Oudot, and Primoz Skraba. Persistence-based clustering in riemannian manifolds. *J. ACM*, 60(6:41):97–106, 2013.

[32] Ulrich Bauer and Maximilian Schmahl. Efficient computation of image persistence. *arXiv preprint arXiv:2201.04170*, 2022.

[33] M. Kashiwara and P. Schapira. *Categories and Sheaves*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2005.

[34] Saunders Mac Lane. *Categories for the working mathematician*, volume 5. Springer Science & Business Media, 2013.

[35] Rachel Levanger. *A Comparison Framework For Interleaved Persistence Modules and Applications of Persistent Homology to Problems in Fluid Dynamics*. Rutgers The State University of New Jersey-New Brunswick, 2017.