Wasserstein distance for generalized persistence modules and abelian categories

Peter Bubenik, Jonathan Scott, and Donald Stanley

ABSTRACT. In persistence theory and practice, measuring distances between modules is central. The Wasserstein distances are the standard family of L^p distances (with $1 \leqslant p \leqslant \infty$) for persistence modules. They are defined in a combinatorial way for discrete invariants called persistence diagrams that are defined for certain persistence modules. We give an algebraic formulation of these distances that applies to all persistence modules. Furthermore, for $p = 1$ this definition generalizes to abelian categories and for arbitrary p it generalizes to Krull-Schmidt categories. In particular, we obtain a definition of Wasserstein distance for multiparameter persistence modules. These distances may be useful for the computation of distance between generalized persistence modules. In our most technical proof, we classify certain maps of persistence modules, which may be of independent interest.

1. Introduction

Persistence modules are certain modules that are the main algebraic object of study in applied algebraic topology. Classically, they are modules over $K[x]$, the polynomial ring in one variable, with coefficients in a field K. More precisely, they are $\mathbb{Z}$-graded modules over the $\mathbb{Z}$-graded algebra $K[x]$, where x has degree one and the unit has degree zero. That is, they are a $\mathbb{Z}$-graded sequence of K-vector spaces together with linear maps given by the action of x [28]. More generally, persistence modules are $\mathbb{R}$-graded modules over the $\mathbb{R}$-graded K-algebra on the monoid of non-negative real numbers with addition. The latter may be viewed as extending the former by allowing x to have arbitrary non-negative real coefficients. Examples of persistence modules are the *interval modules* (Definition 2.9) which are indecomposable.

While all of the above is well studied in algebra, what is novel to the persistence point of view is the emphasis on quantifying the distance between persistence modules, which is crucial for applications in which persistence modules provide an algebraic summary of data. In particular, these distances should have the property that highly persistent elements, that is, nontrivial elements $x^m a$ with m large, are highly weighted.

The most important distances for persistence modules are a family of L^p distances, for $1 \leqslant p \leqslant \infty$, called p-*Wasserstein distances* [13]. For $p = \infty$, this distance is also called the *bottleneck distance* [12]. These are not usually defined directly for persistence modules, but for a certain set called a *persistence diagram* [12] that is obtained from a persistence module[1].

The bottleneck distance has been the most well studied. It has an equivalent algebraic formulation called *interleaving distance* [10, 22] that has a categorical interpretation [5], that has been extended in great generality [3, 17, 2, 4]. More concretely, it has been used for various generalized persistence modules [24, 16, 25].

However, from the metric point of view, these distances, being L^∞ distances, are rather weak. Saying that two persistence modules are close in p-Wasserstein distance for $p < \infty$ is much stronger, with 1-Wasserstein distance giving the strongest notion of proximity.

In this paper, we develop an algebraic version of Wasserstein distance for persistence modules and for more general algebraic settings.

While our focus is on algebra, we will adopt a somewhat representation-theoretic or categorical framework. We will consider persistence modules to be functors from an ordered set of real numbers (e.g. $(\mathbb{R}, \leqslant)$, $(\mathbb{Z}, \leqslant)$, $(\mathbb{N}, \leqslant)$, or $\{0 \leqslant 1 \leqslant \cdots \leqslant n\}$), thought of as a category, to the category of finite-dimensional vector spaces over the fixed field K.

All of our theory applies more generally to abelian categories (Section 2.3) that are Krull-Schmidt (Section 2.2) and part of our theory applies to arbitrary abelian categories. In particular, we provide a Wasserstein distance for multiparameter persistence modules [8].

We introduce two main constructions. The first we call a *path metric* and can be defined in any abelian category. It assumes that a collections of objects has been assigned non-negative weights. From this data, a distance is defined between any two objects, using certain zigzags of paths and the weights of their kernels and cokernels (Section 4).

The second construction requires the Krull-Schmidt property. It assumes that one has a distance, d, for indecomposable objects, and for $1 \leqslant p \leqslant \infty$. This distance is extended to a p-*Wasserstein distance*, $W_p(d)$ (Section 5).

We define what it means for a metric to be p-*subadditive* (Definition 5.8) and prove the following.

THEOREM 1.1 (Theorem 5.13). *The metric $W_p(d)$ is the largest p-subadditive metric which agrees with d on the indecomposable objects.*

1.1. Persistence modules. For persistence modules with ordered indexing set $\mathbb{Z}$, $\mathbb{N}$, or $\mathbf{n} := \{0, 1, \ldots, n\}$, we consider the path metric d_w that starts by assigning each simple module (Definition 2.8) weight one. For persistence modules with ordered indexing set $\mathbb{R}$, we consider the path metric d_w that starts by assigning each interval module (Definition 2.9) the length of the corresponding interval.

[1]These sets are subsets (with multiplicity) of $\mathbb{R}^2$. In [13], the Wasserstein distance uses the ∞-norm to measure distances in $\mathbb{R}^2$, we will use a version of the Wasserstein distance where distances in $\mathbb{R}^2$ are measured using the 1-norm.

We prove the following.

THEOREM 1.2 (Proposition 3.1 and Theorem 4.5). *For interval persistence modules* I *and* J,

$$W_1(I, J) = \lambda(I \triangle J) = d_w(I, J),$$

where W_1 *denotes the* 1*-Wasserstein distance,* λ *denotes the counting measure or Lebesgue measure, as appropriate, and* $\triangle$ *denotes the symmetric difference (we abuse notation and refer to the interval module by its support).*

THEOREM 1.3 (Theorem 6.1). *For persistence modules, the path metric* d_w *equals the* 1*-Wasserstein distance.*

THEOREM 1.4 (Theorem 5.5). *Furthermore, for* $1 \leqslant p \leqslant \infty$, $W_p(d_w)$ *equals the* p*-Wasserstein distance for persistence modules.*

Since our definition of path metric can be applied to any abelian category (Definition 4.1 and Lemma 4.2), it can be applied to arbitrary persistence modules. That is, to any functors from the ordered set of real numbers to the category of (not-necessarily finite dimensional) K-vector spaces. We thus have the following.

THEOREM 1.5. *By Theorem 1.3,* d_w *extends the* 1*-Wasserstein distance from pointwise finite-dimensional persistence modules to arbitrary persistence modules.*

In particular, our definition applies to persistence modules even if they do not have a persistence diagram [11].

1.2. Zigzag persistence modules. Zigzag persistence modules are linear sequences of vector spaces in which the maps are allowed to go in either direction (in a specified pattern). For example, consider the three following three zigzag persistence modules L, M, and N,

$$\begin{array}{ccccccccccc} L & = & K & \to & K & \to & K & \leftarrow & K & \leftarrow & K \\ M & = & K & \to & K & \to & K & \leftarrow & 0 & \leftarrow & 0 \\ N & = & 0 & \to & 0 & \to & K & \leftarrow & K & \leftarrow & K \end{array}$$

where in each case the maps are the identity if possible and are otherwise 0. These may be viewed as representations of the following quiver,

$$\bullet \to \bullet \to \bullet \leftarrow \bullet \leftarrow \bullet \tag{1.6}$$

or modules over the corresponding path algebra, or functors from the category (1.6) to the category of K-vector spaces. The zigzag persistence modules L, M, and N, are indecomposable. In fact, the indecomposable modules for such linear quivers are exactly the interval modules [19]. However, we will show that our distances for this quiver behave differently then for the corresponding ordered quiver $\bullet \to \bullet \to \bullet \to \bullet \to \bullet$.

As we did for persistence modules, we can assign the simple modules weight one and consider the corresponding path metric, d_w, and p-Wasserstein metric, $W_p(d_w|_{\mathcal{I}})$, where $\mathcal{I}$ denotes the interval zigzag modules. However, unlike for persistence modules, $W_1(d_w|_{\mathcal{I}}) \neq d_w$. Indeed, there is a surjective map $M \oplus N \to L$ with simple kernel and so $d_w(M \oplus N, L) = 1$ but for $W_1(d_w|_{\mathcal{I}})$ we need to match indecomposables (see

Definition 5.1), so $W_1(d_w|_{\mathcal{I}})(M \oplus N, L) = d_w(M, L) + d_w(N, 0) = 2 + 3 = 5$. Which of these metrics is most appropriate will depend on the application.

1.3. Classification of certain maps of persistence modules. In our proof of Theorem 1.3, we prove the following two classification results for morphisms of persistence modules into or out of interval modules, which may be of independent interest. Let M be a persistence module and $[a, b)$ be an interval module. The maps $p_\bullet$ and $i_\bullet$ are the maps to and from the direct summands in a direct sum.

THEOREM 1.7 (Theorem 6.6). *Given* $f\colon [a, b) \to M$, *there exists* $a_1 < a_2 \cdots < a_n \leqslant a < b_n < \cdots < b_2 < b_1 \leqslant b$, *a module* N *and an isomorphism* $\theta\colon M \to N \oplus_{j=1}^{n} [a_j, b_j)$ *such that* $p_N \theta f = 0$ *and* $p_j \theta f$ *is nonzero for all* j. *Furthermore,* $\ker f \cong [b_1, b)$ *and*

$$\operatorname{coker} f \cong N \oplus [a_1, b_2) \oplus \cdots \oplus [a_{n-1}, b_n) \oplus [a_n, a).$$

THEOREM 1.8 (Theorem 6.8). *Given* $f : M \to [a, b)$ *there exist* $a_1 < a_2 < \cdots < a_n < b \leqslant b_n < \cdots < b_2 < b_1$, *a module* N, *and an isomorphism* $\theta : N \oplus \bigoplus_{j=1}^{n} [a_j, b_j) \to M$ *such that* $f\theta i_N = 0$ *and* $f\theta i_j$ *is nonzero for all* j. *Furthermore* $\operatorname{coker} f \cong [a, a_1)$ *and*

$$\ker f \cong [a_2, b_1) \oplus \cdots \oplus [a_{n-1}, b_n) \oplus [b, b_n).$$

The map f in the first theorem can be thought of as the spreading the effect of the interval module $[a, b)$ to a number of summands of the codomain of f in a constrained way. It gives a canonical form to a map from an interval module to a persistence module. This could be used to give indecomposable modules in commutative ladders [18].

1.4. Computations for persistence modules and multiparameter persistence modules. An important advantage of our path metric is that is does not require a decomposition into indecomposables. This is important in practice, since for generalized persistence modules, there is no effective algorithm for finding such a decomposition [7]. Even for persistence modules, computing interval decompositions is computationally expensive.

Since the path metric is defined as an infimum over paths, the cost of any path provides an upper bound for the distance. Furthermore, for many choices of path metric, the dimension vectors can be used to compute a lower bound and perhaps an upper bound for the distance.

(1) Can we compute or estimate $d_w(M, N)$ for persistence modules arising from the homology of filtered complexes without computing an interval decomposition?
(2) What is a good choice of weight function for multi-parameter persistence modules, and can we compute or estimate $d_w(M, N)$ for such a weight?

For example, consider the multiparameter persistence modules indexed by $(\mathbb{Z}, \leqslant)^n$. Now take those modules M for which there exists a set $S \subset \mathbb{Z}^n$ and $M(a) = K$ if $a \in S$, $M(a) = 0$ otherwise, and $M(a \leqslant b)$ is the identity on K if $a \leqslant b \in S$. Let $w(M) = \lambda(S)$

where λ denotes the counting measure on $\mathbb{Z}^n$. Then for multiparameter persistence modules M and N, the proof of Corollary 4.15 generalizes to show that

$$\int |\dim(M) - \dim(N)| \leqslant d_w(M, N).$$

More generally we can assign to any persistence module the weight $w(M) = \lambda(\dim(M))$.

We can do the same for multiparameter persistence modules indexed by $(\mathbb{R}^n, \leqslant)$, taking λ to be the Lebesgue measure on $\mathbb{R}^n$. We just need to take care that we only use sets that are Lebesgue measurable.

1.5. Representations of quivers. We conclude this section with one more question for future work.

For many quivers (those of wild type [1]) understanding all of their representations is hopeless. However, given a path metric, d_w, the balls centered at the zero representation give a filtration of the set of isomorphism classes of representations.

(3) Is there a choice of weight w for which the path metric d_w can help one understand the representations of a quiver?

Related work. The reader is encouraged to consider two recent extensive algebraic treatments of multiparameter persistence modules [20, 23]. In the first, numerous tools from commutative algebra are used to study $\mathbb{N}^d$-graded persistence modules. In the second, machinery of commutative algebra is redeveloped for the $\mathbb{R}^d$-graded setting and used to study $\mathbb{R}^d$-graded persistence modules.

Outline of the paper. In Section 2 we provide necessary background. In Section 3 we reformulate the Wasserstein distance for persistence modules. In Section 4 we define path metrics for abelian categories and work out the details for a path metric for persistence modules. In Section 5 we define Wasserstein distance for Krull-Schmidt categories. Finally, in Section 6 we show that for persistence modules the 1-Wasserstein distance and our path metric agree.

2. Background

In Sections 2.1, 2.2, and 2.3, we summarize elementary definitions and properties of additive, abelian, and Krull-Schmidt categories, closely following the development in [21]. In Sections 2.4 and 2.5, we specialize to categories of persistence modules and multiparameter persistence modules. In Section 2.6 we define metrics for abelian categories. We review p-norms in Section 2.7. In Section 2.8 we define the Wasserstein distance for persistence modules. Finally, in Section 2.9 we define and give notation for zigzags of morphisms.

2.1. Additive categories. An *additive category* is one that is enriched in abelian groups (i.e. hom sets are abelian groups, and composition of morphisms is biadditive) and that has all finite products and a *zero object* 0 such that for every object X there are unique morphisms $0 \to X$ and $X \to 0$.

Let $\mathbf{A}$ be an additive category. Following Krause [21], we say that X is the *direct sum* of Y and Z in $\mathbf{A}$ if there are morphisms $i : Y \to X$, $j : Z \to X$, $p : X \to Y$, and $q : X \to Z$ such that $ip + jq = 1_X$, $pi = 1_Y$, and $qj = 1_Z$. Thus p and q are epimorphisms, i and j are monomorphisms, and we consider Y and Z to be subobjects of X. We write $X \cong Y \oplus Z$.

One can show that $qi = pj = 0$, from which it is easy to deduce that i and j determine an isomorphism $X \cong Y \amalg Z$, and that p and q determine an isomorphism $X \cong Y \times Z$.

An object $X \in \mathbf{A}$ is *indecomposable* if $X \cong Y \oplus Z$ implies that either Y or Z is 0.

2.2. Krull-Schmidt categories. In this section we introduce Krull-Schmidt categories. A good reference is [21]. Recall that a ring is *local* if its set of non-unit elements is closed under addition.

Definition 2.1. An additive category $\mathbf{A}$ is said to be a *Krull-Schmidt category* if every object decomposes into a finite direct sum of objects having local endomorphism rings.

Proposition 2.2. *In a Krull-Schmidt category $\mathbf{A}$, for all $M \in \mathbf{A}$,*

- *M is indecomposable if and only if its endomorphism ring is local;*
- *M is isomorphic to a finite direct sum of indecomposables; and*
- *if $M \cong \bigoplus_{i=1}^m M_i$ and $M \cong \bigoplus_{j=1}^n N_j$, where each direct summand is indecomposable, then $m = n$ and $M_i \cong N_{\sigma(i)}$ for all i and some permutation σ.*

Theorem 2.3 (Krull-Remak-Schmidt-Azuyama). *Let A be a finite-dimensional associative algebra over a field K. The category A-**mod** of finite-dimensional left A-modules over A is a Krull-Schmidt category.*

2.3. Abelian categories. In this section, we define abelian category and give several examples arising from persistent homology. We will introduce a broader class of examples in the next section.

Definition 2.4. An additive category is *abelian* if it has all kernels and cokernels, and if for every $f : M \to N$, the induced morphism $\bar{f}$ in the natural factorization,

$$\begin{tikzcd} \ker f \arrow[r, "j"] & M \arrow[r, "f"] \arrow[d] & N \arrow[r, "q"] & \operatorname{coker} f \\ & \operatorname{coker} j \arrow[r, "\bar{f}"] & \ker q \arrow[u] & \end{tikzcd}$$

is an isomorphism.

If $j : N \to M$ is the inclusion of a subobject, then we will write $\operatorname{coker} j = M/N$. Although subobjects are technically equivalence classes of such monomorphisms j, if j'

is an equivalent monomorphism, then there is a unique isomorphism $\operatorname{coker} j \cong \operatorname{coker} j'$. It follows from the definition that if f is a monomorphism (respectively, epimorphism) then $M \cong \ker q$ (respectively, $N \cong \operatorname{coker} j$).

EXAMPLE 2.5. Let K be a field. The category $\mathbf{Vect}_K$ of vector spaces over K and K-linear maps is an abelian category, as is the category $\mathbf{vect}_K$ of finite-dimensional vector spaces over K and K-linear maps.

EXAMPLE 2.6. If $\mathbf{A}$ is abelian and $\mathbf{D}$ is small then the category $\mathbf{A}^{\mathbf{D}}$, of functors from $\mathbf{D}$ to $\mathbf{A}$ and natural transformations, is abelian. $(\mathbb{N}, \leqslant)$, $(\mathbb{Z}, \leqslant)$, $(\mathbb{R}, \leqslant)$, or $(\mathbb{R}^d, \leqslant)$.

THEOREM 2.7 ([26, Prop. 5.92]). *A full subcategory* $\mathbf{S}$ *of an abelian category* $\mathbf{A}$ *is is an abelian category if the zero object of* $\mathbf{A}$ *lies in* $\mathbf{S}$*, and* $\mathbf{S}$ *is closed under binary direct sums, kernels and cokernels.*

DEFINITION 2.8. An object M of an abelian category is *simple* if it has precisely two subobjects, namely 0 and M.

2.4. Persistence modules. In this section we define various categories of persistence modules.

Given a poset $(P, \leqslant)$, there is a corresponding category in which the objects are the elements of P, and for pairs of objects x and y, there is exactly one morphism $x \to y$ if $x \leqslant y$ and otherwise there is no morphism from x to y. For example, we have a category corresponding to the poset $(\mathbb{R}, \leqslant)$ of the real numbers with the usual linear order as a category. Similarly, we have categories corresponding to the sub-posets $(R, \leqslant) \supset (\mathbb{Z}, \leqslant) \supset (\mathbb{N}, \leqslant) \supset (\{1, 2, \dots, n\}, \leqslant)$ where we denote the latter by $\mathbf{n}$.

Let K be a field. By Examples 2.5 and 2.6, each of the functor categories $\mathbf{Vect}_K^{\mathbf{n}}$, $\mathbf{Vect}_K^{(\mathbb{N},\leqslant)}$, $\mathbf{Vect}_K^{(\mathbb{Z},\leqslant)}$, $\mathbf{Vect}_K^{(\mathbb{R},\leqslant)}$, $\mathbf{vect}_K^{\mathbf{n}}$, $\mathbf{vect}_K^{(\mathbb{N},\leqslant)}$, $\mathbf{vect}_K^{(\mathbb{Z},\leqslant)}$, and $\mathbf{vect}_K^{(\mathbb{R},\leqslant)}$, is an abelian category. We call the objects of any of these categories *persistence modules*. The objects of the latter four categories are called *pointwise finite dimensional* persistence modules.

DEFINITION 2.9. Let $(P, \leqslant)$ be a poset. Given an interval $I \subset P$, there is a corresponding persistence module, M, called an *interval module*, given by $M(a) = K$ if $a \in I$ and $M(a) = 0$ otherwise, and $M(a \leqslant b) = 1_K$ if $a, b \in I$ and $M(a \leqslant b) = 0$ otherwise. It will be convenient to abuse notation and denote M by I.

It is a good exercise to check the following.

LEMMA 2.10. *Each interval module is indecomposable.* □

ASSUMPTION 2.11. For convenience we will always denote nonzero interval modules in $\mathbf{vect}_K^{(P,\leqslant)}$ for $P = \mathbf{n}$, $(\mathbb{N}, \leqslant)$, or $(\mathbb{Z}, \leqslant)$, as intervals of the form $[a, b+1)$ where $a \leqslant b \in P$. We do this so that the counting measure of an interval in P coincides with the Lebesgue measure of the corresponding interval in $\mathbb{R}$.

The simple modules in $\mathbf{vect}_K^{\mathbf{n}}$, $\mathbf{vect}_K^{(\mathbb{N},\leqslant)}$, and $\mathbf{vect}_K^{(\mathbb{Z},\leqslant)}$ are the interval modules $S := [i, i+1)$ where i is in $\mathbf{n}$, $\mathbb{N}$, or $\mathbb{Z}$, respectively. That is, $S(j) = K$ if $i = j$ and $S(j) = 0$ otherwise.

Let $V : (P, \leqslant) \to \mathbf{vect}_K$ be a persistence module. Let $(\mathbf{vect}_K)_0$ denote the class of finite-dimensional K-vector spaces (i.e. the objects of $\mathbf{vect}_K$). The *dimension vector* for V (also called the *Hilbert function* for V) is the composite function given by

$$P \xrightarrow{V} (\mathbf{vect}_K)_0 \xrightarrow{\dim} \mathbb{N},$$

where dim denotes the function that gives the dimension of a K-vector space. We denote this composite function by $\dim(V)$. For example, if I is an interval module, then $\dim(I)$ is the indicator function on I.

It is a special case of Gabriel's Theorem [19] that each object in $\mathbf{vect}_K^{\mathbf{n}}$ is isomorphic to a finite direct sum of interval modules. This is no longer true for $\mathbf{vect}_K^{(\mathbb{N},\leqslant)}$, $\mathbf{vect}_K^{(\mathbb{Z},\leqslant)}$, and $\mathbf{vect}_K^{(\mathbb{R},\leqslant)}$, as they contain objects that are infinite direct sums of indecomposables. For example, $\oplus_{j=1}^{\infty}[j, j+1)$. However, it is true that every object in these categories is isomorphic to a direct sum of interval modules [15]. Thus, the category $\mathbf{vect}_K^{\mathbf{n}}$ is Krull-Schmidt, but the categories $\mathbf{vect}_K^{(\mathbb{N},\leqslant)}$, $\mathbf{vect}_K^{(\mathbb{Z},\leqslant)}$, and $\mathbf{vect}_K^{(\mathbb{R},\leqslant)}$, are not. To rectify the situation, we introduce the notion of *finite-type* interval modules.

Definition 2.12. A persistence module has *finite type* if it is isomorphic to a finite direct sum of interval modules. Let $\mathbf{vect}_{K,ft}^{(\mathbb{R},\leqslant)}$, $\mathbf{vect}_{K,ft}^{(\mathbb{Z},\leqslant)}$, and $\mathbf{vect}_{K,ft}^{(\mathbb{N},\leqslant)}$ denote the full subcategories of $\mathbf{vect}_K^{(\mathbb{R},\leqslant)}$, $\mathbf{vect}_K^{(\mathbb{Z},\leqslant)}$, and $\mathbf{vect}_K^{(\mathbb{N},\leqslant)}$, respectively consisting of objects of finite type.

We recall the definition of critical value of a persistence module and then define two one-sided versions of it.

Definition 2.13 ([5]). For a persistence module M call $c \in \mathbb{R}$ a *regular value* of M if there is an open interval I containing c such that for all $a \leqslant b \in I$, $M(a \leqslant b)$ is an isomorphism. If c is not a regular value then call it a *critical value*.

Definition 2.14. For a persistence module M call $c \in \mathbb{R}$ a *upper (respectively, lower) regular value* of M if there is an open interval I containing c such that for all $a \leqslant b \in I$, $M(a \leqslant b)$ is injective (respectively, surjective). If c is not a upper (lower) regular value then call it a *upper (respectively, lower) critical value*.

Example 2.15. If M is a direct sum of interval modules then the upper critical values of M consist of the suprema of the intervals (whenever they exist) and the lower critical values of M consist of the infima of the intervals (whenever they exist).

Lemma 2.16. *Submodules and quotients of finite-type persistence modules are of finite type.*

Proof. Let M be a persistence modules with a submodule $j : N \hookrightarrow M$ and quotient module $p : M \twoheadrightarrow Q$. Then for all $a \in A$, $j_a : N(a) \hookrightarrow M(a)$ is an injection and $q_a : M(a) \twoheadrightarrow Q(a)$ is a surjection. From the commutative diagrams

$$\begin{tikzcd} N(a) \arrow[r, "N(a\leqslant b)"] \arrow[d, hook, "j_a"'] & N(b) \arrow[d, hook, "j_b"] \\ M(a) \arrow[r, "M(a\leqslant b)"] & M(b) \end{tikzcd} \qquad \begin{tikzcd} M(a) \arrow[r, "M(a\leqslant b)"] \arrow[d, two heads, "q_a"'] & M(b) \arrow[d, two heads, "q_b"] \\ Q(a) \arrow[r, "Q(a\leqslant b)"] & Q(b) \end{tikzcd}$$

we see that if $M(a \leqslant b)$ is injective then so is $N(a \leqslant b)$ and that if $M(a \leqslant b)$ is surjective then so is $Q(a \leqslant b)$. Thus upper regular values of M are also upper regular values of N and lower regular values of M are also lower regular values of Q. Hence upper critical values of N are also upper critical values of M and lower critical values of Q are also lower critical values of M.

Now assume that M has finite type. Then M has finitely many upper and lower critical values. By our previous observation, N has finitely many upper critical values and Q has finitely many lower critical values. Since M is pointwise finite-dimensional, so are N and Q. By [15], N and Q are isomorphic to direct sums of nonzero interval modules. Since N and Q are pointwise finite-dimensional, all but finitely many of these intervals are bounded. If N (respectively, Q) is not finite-type, then infinitely many of these intervals must have the same supremum (respectively, infimum). But this contradicts that N and Q are pointwise finite-dimensional. □

Theorem 2.17. *The categories* $\mathbf{vect}_{K,ft}^{(\mathbb{N},\leqslant)}$, $\mathbf{vect}_{K,ft}^{(\mathbb{Z},\leqslant)}$, *and* $\mathbf{vect}_{K,ft}^{(\mathbb{R},\leqslant)}$, *are abelian and Krull-Schmidt.*

Proof. Let M and N be persistence modules of finite type. Then $M \oplus N$ has finite type. Assume $f : M \to N$. Then by Lemma 2.16, $\ker(f)$ and $\operatorname{coker}(f)$ have finite type. Therefore by Theorem 2.7, the full subcategories of finite-type persistence modules are abelian.

By definition, any object of finite type is isomorphic to the direct sum of interval modules. The endomorphism ring of any interval module is isomorphic to the ground field, K, and is therefore local. Hence, the full subcategories of finite-type persistence modules are Krull-Schmidt. □

2.5. Multiparameter persistence modules. In this section we define multiparameter persistence modules. One of our goals is to define Wasserstein distances for multiparameter persistence modules.

Let $(P_i, \leqslant)$ be a poset for $i = 1, \dots, d$, with $d \geqslant 2$. Then $\prod_{i=1}^{d} P_i$ is a poset with the poset structure given by

$$(x_1, \dots, x_d) \leqslant (y_1, \dots, y_d) \text{ if and only if } x_i \leqslant y_i \text{ for all } i = 1, \dots, d.$$

If $(P_i, \leqslant) = (P, \leqslant)$ for all i then we denote this product poset by $(P^d, \leqslant)$. If $(P, \leqslant) = (\mathbb{R}, \leqslant)$, $(\mathbb{Z}, \leqslant)$, $(\mathbb{N}, \leqslant)$, or $\mathbf{n}$ then a functor $M : (P^d, \leqslant) \to \mathbf{vect}_K$ is called a *multiparameter persistence module*.

2.6. Metrics on additive categories. In this section we define a notion of metric for classes of objects in an abelian category.

Definition 2.18. A *symmetric Lawvere metric* on a class $\mathcal{C}$ is a function that assigns to any pair $M, N \in \mathcal{C}$ a number $d(M, N) \in [0, \infty]$ such that for all $M \in \mathcal{C}$, $d(M, M) = 0$, for all $M, N \in \mathcal{C}$, $d(M, N) = d(N, M)$, and for all $M, N, P \in \mathcal{C}$, $d(M, P) \leqslant d(M, N) + d(N, P)$.

This definition relaxes the usual definition of a metric in three ways: it is allowed to take on the value ∞; $d(M,N) = 0$ does not imply that $M = N$; and the class $\mathcal{C}$ is not required to be a set. [2]

Definition 2.19. Let $\mathcal{C}$ be a class of objects in an additive category A. We define a *metric* on $\mathcal{C}$ to be a symmetric Lawvere metric with the additional property that if $M, N \in \mathcal{C}$ with $M \cong N$ then $d(M,N) = 0$. A *metric on an additive category* A is a metric on the class of all objects in A.

Our definition does allow non-isomorphic objects M and N to have $d(M,N) = 0$.

Lemma 2.20. *Let* d *be a metric on a class of objects* $\mathcal{C}$ *in an additive category* $\mathbf{A}$. *Let* $M, M', N, N' \in \mathcal{C}$ *with* $M \cong M'$ *and* $N \cong N'$. *Then* $d(M,N) = d(M',N')$.

Proof. By the triangle inequality $d(M,N) \leqslant d(M,M') + d(M',N') + d(N',N)$. Since we required that $d(A,B) = 0$ for all $A \cong B$ in $\mathcal{C}$, we get that $d(M,N) \leqslant d(M',N')$. Similarly $d(M',N') \leqslant d(M,N)$. □

2.7. Norms on $\mathbb{R}^n$. Since we will frequently use and manipulate p-norms, we briefly recall their definition and a useful basic property.

Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Then for $1 \leqslant p < \infty$

$$\|x\|_p = \left(\sum_{k=1}^{n} |x|^p\right)^{\frac{1}{p}}$$

defines a norm on $\mathbb{R}^n$ as does

$$\|x\|_\infty = \max_{1 \leqslant k \leqslant n} |x|.$$

Lemma 2.21. *If* $x = (x_1, \dots, x_m)$, $y = (y_1, \dots, y_n)$ *and* $z = (x_1, \dots, x_m, y_1, \dots, y_n)$ *then for* $1 \leqslant p \leqslant \infty$,

$$\left\| \left(\|x\|_p, \|y\|_p \right) \right\|_p = \|z\|_p .$$

Proof. For $1 \leqslant p < \infty$, the left hand side equals

$$\left(\|x\|_p^p + \|y\|_p^p\right)^{\frac{1}{p}} = \left(\sum_{i=1}^{m} |x_i|^p + \sum_{j=1}^{n} |y_j|^p\right)^{\frac{1}{p}}$$

which equals $\|z\|_p$. For $p = \infty$, the left hand side equals

$$\max\left(\|x\|_p, \|y\|_p\right) = \max\left(|x_1|, \dots, |x_m|, |y_1|, \dots, |y_n|\right)$$

which equals $\|z\|_p$. □

[2] If we also drop the symmetry requirement then we obtain a Lawvere metric space which may be equivalently defined to be a category enriched in the monoidal poset $(([0,\infty], \geqslant), +, 0)$.

2.8. Wasserstein distances for persistence modules. In this section, we define Wasserstein distances for persistence modules.

These definitions depend on a choice of metric on $\mathbb{R}^2$. While the standard definition of Wasserstein distance [13] uses the ∞-norm, one may use others. If one uses the 1-norm for the 1-Wasserstein distance then one obtains the barcode metric [14, 9], which predates the other Wasserstein distances. In [27], the 2-norm is used. For us, the 1-norm will be the most natural and that is the one we will use.

Let K be a field, and let $\mathbf{A}$ be one of the categories $\mathbf{vect}_K^{\mathbf{n}}$ $\mathbf{vect}_{K,ft}^{(\mathbb{N},\leqslant)}$, $\mathbf{vect}_{K,ft}^{(\mathbb{Z},\leqslant)}$, or $\mathbf{vect}_{K,ft}^{(\mathbb{R},\leqslant)}$.

Let us introduce the following notation which we will use in our definition. Let $\mathcal{I}$ denote the set of nonzero interval modules in $\mathbf{A}$ including the zero module. For a nonzero interval module I, let $x(I) = (\inf\{a : I(a) \neq 0\}, \sup\{a : I(a) \neq 0\}) \in [-\infty,\infty]^2$. Let $\Delta \subset [-\infty,\infty]^2$ denote the diagonal, $\{(x,x) \mid -\infty \leqslant x \leqslant \infty\}$. Given an extended metric d on $[-\infty,\infty]^2$ and $x \in [-\infty,\infty]^2$, let $d(x,\Delta) := \inf_{y\in\Delta} d(x,y)$.

By a *partial matching* between index sets A and B, we mean an injection $\varphi : C \to B$, where $C \subset A$.

Definition 2.22. Let $1 \leqslant p, q \leqslant \infty$. Denote by d the extended metric on $[-\infty,\infty]^2$ given by $d(x,y) = \|x-y\|_q$. Let $M, N \in \mathbf{A}$. Assume $M \cong \oplus_{i\in A} I_a$ and $N \cong \oplus_{j\in B} I'_b$, where each I_a and I'_b is an interval module. Define

$$W_p^q(M,N) = \min_{\varphi:C\to B} \left\| \left(\left\| \left(d(x(I_c), x(I'_{\varphi(c)})) \right)_{c\in C} \right\|_p, \left\| (d(x(I_a),\Delta))_{a\in A-C} \right\|_p, \left\| \left(d(\Delta, x(I'_b)) \right)_{b\in B-\varphi(C)} \right\|_p \right) \right\|_p,$$

where the minimum is over all partial matchings φ between A and B. Call this the (p,q)-*Wasserstein distance* between the persistence modules M and N, or simply the p-*Wasserstein distance*, where q is understood.

2.9. Zigzags of morphisms. Let $\mathbf{A}$ be a category. Let $M, N \in \mathbf{A}$. A *zigzag* of morphisms from M to N is a finite collection of morphisms in $\mathbf{A}$ of the form $M \xleftarrow{f_1} M_1 \xrightarrow{f_2} M_2 \xleftarrow{f_3} M_3 \xrightarrow{f_4} \cdots N$. We will use exponents to indicate the direction of the maps. So the above zigzag will be written $f_1^{-1} f_2 f_3^{-1} f_4 \ldots$. We will also omit identity maps, so we will write a general zigzag from M to N as $f_1^{\varepsilon_1} f_2^{\varepsilon_2} \cdots f_n^{\varepsilon_n}$ for some n and $\varepsilon_1, \ldots, \varepsilon_n \in \pm 1$. Where possible, we can compose zigzags by concatenating them.

3. Wasserstein distance for persistence modules

In this section we show that if we use the 1-norm on $\mathbb{R}^2$, then the Wasserstein distance can be succinctly stated using counting measure on $\mathbb{Z}$ or the Lebesgue measure on $\mathbb{R}$.

Let λ denote the counting measure on $\mathbb{Z}$ or the Lebesgue measure on $\mathbb{R}$. Recall that in $\mathbf{vect}_K^n$, $\mathbf{vect}_K^{(\mathbb{N},\leqslant)}$, or $\mathbf{vect}_K^{(\mathbb{Z},\leqslant)}$, the zero module is assumed to correspond to the empty interval in $\mathbb{R}$ and a nonzero interval module is assumed to correspond to an interval in $\mathbb{R}$ of the form $[a,b)$ where $a,b \in \mathbb{Z}$ (Assumption 2.11). So the two measures agree: $\lambda(\emptyset) = 0$ and $\lambda([a,b)) = b - a$.

Given two intervals I and J in $\mathbb{R}$, let $I \triangle J$ denote their symmetric difference. That is, $I \triangle J = (I \cup J) \setminus (I \cap J)$.

Recall that W_p^q denotes the p-Wasserstein distance for persistence modules using the underlying q-norm for $\mathbb{R}^2$ (Definition 2.22).

We will prove the following two results.

Proposition 3.1. *For interval modules I and J, $W_1^1(I,J) = \lambda(I \triangle J)$.*

Proposition 3.2. *Let M and N be persistence modules. Then*

$$W_p^1(M,N) = \min \|(\lambda(M_k \triangle N_k))_k\|_p, \tag{3.3}$$

where the minimum is taken over all isomorphisms $M \cong \oplus_k M_k$ and $N \cong \oplus_k N_k$, where M_k and N_k are interval modules.

For $x, y \in [-\infty,\infty]^2$, let $d(x,y) = \|x-y\|_1 = |x_1 - y_1| + |x_2 - y_2|$. Recall that for an interval module I, $x(I) = (\inf I, \sup I)$ and that Δ denotes the diagonal in $[-\infty,\infty]^2$. Also, $d(u,\Delta)$ denotes the minimum distance from u to a point in Δ.

Lemma 3.4. *Let I be a nonempty interval. Then $d(x(I),\Delta) = \lambda(I)$.*

Proof. Note that $\lambda(I) = \sup I - \inf I$. Now, for any $c \in \mathbb{R}$, $d((a,b),(c,c)) = |a-c| + |b-c| \geqslant |a-b|$ by the triangle inequality. Since $d((a,b),(b,b)) = |a-b|$, $d((a,b),\Delta) = b-a$. □

Lemma 3.5. *If I, J are nonempty intervals with $I \cap J \neq \emptyset$, then $d(x(I),x(J)) = \lambda(I \triangle J)$.*

Proof. There are a number of cases for intervals I and J with $I \cap J \neq \emptyset$. However, in each case, $\lambda(I \triangle J) = |\sup I - \sup J| + |\inf I - \inf J|$. The latter equals $\|x(I) - x(J)\|_1 = d(x(I),x(J))$. □

Lemma 3.6. *If I and J are nonempty intervals with $I \cap J = \emptyset$, then $d(x(I),x(J)) \geqslant \lambda(I) + \lambda(J)$.*

Proof. Without loss of generality, assume that $\inf(I) \leqslant \sup(I) \leqslant \inf(J) \leqslant \sup(J)$. Then $d(x(I),x(J)) = \sup(J) - \sup(I) + \inf(J) - \inf(I) \geqslant \sup(J) - \inf(J) + \sup(I) - \inf(I) = \lambda(I) + \lambda(J)$. □

Proof of Proposition 3.1. There are only two partial matchings between I and J: one in which I and J are matched to one another, and one in which I and J are both matched to the diagonal. So by Definition 2.22 and Lemma 3.4,

$$\begin{aligned} W_1^1(I,J) &= \min\left(d(x(I),x(J)), d(x(I),\Delta)+d(\Delta,x(J))\right) \\ &= \min\left(d(x(I),x(J)), \lambda(I)+\lambda(J)\right). \end{aligned}$$

If $I \cap J \neq \emptyset$, then by Lemma 3.5, $d(x(I),x(J)) = \lambda(I \triangle J) \leqslant \lambda(I)$ and $\lambda(J)$, so $W_1^1(I,J) = \lambda(I \triangle J)$. If $I \cap J = \emptyset$, then by Lemma 3.6 it follows that $W_1^1(I,J) = \lambda(I)+\lambda(J) = \lambda(I \triangle J)$. $\square$

Proof of Proposition 3.2. Let $1 \leqslant p \leqslant \infty$. Assume $M \cong \oplus_{i \in A} I_a$ and $N \cong \oplus_{j \in B} I'_b$, where each I_a and I'_b is an interval module. By Definition 2.22 and Lemma 3.4,

$$W_p^1(M,N) = \min_{\varphi} \left\| \left(\left\| \left(d(x(I_c), x(I'_{\varphi(c)})) \right)_{c \in C} \right\|_p, \left\| (\lambda(I_a))_{i \in A-C} \right\|_p, \left\| \left(\lambda(I'_b) \right)_{j \in B-\varphi(C)} \right\|_p \right) \right\|_p,$$

where the minimum is over all partial matchings φ between A and B. By Lemma 3.6, this minimum is achieved by a partial matching $\varphi : C \to B$ with the property that $I_c \cap I'_{\varphi(c)} \neq \emptyset$ for all $c \in C$ (where it could well be that $C = \varnothing$). Thus, by Lemma 3.5,

$$W_p^1(M,N) = \min_{\varphi} \left\| \left(\left\| \left(\lambda(I_c \triangle I'_{\varphi(c)}) \right)_{c \in C} \right\|_p, \left\| (\lambda(I_a \triangle \emptyset))_{i \in A-C} \right\|_p, \left\| \left(\lambda(\emptyset \triangle I'_b) \right)_{j \in B-\varphi(C)} \right\|_p \right) \right\|_p.$$

Writing this more compactly we obtain the statement in the proposition. $\square$

4. Path metrics

In this section we use kernels and cokernels to define a large class of metrics on abelian categories.

4.1. Path metrics for abelian categories.

We define a certain graph distance for abelian categories that we will call a path metric.

Definition 4.1.
1. Let $\mathbf{A}$ be an abelian category. Let $\mathcal{S} \subset \operatorname{Ob} \mathbf{A}$ such that $0 \in \mathcal{S}$. Let $w : \mathcal{S} \to [0,\infty)$ with $w(0) = 0$.
2. Let $\mathcal{F}$ denote the set of morphisms f in $\mathbf{A}$ such that either $\ker(f) = 0$ with $\operatorname{coker}(f) \in \mathcal{S}$ or $\operatorname{coker}(f) = 0$ with $\ker(f) \in \mathcal{S}$. Define $w : \mathcal{F} \to [0,\infty)$ by setting $w(f) = w(\ker(f)) + w(\operatorname{coker}(f))$. Note that at least one of these is 0.
3. Let Γ consist of zigzags of arrows (Section 2.9) in $\mathcal{F}$. For a zigzag $\gamma = f_1^{\varepsilon_1} \cdots f_n^{\varepsilon_n}$ where $\varepsilon_k = \pm 1$ and $f_k \in \mathcal{F}$, let $w(\gamma) = \sum_{k=1}^n w(f_k)$. This defines $w : \Gamma \to [0,\infty)$.
4. Define $d_w : \operatorname{Ob} \mathbf{A} \times \operatorname{Ob} \mathbf{A} \to [0,\infty]$ by $d_w(M,N) = \inf_\gamma w(\gamma)$, where the infimum is taken over all zigzags in Γ from M to N. If there are no such zigzags then $d_w(M,N) = \infty$.

Lemma 4.2. *The distance d_w is a metric (Definition 2.19) on $\mathbf{A}$, which we call a* path metric.

Proof. Let $f : M \xrightarrow{\cong} N$. Then $w(f) = w(\ker f) + w(\operatorname{coker} f) = w(0) + w(0) = 0$. Therefore $d_w(M, N) = 0$. It is symmetric since for any zigzag we have the reverse zigzag, which has the same weight. It satisfies the triangle inequality since we can concatenate zigzags. $\square$

4.2. Path metrics for persistence modules. In this section we will define path metrics for persistence modules and give a succinct formula for the distance between a pair of interval modules.

Let $\mathbf{A}$ be be one of the categories $\mathbf{vect}_{K,ft}^{(\mathbb{Z},\leqslant)}$, $\mathbf{vect}_{K,ft}^{(\mathbb{N},\leqslant)}$, or $\mathbf{vect}_K^{\mathbf{n}}$.

Definition 4.3. Let $\mathcal{S}$ denote the set of simple modules in $\mathbf{A}$ together with 0. Define $w : \mathcal{S} \to [0, \infty)$, by $w(0) = 0$ and for $0 \neq S \in \mathcal{S}$, $w(S) = 1$. Let d_w be the corresponding path metric on $\mathbf{A}$ (Definition 4.1 and Lemma 4.2).

Let $\mathbf{A}$ be the category $\mathbf{vect}_{K,ft}^{(\mathbb{R},\leqslant)}$. Let $\mathcal{I}$ denote the set of interval modules in $\mathbf{A}$ together with 0. We will use the same notation for an interval $I \subset \mathbb{R}$ and the corresponding interval module $I \in \mathbf{A}$. The empty interval corresponds to the zero module.

Definition 4.4. Define $w : \mathcal{I} \to [0, \infty)$, by $w(I) = \lambda(I)$ where λ denotes the Lebesgue measure on $\mathbb{R}$. Let d_w be the corresponding path metric on $\mathbf{A}$ (Definition 4.1 and Lemma 4.2).

We will prove the following.

Theorem 4.5. *For the path metrics in Definitions 4.3 and 4.4 and interval modules I and J, $d_w(I, J) = \lambda(I \triangle J)$, where $I \triangle J$ denotes the symmetric difference $(I \cup J) \setminus (I \cap J)$ and where λ denotes the Lebesgue measure.*

Together with Proposition 3.1 we have the following.

Corollary 4.6. *Restricted to interval modules, the path metrics in Definitions 4.3 and 4.4 equal the Wasserstein distance W_1^1.*

For a persistence module M, let $\dim(M)$ denote its dimension vector.

Lemma 4.7. *Let $0 \to M \to N \to P \to 0$ be a short exact sequence of persistence modules. Then* $\dim(N) = \dim(M) + \dim(P)$.

Proof. Since kernels and cokernels of persistence modules are computed pointwise, the result follows from the rank-nullity theorem. $\square$

It is a good exercise to check the following (or see [6, Appendix A]).

Lemma 4.8. *Let I and J be interval modules. Then, after possibly interchanging I and J, we have one of the following two possible cases.*

(1) *There are maps* $I \xrightarrow{f} I \cap J \xrightarrow{g} J$ *with* f *surjective,* $\ker(f) = I \setminus (I \cap J)$*,* g *injective, and* $\operatorname{coker}(g) = J \setminus (I \cap J)$*. (This includes the case* $I \cap J = \emptyset$*.)*
(2) $I \subset J$ *and there is an interval module* K *and maps* $I \xleftarrow{f} K \xrightarrow{g} J$ *with* f *surjective,* g *injective and* $J \setminus I$ *is the disjoint union of* $\ker(f)$ *and* $\operatorname{coker}(g)$*.*

4.2.1. *Path metrics for discrete persistence modules.* First we prove Theorem 4.5 for discrete persistence modules.

In this section we assume that all persistence modules objects in one of the categories, $\mathbf{vect}_K^n$, $\mathbf{vect}_{K,ft}^{(\mathbb{N},\leqslant)}$, or $\mathbf{vect}_{K,ft}^{(\mathbb{Z},\leqslant)}$.

Let P be a countable set. For $f : P \to \mathbb{Z}_{\geqslant 0}$, let $\sum f = \sum_{a \in P} f(a)$ if the sum is well defined, and $\sum f = \infty$ otherwise.

From Lemma 4.7 we have the following corollaries.

Corollary 4.9. *Let* γ *be a zigzag of persistence modules of length* n *from* M *to* N*. Then* $\sum |\dim(M) - \dim(N)| \leqslant n$*.*

Proof. By Definition 4.3 and Lemma 4.7, for each step in a zigzag, the dimension vector can change by at most one in one coordinate. □

Corollary 4.10. *Let* M *and* N *be persistence modules. Then* $d_w(M, N) \geqslant \sum |\dim(M) - \dim(N)|$*.*

Proof. By Corollary 4.9, any zigzag from M to N has length at least $\sum |\dim(M) - \dim(N)|$. The result follows from Definition 4.1. □

Next we consider interval modules. As a special case of Corollary 4.10, we have the following.

Corollary 4.11. *Let* I *and* J *be interval modules. Then* $d_w(I, J) \geqslant \sum \dim(I \triangle J)$*.*

Lemma 4.12. *Let* I *and* J *be interval modules and let* λ *denote the Lebesgue measure.*

(1) *If there a surjection* $f : I \twoheadrightarrow J$ *then* $d_w(I, J) \leqslant \lambda(I \setminus J)$*.*
(2) *If there a injection* $f : I \hookrightarrow J$ *then* $d_w(I, J) \leqslant \lambda(J \setminus I)$*.*

Proof. In the first case, by Lemma 4.8, $J = [a, b)$ and $I = [a, b+n)$ for some $a \leqslant b$ and $n \geqslant 0$. Then there is a zigzag of length n from I to J given by a sequence of surjections each with simple kernel, $[a, b+n) \to [a, b+n-1) \to [a, b+n-2) \to \cdots \to [a, b)$. Thus $d_w(I, J) \leqslant n$.

In the second case, by Lemma 4.8, $I = [a, b)$ and $J = [a-n, b)$ for some $a \leqslant b$ and $n \geqslant 0$. Then there is a zigzag of length n from I to J given by a sequence of injections each with simple cokernel, $[a, b) \to [a-1, b) \to [a-2, b) \to \cdots \to [a-n, b)$. Thus $d_w(I, J) \leqslant n$. □

Proposition 4.13. *For* $I, J \in \mathcal{I}$, $d_w(I,J) = \lambda(I \triangle J) = \lambda(I) + \lambda(J) - 2\lambda(I \cap J)$.

Proof. Let I and J be interval modules. By Lemma 4.8, we have a pair of maps connecting I and J. Combining the triangle inequality and Lemma 4.12, we have that $d_w(I,J) \leqslant \lambda(I \triangle J)$. Together with Corollary 4.11, we have that $d_w(I,J) = \lambda(I \triangle J)$.

Finally, $\lambda((I - (I \cap J)) \cup (J - (I \cap J))) = \lambda(I - (I \cap J)) + \lambda(J - (I \cap J)) = \lambda(I) + \lambda(J) - 2\lambda(I \cap J)$. □

4.2.2. *Path metrics for continuous persistence modules.* Next we prove Theorem 4.5 for continuous persistence modules.

In this section we assume that all persistence modules are objects in the category $\mathbf{vect}_{K,ft}^{(\mathbb{R},\leqslant)}$.

For $f : \mathbb{R} \to \mathbb{Z}_{\geqslant 0}$, let $\int f = \int f(t)\,dt$ if the integral is well defined, and $\int f = \infty$ otherwise.

From Lemma 4.7 we have the following corollaries.

Corollary 4.14. *Let* γ *be a zigzag of persistence modules from* M *to* N. *Then* $\int |\dim(M) - \dim(N)| \leqslant w(\gamma)$.

Proof. By Lemma 4.7, for each $f : M' \to M'' \in \mathcal{F}$ (Definition 4.1), $w(f) = \int |\dim(M') - \dim(M'')|$.

Let $\gamma = f_1^{\varepsilon_1} \cdots f_n^{\varepsilon_n}$ be zigzag from M to N with f_i a morphism (in either direction) between M_{i-1} and M_i. Then

$$\begin{aligned}
\dim(M) - \dim(N) &= \dim(M_0) - \dim(M_1) + \cdots + \dim(M_{n-1}) - \dim(M_n) \\
|\dim(M) - \dim(N)| &\leqslant |\dim(M_0) - \dim(M_1)| + \cdots + |\dim(M_{n-1}) - \dim(M_n)| \\
\int |\dim(M) - \dim(N)| &\leqslant \int |\dim(M_0) - \dim(M_1)| + \cdots + \int |\dim(M_{n-1}) - \dim(M_n)| \\
&= w(\gamma).
\end{aligned}$$
□

Corollary 4.15. *Let* M *and* N *be persistence modules. Then* $d_w(M,N) \geqslant \int |\dim(M) - \dim(N)|$.

Proof. The result follows from Corollary 4.14 and Definition 4.1. □

As a special case we have the following.

Corollary 4.16. *Let* I *and* J *be interval modules. Then* $d_w(I,J) \geqslant \int \dim(I \triangle J)$.

Proposition 4.17. *For* $I, J \in \mathcal{I}$, $d_w(I,J) = \lambda(I \triangle J) = \lambda(I) + \lambda(J) - 2\lambda(I \cap J)$, *where* $I \triangle J$ *denotes the symmetric difference* $(I \setminus (I \cap J)) \cup (J \setminus (I \cap J))$ *and where* λ *denotes the Lebesgue measure.*

PROOF. Let I and J be interval modules. By Lemma 4.8 we have a zigzag γ of length 2 between I and J with $w(\gamma) = \int \dim(I \triangle J)$. Thus $d_w(I, J) \leqslant \int \dim(I \triangle J)$. Combining this with Corollary 4.16, we have that $d_w(I, J) = \int \dim(I \triangle J)$.

Since $\dim(I \triangle J)$ is the indicator function on $I \triangle J$, $\int \dim(I \triangle J) = \lambda(I \triangle J)$. □

5. Wasserstein distances for Krull-Schmidt categories

In this section we generalize the p-Wasserstein distance for persistence modules (Section 2.8) to Krull-Schmidt categories. We assume that we have a metric on the class of indecomposable objects together with the zero object, and then extend this to a metric on all objects. The main result of the section is Theorem 5.13, which states that the Wasserstein is universal among metrics that agree on the indecomposable objects.

Let $\mathbf{A}$ be a Krull-Schmidt category. Let $\mathcal{I}$ denote the class of indecomposable objects in $\mathbf{A}$ together with the zero object. Let $1 \leqslant p \leqslant \infty$.

DEFINITION 5.1. Let d be a metric (Definition 2.19) on $\mathcal{I}$. For $M, N \in \mathbf{A}$, define

$$W_p(d)(M, N) = \min \|(d(M_k, N_k))\|_p, \tag{5.2}$$

where the minimum is taken over all isomorphisms $M \cong \oplus_k M_k$ and $N \cong \oplus_k N_k$, where M_k and N_k are in $\mathcal{I}$. Call $W_p(d)$ a p-*Wasserstein* distance.

REMARK 5.3. Recall that M and N are isomorphic to a finite sum of indecomposables that is unique of to isomorphism and reordering. Note that the direct sum in Definition 5.1 also allows zero objects. So the minimum in (5.2) is over all partial matchings of the indecomposable direct summands of M and N, where the unmatched direct summands are matched with the zero object. We make this precise in the following lemma.

LEMMA 5.4. *Let $M, N \in \mathbf{A}$. Assume $M \cong \oplus_{a \in A} M_a$ and $N \cong \oplus_{b \in B} N_b$, where each M_a and N_b is indecomposable. Let d be a metric on $\mathcal{I}$. Then*

$$W_p(d)(M, N) = \min_{\varphi} \left\| \left(\left\| (d(M_c, N_{\varphi(c)}))_{c \in C} \right\|_p, \left\| (d(M_a, 0))_{a \in A - C} \right\|_p, \left\| (d(0, N_b))_{b \in B - \varphi(C)} \right\|_p \right) \right\|_p,$$

where the minimum is over all partial matchings: $C \subset A$ and $\varphi : C \to B$ is injective.

By Proposition 3.2 and Theorem 4.5, we have the following.

THEOREM 5.5. *For the path metrics d_w in Definitions 4.3 and 4.4 and persistence modules M and N,*

$$W_p^1(M, N) = W_p(d_w)(M, N).$$

LEMMA 5.6. *Restricted to $\mathcal{I}$, $W_p(d)$ equals d.*

PROOF. Let $M, N \in \mathcal{I}$. Then $W_p(d)(M, N) = \min \left(d(M_1, N_1), \|(d(M_1, 0), d(0, N_2))\|_p \right)$, where $M_1 \cong M$ and $N_1 \cong N_2 \cong N$. By Lemma 2.20 and the triangle inequality, this equals $d(M, N)$. □

PROPOSITION 5.7. *For any metric d on $\mathcal{I}$, $W_p(d)$ is a metric on $\mathbf{A}$.*

Proof. Since $d(I,I) = 0$ for all $I \in \mathcal{I}$, it follows that $W_p(d)(M,M) = 0$ for all $M \in \mathbf{A}$. Since d is symmetric, it follows that $W_p(d)$ is symmetric.

The proof of the triangle inequality uses the Krull-Schmidt property. Let $M, N, P \in \mathbf{A}$. By including sufficiently many zero modules and reordering the direct summands, we may assume that $M \cong \oplus_k M_k$, $N \cong \oplus_k N_k$, $P \cong \oplus_k P_k$, and that $W_p(d)(M,N) = \|(d(M_k,N_k))_k\|_p$ and $W_p(d)(N,P) = \|(d(N_k,P_k))_k\|_p$. Then

$$\begin{aligned} W_p(d)(M,P) &\leqslant \|(d(M_k,P_k))_k\|_p \leqslant \|(d(M_k,N_k) + d(N_k,P_k))_k\|_p \\ &\leqslant \|(d(M_k,N_k))_k\|_p + \|(d(N_k,P_k))_k\|_p = W_p(M,N) + W_p(N,P), \end{aligned}$$

where the first inequality is by definition, the second inequality is by the triangle inequality for d, and the third inequality is by the Minkowski inequality.

Finally, suppose that $M \cong N$. Then there are isomorphisms $M \cong \oplus_k M_k$ and $N \cong \oplus_k N_k$ where for all k, $M_k \cong N_k$. Since d is a metric on $\mathcal{I}$, $d(M_k, N_k) = 0$, and so $W_p(d)(M,N) = 0$. □

Definition 5.8. Call a metric d on $\mathbf{A}$ p-*subadditive* if for any $M, M', N, N' \in \mathbf{A}$, $d(M \oplus M', N \oplus N') \leqslant \|(d(M,N), d(M',N'))\|_p$.

Lemma 5.9. *If a metric d is p-subadditive and $M \cong \oplus_{k=1}^n M_k$ and $N \cong \oplus_{k=1}^n N_k$ then*

$$d(M,N) \leqslant \|(d(M_k,N_k))_{k=1}^n\|_p .$$

Proof. The proof is by induction on n. If $n = 1$ then we have equality. For the inductive step,

$$\begin{aligned} d(M,N) &\leqslant \left\| \left(d(\oplus_{k=1}^{n-1} M_k, \oplus_{k=1}^{n-1} N_k), d(M_n,N_n) \right) \right\|_p \\ &\leqslant \left\| \left(\left\| (d(M_k,N_k))_{k=1}^{n-1} \right\|_p, d(M_n,N_n) \right) \right\|_p = \|(d(M_k,N_k))_{k=1}^n\|_p , \end{aligned}$$

where the first inequality is by definition and the second is by the induction hypothesis. □

Proposition 5.10. *For any metric d on $\mathcal{I}$, $W_p(d)$ is p-subadditive.*

Proof. Fix objects M, M', N, N'. Then there exist isomorphisms $M \cong \oplus_k M_k$, $N \cong \oplus_k N_k$, $M' \cong \oplus_j M'_j$, and $N' \cong \oplus_j N'_j$ such that $W_p(d)(M,N) = \|(d(M_k,N_k))\|_p$ and $W_p(d)(M',N') = \left\|(d(M'_j,N'_j))\right\|_p$. Then by Definition 5.1,

$$\begin{aligned} W_p(d)(M \oplus M', N \oplus N') &\leqslant \left\| (\|(d(M_k,N_k))\|_p, \left\|(d(M'_j,N'_j))\right\|_p) \right\|_p \\ &= \left\| (W_p(d)(M,N), W_p(d)(M',N')) \right\|_p . \end{aligned}$$ □

Proposition 5.11. *For any p-subadditive metric d on $\mathbf{A}$, $d \leqslant W_p(d|_{\mathcal{I}})$.*

Proof. Assume $M \cong \oplus_{a\in A} M_a$ and $N \cong \oplus_{b\in B} N_b$ where each M_a and N_b is indecomposable. For all partial matchings $C \subset A$, $\varphi : C \to B$,

$$d(M,N) = d\left(\bigoplus_{c\in C} M_c \oplus \bigoplus_{a\in A-C} M_a \oplus \bigoplus_{b\in B-\varphi(C)} 0, \bigoplus_{c\in C} N_{\varphi(c)} \oplus \bigoplus_{a\in A-C} 0 \oplus \bigoplus_{b\in B-\varphi(C)} N_b\right)$$
$$\leqslant \left\| \left(\left\| \left(d(M_c, N_{\varphi(c)})\right)_{c\in C} \right\|_p, \left\| (d(M_a,0))_{a\in A-C} \right\|_p, \left\| (d(0,N_b))_{b\in B-\varphi(C)} \right\|_p \right) \right\|_p,$$

by Lemma 5.9. Therefore by Lemma 5.4, $d(M,N) \leqslant W_p(d|_{\mathcal{I}})(M,N)$. □

The last two propositions give us an alternative definition of Wasserstein distance.

Corollary 5.12. *Let* d *be a metric on* $\mathcal{I}$ *and let* $M, N \in \mathbf{A}$. *Then*

$$W_p(d)(M,N) = \sup\{D(M,N) \mid \textit{where } D \textit{ is } p\textit{-subadditive metric on } \mathbf{A} \textit{ and } D|_{\mathcal{I}} = d\}.$$

Proof. ($\leqslant$) Proposition 5.10 and Lemma 5.6 imply that $W_p(d)$ is an element of the set of the right hand side. ($\geqslant$) Proposition 5.11 implies $W_p(d)$ is greater than or equal to the each term on the right hand side. □

Combining these results have the following.

Theorem 5.13. *For any metric* d *on* $\mathcal{I}$, $W_p(d)$ *is a* p*-subadditive metric on* $\mathbf{A}$ *that agrees with* d *on* $\mathcal{I}$. *Furthermore it is the largest such metric.*

6. Agreement of 1-Wasserstein distance and path metric for persistence modules

Let $\mathbf{A}$ be one of the categories $\mathbf{vect}_{K,ft}^{(\mathbb{R},\leqslant)}$, $\mathbf{vect}_{K,ft}^{(\mathbb{Z},\leqslant)}$, $\mathbf{vect}_{K,ft}^{(\mathbb{N},\leqslant)}$, or $\mathbf{vect}_K^{\mathbf{n}}$. Let d_w be the path metric on $\mathbf{A}$ defined in Section 4.2. The goal of this section is to establish the following.

Theorem 6.1. *Let* d_w *be the path metric on* $\mathbf{A}$ *defined in Section 4.2. Then* $W_1(d_w) = d_w$.

One direction is easy.

Lemma 6.2. *The path metric* d_w *is* 1*-subadditive. That is,* $d_w(M\oplus M', N\oplus N') \leqslant d_w(M,N) + d(M',N')$.

Proof. Let $\gamma = f_1^{\varepsilon_1}\cdots f_n^{\varepsilon_n}$ be a zigzag from M to N and let $\gamma' = g_1^{\delta_1}\cdots g_m^{\delta_m}$ be a zigzag from M' to N'. Then we have the zigzag $(f_1^{\varepsilon_1}\oplus 1_{M'})\cdots(f_n^{\varepsilon_n}\oplus 1_{M'})(1_N\oplus g_1^{\delta_1})\cdots(1_N\oplus g_m^{\delta_m})$ from $M\oplus M'$ to $N\oplus N'$. The result follows. □

Combining Lemma 6.2 and Corollary 5.12, we have that $W_1(d_w) \geqslant d_w$. It remains to show that $W_1(d_w) \leqslant d_w$. We will use the remainder of this section to establish this result.

LEMMA 6.3. *Let* I *and* J *be interval modules with* $\inf I \geqslant \inf J$. *Let* $a \in I \cap J$. *A nonzero map* $\phi(a) : I(a) \to J(a)$ *extends to a unique map* $\phi : I \to J$.

PROOF. Assume $a \leqslant b$ and $b \in I \cap J$. Then let $\phi(b) = J(a \leqslant b) \circ \phi(a) \circ I(a \leqslant b)^{-1}$. Similarly if $c \leqslant a$ and $c \in I \cap J$, then $\phi(c) = J(c \leqslant a)^{-1} \circ \phi(a) \circ I(c \leqslant a)$. For all $c \notin I \cap J$ let $\phi(c) = 0$. Then by construction, $\phi : I \to J$ is a map of persistence modules. □

DEFINITION 6.4. Let I be an interval module. A *coherent basis* for I is a choice of nonzero element $x_a \in I(a)$ for each $a \in I$ such that $I(a \leqslant b)(x_a) = x_b$ whenever $a, b \in I$ and $a \leqslant b$.

Let I be an interval module. For any $a \in I$, then any nonzero $x_a \in I(a)$ determines a unique coherent basis, with

$$x_b = \begin{cases} I^{-1}_{(b \leqslant a)}(x_a) & b \leqslant a \\ I_{(a \leqslant b)}(x_a) & a \leqslant b. \end{cases}$$

LEMMA 6.5. *Let* $\phi : I \to J$ *be a nonzero map. Any choice of coherent basis for* I *determines a unique coherent basis for* J *such that* $\phi(a) = 1$ *whenever it is nonzero (i.e. whenever* $a \in I \cap J$*). Similarly, any coherent basis for* J *determines a unique coherent basis for* I *such that* $\phi(a) = 1$ *whenever it is nonzero.*

PROOF. Since ϕ is nonzero, $I \cap J \neq \varnothing$, and for $a \in I \cap J$, ϕ_a is an isomorphism. Let $\{x_t : t \in I\}$ be a coherent basis for I, and choose $a \in I \cap J$. Then a unique coherent basis for J is determined by $\phi_a(x_a)$.

Similarly, if $\{y_t : t \in J\}$ is a coherent basis for J, and $a \in I \cap J$, then a unique coherent basis for I is determined by $\phi_a^{-1}(y_a)$. □

Given a persistence module of the form $L = N \oplus \bigoplus_{j=1}^{n} [a_j, b_j)$, let $i_N : N \to L$, $p_N : L \to N$ denote the injection and projection maps. Similarly, for $j = 1, \ldots, n$, set $i_j : [a_j, b_j) \to L$ and $p_j : L \to [a_j, b_j)$.

THEOREM 6.6. *Given the notation above, if* $f \colon [a, b) \to M$ *is nonzero, then there exists* $a_1 < a_2 \cdots < a_n \leqslant a < b_n < \cdots < b_2 < b_1 \leqslant b$, *a module* N *and an isomorphism* $\theta \colon M \to N \oplus_{j=1}^{n} [a_j, b_j)$ *such that* $p_N f = 0$ *and* $p_j f$ *is nonzero for all* j. *Furthermore,* $\ker f \cong [b_1, b)$ *and*

$$\operatorname{coker} f \cong N \oplus [a_1, b_2) \oplus \cdots \oplus [a_{n-1}, b_n) \oplus [a_n, a).$$

PROOF. There is an isomorphism $\theta : M \to \oplus_{i=1}^{m} [a_i, b_i)$. By reordering summands if necessary, we may suppose that $p_i \theta f \neq 0$ if and only if $i \leqslant n$. Let $N = \oplus_{i=n+1}^{m} [a_i, b_i)$. Note that n may vary for different choices of θ. However, since f is nonzero, $n \geqslant 1$. Assume that we have made a choice for θ such that n is minimal. Order the direct summands so that $b_1 \geqslant b_2 \geqslant \cdots \geqslant b_n$. Since $p_j \theta f : [a, b) \to [a_j, b_j)$ is nonzero, $a \in [a_j, b_j)$.

If $n = 1$ then $f = (i_N p_N + i_1 p_1)\theta f = i_1 p_1 \theta f$, so $f_t = 0$ if and only if $p_1 \theta f_t = 0$ if and only if $t \notin [a, b) \cap [a_1, b_1)$. Since $a_1 \leqslant a < b_1 \leqslant b$, $\ker f = [b_1, b)$ and $\operatorname{coker} f \cong N \oplus [a_1, a)$, as desired.

Now assume that $n \geqslant 2$. We will first show that the minimality of n implies that $b_i > b_{i+1}$ for all i. Suppose to the contrary that $b_k = b_{k+1}$. Without loss of generality, we may assume that $a_k \geqslant a_{k+1}$.

For $j = k, k+1$, $p_j \theta f : [a, b) \to [a_j, b_j)$ is nonzero and $a \in [a_j, b_j)$. So $p_j \theta f(a)$ is a nonzero map of one-dimensional vector spaces and hence invertible. Let $\alpha : [a_k, b_k) \to [a_{k+1}, b_{k+1})$ be the nonzero map given by Lemma 6.3 determined by the map $(p_{k+1}\theta f)(a) \circ (p_k \theta f)(a)^{-1}$ using Lemma 6.3.

Let $\phi\colon \oplus_{i=1}^{n} [a_i, b_i) \oplus N \to \oplus_{i=1}^{n} [a_i, b_i) \oplus N$ to be the unique map determined by the rules $\phi i_N = i_N$ and $\phi i_j = i_j$ for $j \neq k$, while

$$p_j \phi i_k = \begin{cases} 1 & j = k \\ -\alpha & j = k+1 \\ 0 & \text{otherwise.} \end{cases}$$

We claim that ϕ is an isomorphism. Let $M' = [a_k, b_k) \oplus [a_{k+1}, b_{k+1})$ and $M'' = N \oplus \bigoplus_{j=1}^{k-1} [a_j, b_j) \oplus \bigoplus_{j=k+2}^{n} [a_j, b_j)$. By construction, $\theta^{-1}\phi$ decomposes into the identity map on M'' and a nonidentity map ϕ' on M'. The map ϕ' is nontrivial for $a_k \leqslant a \leqslant b_k = b_{k+1}$. For each such a, one can choose a basis for $M'(a)$ such that $\phi'(a)$ is given by $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ which is an isomorphism.

Consider the following computation.

$$\begin{aligned} p_{k+1}\phi\theta f(a) &= p_{k+1}\phi i_k p_k \theta f(a) + p_{k+1}\phi i_{k+1} p_{k+1}\theta f(a) \\ &= -p_{k+1}\theta f(a) \circ p_k \theta f(a)^{-1} \circ p_k \theta f(a) + p_{k+1}\theta f(a) \\ &= 0. \end{aligned}$$

Therefore $p_{k+1}\phi\theta f = 0$. Thus, the isomorphism $\phi\theta$ contradicts the minimality of n.

The exact same argument shows that our assumptions imply that for all i, $a_i < a_{i+1}$.

Choose a coherent basis for $[a, b)$. By Lemma 6.5, for each j, there is a coherent basis for $[a_j, b_j)$ such that $p_j \theta f(c) = 1$ whenever $c \in [a, b) \cap [a_j, b_j)$. Let K denote $[a, b)$ and let M_j denote $[a_j, b_j)$. Then for $c \in [a, b)$,

$$\theta f(c) : 1_{K(c)} \mapsto \sum_{\substack{1 \leqslant j \leqslant n \\ c \in [a_j, b_j) \cap [a, b)}} 1_{M_j(c)}.$$

Hence $\ker f \cong \ker \theta f \cong [b_1, b)$.

Let $p = p_1 + \cdots + p_n$. Then $\operatorname{coker} f \cong \operatorname{coker} \theta f \cong N \oplus p\theta f$.

Consider the following basis for the vector spaces in $\bigoplus_{j=1}^{n}[a_j, b_j)$. For $k = 1, \dots, n$ and $c \in [a_k, b_k) \setminus [a_{k+1}, b_{k+1})$ (where $[a_{n+1}, b_{n+1}) := \emptyset$), $\bigoplus_{j=1}^{n}[a_j, b_j)(c)$ has basis

$$\left\{ 1_{M_1(c)}, 1_{M_1(c)} + 1_{M_2(c)}, \dots, \sum_{j=1}^{k} 1_{M_j(c)} \right\}.$$

Thus, if $c \in [a_k, b_k) \setminus [a_{k+1}, b_{k+1})$ then $\operatorname{coker} p\theta f(c)$ has basis $[1_{M_1(c)}]$, $[1_{M_1(c)} + 1_{M_2(c)}]$, $\dots$, $[\sum_{j=1}^{k-1} 1_{M_j(c)}]$. Therefore $\operatorname{coker} f \cong N \oplus [a_1, b_2) \oplus [a_2, b_3) \oplus \cdots \oplus [a_{n-1}, b_n) \oplus [a_n, a)$. □

Corollary 6.7. *Given a short exact sequence $0 \to [a, b) \to M \to N \to 0$, it follows that $W_1(d_w)(M, N) \leqslant b - a$.*

Proof. Applying Theorem 6.6, we can compare the direct sum decompositions of M and N. It follows that $W_1(d_w)(M, N) \leqslant b - a$. □

Dual to Theorem 6.6 we have the following.

Theorem 6.8. *Given $f : M \to [a, b)$ there exist $a_1 < a_2 < \cdots < a_n < b \leqslant b_n < \cdots < b_2 < b_1$, a module N, and an isomorphism $\theta : N \oplus \bigoplus_{j=1}^{n}[a_j, b_j) \to M$ such that $f\theta i_N = 0$ and $f\theta i_j$ is nonzero for all j. Furthermore* $\operatorname{coker} f \cong [a, a_1)$ *and*

$$\ker f \cong [a_2, b_1) \oplus \cdots \oplus [a_{n-1}, b_n) \oplus [b, b_n).$$

Proof. Let $\theta : N \oplus \bigoplus_{j=1}^{n}[a_j, b_j) \to M$ be an isomorphism (with N possibly zero) such that $f\theta i_N = 0$, for all j, $f\theta i_j$ is nonzero, and n is as small as possible. Without loss of generality, assume that $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$. Since for all j, $f\theta i_j$ is nonzero, $a_j \in [a, b)$.

If $n = 0$ or 1 then we are done. So assume that $n \geqslant 2$. We claim that the hypotheses imply that $a_1 < a_2 < \cdots < a_n$. If not, there is a k such that $a_k = a_{k+1}$. Without loss of generality, assume that $b_k \leqslant b_{k+1}$. Let $\alpha : [a_{k+1}, b_{k+1}) \to [a_k, b_k)$ be the nonzero map obtained from $((f\theta i_k)(a_{k+1}))^{-1} \circ (f\theta i_{k+1})(a_{k+1})$ and Lemma 6.3.

Let $\phi : N \oplus \bigoplus_{j=1}^{n}[a_j, b_j) \to N \oplus \bigoplus_{j=1}^{n}[a_j, b_j)$ be the unique map determined by

$$\begin{aligned} p_N\phi &= p_N & \text{for } j \neq k, p_j\phi = p_j \\ p_k\phi i_{k+1} &= -\alpha & \text{for } j \neq k+1, p_k\phi i_j = \delta_j^k \end{aligned}$$

We claim that ϕ is an isomorphism. Let $M' = [a_k, b_k) \oplus [a_{k+1}, b_{k+1})$ and $M'' = N \oplus \bigoplus_{j=1}^{k-1}[a_j, b_j) \oplus \bigoplus_{j=k+2}^{n}[a_j, b_j)$. By construction, ϕ decomposes to the identity map on M'' and a nonidentity map ϕ' on M'. The map ϕ' is nontrivial for $a_k = a_{k+1} \leqslant a < b_k$. For each such a, one can choose a basis for $M'(a)$ such that $\phi'(a)$ is given by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, which is an isomorphism.

The exact same argument shows that our assumptions imply that for all j, $b_j > b_{j+1}$.

Since $[a,b)\cap[a_1,b_1)=[a_1,b)$ and for all $c\in[a,a_1)$, and all $j=1,\dots,n$, $c\notin[a_j,b_j)$, it follows that $\operatorname{im} f\cong\operatorname{im} f\theta\cong[a_1,b)$ and thus $\operatorname{coker} f\cong\operatorname{coker} f\theta\cong[a,a_1)$.

Choose a coherent basis for $[a,b)$. By Lemma 6.5, for each j, there is a coherent basis for $[a_j,b_j)$ such that $f\theta i_j(c)=1$ whenever $c\in[a,b)\cap[a_j,b_j)$. Let M_j denote $[a_j,b_j)$. It follows that for $j=1,\dots,n-1$, and for $c\in[a_{j+1},b_j)$, $f\theta i_j(c)1_{M_j}=f\theta i_{j+1}(c)1_{M_{j+1}}$.

Let $M'=N\oplus[a_1,b_1)\oplus\cdots\oplus[a_{n-1},b_{n-1})\oplus[a_n,b_n)$. Let $K=N\oplus[a_2,b_1)\oplus\cdots\oplus[a_n,b_{n-1})\oplus[b,b_n)$. For $j=1,\dots,n-1$, denote $[a_{j+1},b_j)$ by L_j. For $j=1,\dots,n-1$ let ι_j denote the inclusion (coprojection) $[a_{j+1},b_j)\to K$, and let ι_n denote the inclusion $[b,b_n)\to K$. Let ι_N denote the inclusion $N\to K$. For $j=1,\dots,n-1$, let β_j denote the inclusion $[a_{j+1},b_j)\hookrightarrow[a_j,b_j)$, and let β_n denote the inclusion $[b,b_n)\hookrightarrow[a_n,b_n)$. For $j=1,\dots,n-1$, let $\gamma_j:L_j\to M_{j+1}$ be the map given by $\gamma_j(c):1_{L_j(c)}\mapsto 1_{M_{j+1}(c)}$ for all $c\in[a_{j+1},b_{j+1})$.

Let ψ_N be the unique map $N\to M'$ determined by $p_N\psi_N$ is the identity on N and $p_j\psi_N=0$ for all j. For $k=1,\dots,n-1$, let ψ_k be the unique map $L_k\to M'$ determined by $p_N\psi_k=0$, $p_k\psi_k=\beta_k$, $p_{k+1}\psi_k=-\gamma_k$, and for all $j\neq k,k+1$, $p_j\psi_k=0$. Let ψ_n be the unique map $L_n\to M'$ determined by $p_N\psi_n=0$, $p_n\psi_n=\beta_n$ and for all $j=1,\dots,n-1$, $p_j\psi_n=0$. Let $\psi:K\to M'$ be the unique map determined by $\psi\iota_N=\psi_N$, and for $j=1,\dots,n$, $\psi\iota_n=\psi_n$.

Then we have a short exact sequence $0\to K\xrightarrow{\psi}M'\xrightarrow{f\theta}[a_1,b)\to 0$. It follows that $\ker f\cong\ker f\theta\cong K$. ☐

Corollary 6.9. *Given a short exact sequence $0\to M\to N\to[a,b)\to 0$, it follows that $W_1(d_w)(M,N)\leqslant b-a$.*

Proof. Applying Theorem 6.8, we can compare the direct sum decompositions of M and N. The result follows. ☐

Combining Corollaries 6.7 and 6.9, we have that $W_1(d_w)\leqslant d_w$, which completes the proof of Theorem 6.1.

Acknowledgments. The first author would like to acknowledge the support of the Army Research Office [Award W911NF1810307] and the Simons Foundation [Grant number 594594].

References

[1] Michael Barot. *Introduction to the representation theory of algebras*. Springer, Cham, 2015.

[2] Andrew J. Blumberg and Michael Lesnick. Universality of the homotopy interleaving distance. 05 2017, 1705.01690.

[3] Peter Bubenik, Vin de Silva, and Jonathan Scott. Metrics for Generalized Persistence Modules. *Found. Comput. Math.*, 15(6):1501–1531, 2015.

[4] Peter Bubenik, Vin de Silva, and Jonathan Scott. Interleaving and gromov-hausdorff distance. 07 2017, 1707.06288.

[5] Peter Bubenik and Jonathan A. Scott. Categorification of persistent homology. *Discrete Comput. Geom.*, 51(3):600–627, 2014.
[6] Peter Bubenik and Tane Vergili. Topological spaces of persistence modules and their properties. *Journal of Applied and Computational Topology*, page 26pp., (accepted), 1802.08117.
[7] Gunnar Carlsson, Gurjeet Singh, and Afra Zomorodian. Computing multidimensional persistence. *J. Comput. Geom.*, 1(1):72–100, 2010.
[8] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete Comput. Geom.*, 42(1):71–93, 2009.
[9] Gunnar Carlsson, Afra Zomorodian, Anne Collins, and Leonidas Guibas. Persistence barcodes for shapes. In *SGP '04: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 124–135, New York, NY, USA, 2004. ACM Press.
[10] 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 25th annual symposium on Computational geometry*, SCG '09, pages 237–246, New York, NY, USA, 2009. ACM.
[11] Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*. SpringerBriefs in Mathematics. Springer, [Cham], 2016.
[12] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, 2007.
[13] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have L_p-stable persistence. *Found. Comput. Math.*, 10(2):127–139, 2010.
[14] Anne Collins, Afra Zomorodian, Gunnar Carlsson, and Leonidas J. Guibas. A barcode shape descriptor for curve point cloud data. *Computers & Graphics*, 28(6):881 – 894, 2004.
[15] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *J. Algebra Appl.*, 14(5):1550066, 8, 2015.
[16] Vin de Silva, Elizabeth Munch, and Amit Patel. Categorified Reeb graphs. *Discrete Comput. Geom.*, 55(4):854–906, 2016.
[17] Vin de Silva, Elizabeth Munch, and Anastasios Stefanou. Theory of interleavings on categories with a flow, 2017, arXiv:1706.04095.
[18] Emerson G. Escolar and Yasuaki Hiraoka. Persistence Modules on Commutative Ladders of Finite Type. *Discrete Comput. Geom.*, 55(1):100–157, 2016.
[19] Peter Gabriel. Unzerlegbare Darstellungen. I. *Manuscripta Math.*, 6:71–103; correction, ibid. 6 (1972), 309, 1972.
[20] Heather A. Harrington, Nina Otter, Hal Schenck, and Ulrike Tillmann. Stratifying multiparameter persistent homology. 08 2017, 1708.07390.
[21] Henning Krause. Krull-Schmidt categories and projective covers. *Expo. Math.*, 33(4):535–549, 2015.
[22] Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Found. Comput. Math.*, 15(3):613–650, 2015.
[23] Ezra Miller. Data structures for real multiparameter persistence modules. 09 2017, 1709.08155.
[24] Dmitriy Morozov, Kenes Beketayev, and Gunther H. Weber. Interleaving distance between merge trees. In *Proceedings of TopoInVis*, 2013.
[25] Elizabeth Munch and Anastasios Stefanou. The ℓ^∞-cophenetic metric for phylogenetic trees as an interleaving distance. 03 2018, 1803.07609.
[26] Joseph J. Rotman. *An introduction to homological algebra*. Universitext. Springer, New York, second edition, 2009.
[27] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *Discrete Comput. Geom.*, 52(1):44–70, 2014.
[28] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.

(Peter Bubenik) Department of Mathematics, University of Florida

E-mail address: peter.bubenik@ufl.edu

(Jonathan Scott) Department of Mathematics, Cleveland State University

E-mail address: j.a.scott3@csuohio.edu

(Donald Stanley) Department of Mathematics, University of Regina

E-mail address: donald.stanley@uregina.ca