



Exact weights, path metrics, and algebraic Wasserstein distances

Peter Bubenik¹ · Jonathan Scott² · Donald Stanley³

Received: 10 July 2020 / Revised: 23 September 2022 / Accepted: 25 September 2022 /
Published online: 19 October 2022

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

We use weights on objects in an abelian category to define what we call a path metric. We introduce three special classes of weight: those compatible with short exact sequences; those induced by their path metric; and those which bound their path metric. We prove that these conditions are in fact equivalent, and call such weights exact. As a special case of a path metric, we obtain a distance for generalized persistence modules whose indexing category is a measure space. We use this distance to define Wasserstein distances, which coincide with the previously defined Wasserstein distances for one-parameter persistence modules. For one-parameter persistence modules, we also describe maps to and from an interval module, and we give a matrix reduction for monomorphisms and epimorphisms.

Keywords Distances for abelian categories · Persistence modules · Wasserstein distance

Mathematics Subject Classification 55N31 · 18E10

✉ Peter Bubenik
peter.bubenik@ufl.edu

Jonathan Scott
j.a.scott3@csuohio.edu

Donald Stanley
donald.stanley@uregina.ca

¹ Department of Mathematics, University of Florida, Gainesville, USA

² Department of Mathematics, Cleveland State University, Cleveland, USA

³ Department of Mathematics, University of Regina, Regina, Canada

1 Introduction

In nice cases, one-parameter persistence modules are isomorphic to a direct sum of interval modules (Crawley-Boevey 2015; Botnan and Crawley-Boevey 2020) and they have a combinatorial description called a persistence diagram (Cohen-Steiner et al. 2007; Patel 2018). Persistence diagrams have a family of L^p distances, for $1 \leq p \leq \infty$, called p -Wasserstein distances (Cohen-Steiner et al. 2010). For $p = \infty$, this distance is also called the bottleneck distance (Cohen-Steiner et al. 2007). These distances have a common generalization with Wasserstein distances for probability measures (Divol and Lacombe 2021; Bubenik and Elchesen 2022). The bottleneck distance for one-parameter persistence modules has an equivalent linear-algebra formulation called interleaving distance (Chazal et al. 2009; Lesnick 2015; Bubenik and Scott 2014; Bauer and Lesnick 2015; Harker et al. 2019) which has been extended to various generalized persistence modules (Morozov et al. 2013; Bubenik et al. 2015; de Silva et al. 2016; de Silva et al. 2018; Blumberg and Lesnick 2017; Bubenik et al. 2017; Munch et al. 2019; Botnan et al. 2020). 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.

We generalize the 1-Wasserstein distance for one-parameter persistence modules to abelian categories. If these abelian categories satisfy some additional standard axioms we also obtain a generalization of the p -Wasserstein distances.

For an abelian category \mathbf{A} , a weight assigns each object $A \in \mathbf{A}$ an associated weight $w(A) \in [0, \infty]$ such that $w(0) = 0$ and if $A \cong B$ then $w(A) = w(B)$. For example, for a field K and the category of K -vector spaces, we have the weight $w(A)$ given by the dimension of A . For another example, for a ring R and the category of left R -modules, we have the weight $w(M) = \text{pd}(M) + 1$, where $\text{pd}(M)$ denotes the projective dimension of M . Say that a weight w is exact (Definition 3.9) if for each short exact sequence $0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$, $w(A) \leq w(B) + w(C)$, $w(B) \leq w(A) + w(C)$, and $w(C) \leq w(A) + w(B)$. Both of the previous two examples of weights are exact.

Given $A, B \in \mathbf{A}$, a zigzag from A to B consists of a sequence of morphisms $\gamma : A = A_0 \xrightarrow{\gamma_1} A_1 \xleftarrow{\gamma_2} A_2 \xrightarrow{\gamma_3} \dots \xleftarrow{\gamma_n} A_n = B$ for some $n \geq 0$. Define the cost of a zigzag by

$$\text{cost}_w(\gamma) = \sum_{i=1}^n (w(\ker \gamma_i) + w(\text{coker } \gamma_i)),$$

and let $d_w(A, B) = \inf_\gamma \text{cost}_w(\gamma)$, where the infimum is taken over all zigzags between A and B (Definition 3.4). We show (Lemma 3.5) that d_w is a metric (Definition 3.3) which we call the path metric.

Given a metric d on \mathbf{A} , there is a weight given by $|d|(A) = d(A, 0)$ (Definition 3.19). Therefore, given a weight w , we obtain a sequence of weights w_1, w_2, w_3, \dots with $w_1 = w$ and $w_{n+1} = |d_{w_n}|$ for $n \geq 1$. We prove that $w_1 \geq w_2 \geq w_3 \geq \dots$ (Lemma 3.20). This sequence stabilizes if there exists an $n \geq 1$ such that $w_{n+1} = w_n$. We call a weight w stable if $|d_w| = w$ (Definition 3.21).

We prove that any weight provides an upper bound for its path metric: $d_w(A, B) \leq w(A) + w(B)$ (Proposition 3.25). We say that a weight *bounds its path metric* if in addition $|w(A) - w(B)| \leq d_w(A, B)$ (Definition 3.26).

We prove that the three seemingly unrelated conditions on weights we have introduced are in fact equivalent.

Theorem 1.1 (Theorem 3.28) *For a weight w the following are equivalent:*

- w is exact;
- w is stable; and
- w bounds its path metric.

We also show (Definition 3.17) that for each weight there is a canonical exact weight and that for each exact weight there is a canonical *amplitude*, a strengthening of our notion of exact weight introduced by Giunti et al. (2021) (Definition 3.12).

A *persistence module* indexed by a small category \mathbf{P} and valued in \mathbf{A} is a functor from \mathbf{P} to \mathbf{A} and a *morphism* of persistence modules is a natural transformation. For example, consider (\mathbb{R}^n, \leq) or (\mathbb{Z}^n, \leq) with the coordinatewise partial order, viewed as a category. The category of such persistence modules and their morphisms is an abelian category.

To define a path metric on this category of persistence modules, we use one additional ingredient. We assume that the underlying set P of the small category \mathbf{P} has a measure μ . For example, consider \mathbb{R}^n with the Lebesgue measure or \mathbb{Z}^n with the counting measure. Then a persistence module M has an associated weight defined by $W(M) = \mu(w(M)) = \int_P w(M) d\mu$ (Definition 4.1). If w is exact or an amplitude then so is W (Lemmas 4.2 and 4.3). Using this weight we obtain the path metric $d_W = d_{\mu \circ w}$.

We prove that exact weights may be used to bound the path metric for persistence modules.

Theorem 1.2 (Proposition 4.6 and Theorem 4.7) *If w is an exact weight and $w(M)$ and $w(N)$ are μ -integrable then*

$$\int_P |w(M) - w(N)| d\mu \leq d_W(M, N) \leq \int_P (w(M) + w(N)) d\mu.$$

Now assume that the persistence modules have values in a Grothendieck category (Sect. 2.3) such as \mathbf{Vect}_K , the category of vector spaces over a field K , or \mathbf{Mod}_R , the category of left R -modules for some ring R , and that they have a decomposition into a direct sum of persistence modules with local endomorphism rings (Sect. 2.4). Given a metric d and $p \in [1, \infty]$, we define the associated *p-Wasserstein distance* (Definition 5.1),

$$W_p(d)(M, N) = \inf \|\{d(M_a, N_a)\}_{a \in A}\|_p,$$

where the infimum is taken over all isomorphisms $M \cong \bigoplus_{a \in A} M_a$ and $N \cong \bigoplus_{a \in A} N_a$, where each M_a and N_a is either 0 or has a local endomorphism ring and is thus indecomposable (Lemma 2.1). We show that $W_p(d)$ is a metric (Proposition 5.4), which has the following universal property.

Theorem 1.3 (Theorem 5.15) *The metric $W_p(d)$ is the largest p -subadditive metric that is bounded above by d on indecomposables.*

For one-parameter persistence modules we prove the following two isometry theorems.

Theorem 1.4 (Theorem 5.9) *For persistence modules indexed by the integers or the real numbers with values in \mathbf{Vect}_K , $W_p(d_W)$ agrees with the p -Wasserstein distance of the corresponding persistence diagrams (Sect. 2.7).*

Theorem 1.5 (Theorem 6.17) *For persistence modules indexed by the integers or the real numbers with values in \mathbf{Vect}_K , $W_1(d_W)$ agrees with the path metric d_W .*

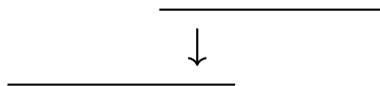
As part of the proof we show that monomorphisms and epimorphisms of one-parameter persistence modules have the following representations which imply that there is an induced matching of interval modules.

Theorem 1.6 (Theorem 6.11) *A monomorphism between persistence modules given by finite direct sums of interval modules can be represented by a matrix in which blocks corresponding to interval modules with the same right end are diagonal.*

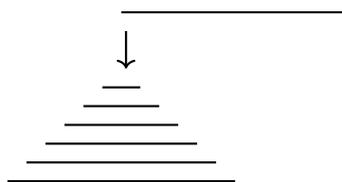
Theorem 1.7 (Theorems 6.13) *An epimorphism between persistence modules given by finite direct sums of interval modules can be represented by a matrix in which blocks corresponding to interval modules with the same left end are diagonal.*

We generalize the following well-known important elementary result for nonzero maps between persistence modules.

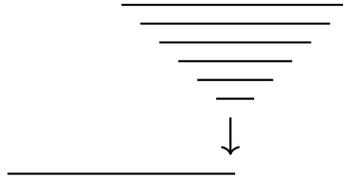
Lemma 1.8 (Lemma 4.9) *Nonzero maps between interval modules may be visualized as follows.*



Theorem 1.9 (Theorem 6.7) *Nonzero maps from an interval module to a finite direct sum of interval modules may be visualized as follows.*



Theorem 1.10 (Theorem 6.9) *Nonzero maps from a finite direct sum of interval modules to an interval module may be visualized as follows.*



Open questions

We have not addressed algorithms for computing our path metric or our algebraic Wasserstein distance, under suitable finiteness conditions (Lesnick et al. 2015; Miller 2019). For example, is there an effective algorithm for computing the distance d_W between two finitely-presented two-parameter persistence modules? Furthermore, for particular applications in which generalized persistence modules arise, one may ask whether or not our distances are stable.

Related work

Patel (2018) defines persistence diagrams for functors on (\mathbb{R}, \leq) (which are obtained from functors on (\mathbb{N}, \leq) by a left Kan extension) to essentially small symmetric monoidal categories with images and more generally to essentially small abelian categories. In the latter case one can apply the tools developed here. Note that our metric d_W is similar in spirit to the construction of the Grothendieck group of an abelian category. Also note that the distances considered in Patel (2018) and the follow-up paper by McCleary and Patel (2019) (interleaving distance, erosion distance, and bottleneck distance) are L^∞ distances. Elchesen and Mémoli (2019) define a distance for zigzag persistence modules (the reflection distance) that is similar to our metric d_W . Related recent papers on the algebra of persistence modules include (Harrington et al. 2019; Bauer et al. 2020; Miller 2019, 2020a, b; Bubenik and Milićević 2021). The first author and Elchesen have also shown a universality result for Wasserstein distance for persistence diagrams (Bubenik and Elchesen 2022).

Skraba and Turner (2020) have independently defined an algebraic p -Wasserstein distance for pointwise-finite-dimensional one-parameter persistence modules and showed that for diagrams with finite total p -persistence it is isometric to the usual p -Wasserstein of the corresponding persistence diagrams.

Scolamiero et al. (2017) define what they call a noise system for tame persistence modules indexed by (\mathbb{Q}^r, \leq) and valued in $\mathbf{Vect}_{\mathbf{K}}$ and use it to define a path metric. Giunti et al. (2021) have defined axioms for a weight on an abelian category which they call an amplitude. Their requirements are closely related to our conditions for an exact weight, but are more restrictive. They observe that noise systems generalize to abelian categories and they prove that an abelian category with an amplitude is equivalent to an abelian category with a noise system. Thus, exact weights may be considered to be generalizations of amplitudes and noise systems. For the path metric on noise systems and amplitudes, it is sufficient to consider zigzags which are cospans (or spans) (Scolamiero et al. 2017; Giunti et al. 2021).

In Sect. 6.3 we show that for monomorphisms and epimorphisms of persistence modules there is an induced algebraic matching of interval modules. Compare this with the induced combinatorial matchings of (Bauer and Lesnick 2015, Theorem 4.2) and the related result by Skraba and Vejdemo (Bauer and Lesnick 2015, Remark 4.4). A closely related result has been proved by Ezra Miller (Miller 2020a, Remark 9.24). Miller's result holds in greater generality, though in our case his result is perhaps slightly weaker or at least less explicit. Our proof is elementary, using a matrix reduction argument.

Outline of the paper

Section 2 consists of background material. In Sect. 3 we define weights and path metrics and study exact weights and their properties. In Sect. 4 we define metrics for generalized persistence modules indexed by a measure space and consider some of their properties. In Sect. 5 we define Wasserstein distances for persistence modules with values in a Grothendieck category, prove that it extends the usual definition, and establish a universal property. In Sect. 6 we show that for one-parameter persistence modules our algebraic 1-Wasserstein distance agrees with the path metric. We also prove structure theorems for maps into and out of an interval module and show that monomorphisms and epimorphisms of persistence modules can be represented by matrices whose form induces a matching of interval modules. Finally, in Sect. 7, we apply our metrics to three examples of two-parameter persistence modules and a pair of zigzag persistence modules.

2 Background

In this section we give background material that will be used later.

2.1 Additive categories

A *zero object* in a category is an object 0 such that for every object X there are unique morphisms $0 \rightarrow X$ and $X \rightarrow 0$. In a category with a zero object, for any two objects A, B there is a unique *zero morphism* given by the composition $A \rightarrow 0 \rightarrow B$. 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*.

Let \mathbf{A} be an additive category. We say that X is the *direct sum* of Y and Z in \mathbf{A} if there are morphisms $i : Y \rightarrow X$, $j : Z \rightarrow X$, $p : X \rightarrow Y$, and $q : X \rightarrow Z$ 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 $qi = 0$ and $pj = 0$, 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 . See Krause (2015) for more details.

In an additive category \mathbf{A} , the *kernel* of a morphism $f : A \rightarrow B$, if it exists, is the equalizer of f and the zero morphism between A and B . Dually, the *cokernel* of f , if it exists, is the coequalizer of f and the zero morphism.

2.2 Abelian categories

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

$$\begin{array}{ccccc} \ker f & \xrightarrow{j} & M & \xrightarrow{f} & N & \xrightarrow{q} & \text{coker } f \\ & & \downarrow & & \uparrow & & \\ & & \text{coker } j & \xrightarrow{\bar{f}} & \ker q & & \end{array}$$

is an isomorphism. Note that $\ker q$ is called the image of f and $\text{coker } j$ is called the coimage of f .

Let R be a commutative ring (with identity). Then the category \mathbf{Mod}_R of R -modules and R -module homomorphisms is an abelian category. As a special case, let K be a field. The category \mathbf{Vect}_K of vector spaces over K and K -linear maps is an abelian category. If \mathbf{A} is an abelian category and \mathbf{D} is a small category then the category $\mathbf{A}^{\mathbf{D}}$, of functors from \mathbf{D} to \mathbf{A} and natural transformations, is an abelian category.

2.3 Grothendieck categories

An *AB5 category* is an abelian category with all coproducts (and hence all colimits) in which filtered colimits of exact sequences are exact. A *Grothendieck category* is an AB5 category which has a generator (i.e. separator).

For example, for any unital ring R , the category \mathbf{Mod}_R of left R -modules and R -module homomorphisms is a Grothendieck category. This includes the cases \mathbf{Vect}_K (where R is a field K) and \mathbf{Ab} the category of abelian groups and group homomorphisms (where $R = \mathbb{Z}$). Let \mathbf{P} be a small category. For any Grothendieck category \mathbf{A} , the category $\mathbf{A}^{\mathbf{P}}$ is a Grothendieck category. In particular, $\mathbf{Vect}_K^{\mathbf{P}}$ is a Grothendieck category.

Let \mathbf{A} be a Grothendieck category. For an arbitrary set A and a collection of objects $\{M_a\}_{a \in A}$ in \mathbf{A} , by definition we have the direct sum (i.e. coproduct) $\bigoplus_{a \in A} M_a$, and canonical maps $i_a : M_a \rightarrow \bigoplus_{a \in A} M_a$ for all $a \in A$. It follows from the Gabriel-Popescu Theorem that \mathbf{A} also has all limits (Stenström 1975, Chapter X), and thus products, in particular. Therefore, we have the product $\prod_{a \in A} M_a$. For $a, b \in A$ define $\tau_{a,b} : M_a \rightarrow M_b$ to be the identity on M_a if $a = b$ and to be the zero map otherwise. For $b \in A$ the maps $\tau_{a,b}$ induce a canonical projection map $p_b : \bigoplus_{a \in A} M_a \rightarrow M_b$. These maps induce a canonical map $\bigoplus_{a \in A} M_a \rightarrow \prod_{a \in A} M_a$.

2.4 Krull-Remak-Schmidt-Azumaya Theorem

An element r in a ring R is a *nonunit* if $Rr \neq R$ and $rR \neq R$. A *local ring* is a ring in which the sum of two nonunits is a nonunit.

Lemma 2.1 *Let \mathbf{A} be an abelian category. If $M \in \mathbf{A}$ has a local endomorphism ring, then M is indecomposable.*

Proof Assume $M \cong M_1 \oplus M_2$, with corresponding maps i_1, p_1, i_2, p_2 . Then $i_1 p_1$ and $i_2 p_2$ are nonunits but their sum is not. □

Theorem 2.2 (Krull–Remak–Schmidt–Azumaya Theorem) (Bucur et al. 1968, Section 6.7), (Pareigis 1970, Section 4.8), (Popescu 1973, Section 5.1). *Let \mathbf{A} be an AB5 category and $M \in \mathbf{A}$. If*

$$M \cong \bigoplus_{i \in I} A_i \cong \bigoplus_{j \in J} B_j,$$

where each A_i and B_j has a local endomorphism ring, then there is a bijection $\varphi : I \rightarrow J$ such that for all $i \in I$, $A_i \cong B_{\varphi(i)}$.

Definition 2.3 For a Grothendieck category \mathbf{A} , let \mathbf{A}_ℓ denote the full additive subcategory of \mathbf{A} whose objects are those objects of \mathbf{A} that are isomorphic to a direct sum of objects with a local endomorphism ring.

2.5 Persistence modules

Let \mathbf{P} be a small category and let \mathbf{A} be an abelian category. Functors $M : \mathbf{P} \rightarrow \mathbf{A}$ are called *persistence modules* indexed by \mathbf{P} with values in \mathbf{A} . Natural transformations of such functors are called *morphisms of persistence modules*. Of particular interest are the cases that \mathbf{A} is \mathbf{Mod}_R or its special case \mathbf{Vect}_K . Let P denote the set of objects of \mathbf{P} . For a persistence module $M : \mathbf{P} \rightarrow \mathbf{Vect}_K$ the *dimension vector* or *Hilbert function* for M is the function $\dim M : P \rightarrow [0, \infty]$ given by $p \mapsto \dim M(p)$.

Among persistence modules with values in \mathbf{Vect}_K , of greatest interest is the case where $P \subseteq \mathbb{R}^d$ for some d and the morphisms are given by the coordinate-wise/product partial order: $(x_1, \dots, x_d) \leq (y_1, \dots, y_d)$ iff $x_i \leq y_i$ for all $1 \leq i \leq d$. When $d \geq 2$ these are called *multi-parameter persistence modules* and when $d = 1$ these are called *one-parameter persistence modules* or just *persistence modules*.

Definition 2.4 Let P be a poset. A subset $C \subseteq P$ is *convex* if for all $p \leq q \leq r$ with $p, r \in C$, we have $q \in C$. A subset $C \subseteq P$ is *connected* if for each $p, q \in C$ there is a sequence $p = p_0, p_1, \dots, p_n = q$ in C such that for each $1 \leq j \leq n$, either $p_{j-1} \leq p_j$ or $p_j \leq p_{j-1}$. An *interval* in P is a convex connected subset. Note that if P is totally ordered then an interval is just a convex subset. Let I be an interval in P . Define a persistence module M indexed by P with values in \mathbf{Vect}_K as follows. For each $p \in P$, let $M(p) = K$ if $p \in I$ and $M(p) = 0$ if $p \notin I$. For $p \leq q$ with $p, q \in I$, let $M(p \leq q)$ be the identity map on K . All other maps $M(p \leq q)$ are

zero, since either the domain or codomain is zero. Call M an *interval module* and it is convenient to abuse notation and denote M by I .

Lemma 2.5 *Each interval module has a local endomorphism ring and is thus indecomposable.*

Proof The endomorphism ring of an interval module, which by definition has values in \mathbf{Vect}_K , is isomorphic to K . □

2.6 p -Norms

It is customary to restrict p -norms to those elements for which they have a finite value; we will not do so. Let $x = \{x_a\}_{a \in A}$, where each $x_a \in [0, \infty]$. Then for $1 \leq p < \infty$, let $\|x\|_p = (\sum_{a \in A} |x_a|^p)^{\frac{1}{p}}$ and $\|x\|_\infty = \sup_{a \in A} |x_a|$.

Lemma 2.6 *Let A and B be disjoint indexing sets. Let $x = \{x_a\}_{a \in A}$, $y = \{x_b\}_{b \in B}$ and $z = \{x_c\}_{c \in A \cup B}$. Then for $1 \leq p \leq \infty$, $\|(\|x\|_p, \|y\|_p)\|_p = \|z\|_p$.*

2.7 Persistence diagrams and their Wasserstein distances

Let $P \subseteq \mathbb{R}$, where \mathbb{R} is given the usual total order. For an interval I in P , let $P_{>I} = \{p \in P \mid \forall x \in I, x < p\}$. For an interval module I indexed by P , let $x(I) = (\inf I, \inf P_{>I}) \in [-\infty, \infty]^2$, where $\inf \emptyset = \infty$. For $x, y \in [-\infty, \infty]^2$, let $d(x, y) = \|x - y\|_1$. Let $\Delta \subset [-\infty, \infty]^2$ denote the diagonal, $\{(x, x) \mid -\infty \leq x \leq \infty\}$ and for $x \in [-\infty, \infty]^2$, let $d(x, \Delta) := \inf_{y \in \Delta} d(x, y)$. By a *matching* between index sets A and B , we mean an injection $\varphi : C \rightarrow B$, where $C \subset A$.

Let $P \subset \mathbb{R}$ and let M be a persistence module indexed by P with values in \mathbf{Vect}_K . Assume that $M \cong \bigoplus_{j \in J} I_j$ where each I_j is an interval module. By Lemma 2.5 and Theorem 2.2, there is a well-defined multiset $\text{Dgm } M := \{x(I_j)\}_{j \in J}$, called the *persistence diagram* of M .

Definition 2.7 Let $1 \leq p \leq \infty$. Let M, N be persistence modules indexed by P with values in \mathbf{Vect}_K that have persistence diagrams $\text{Dgm } M = \{x_a\}_{a \in A}$ and $\text{Dgm } N = \{x'_b\}_{b \in B}$. Define

$$W_p(M, N) = \inf_{\varphi: C \rightarrow B} \left\| \left(\left\| \{d(x_c, x'_{\varphi(c)})\}_{c \in C} \right\|_p, \left\| \{d(x_a, \Delta)\}_{a \in A - C} \right\|_p, \left\| \{d(\Delta, x'_b)\}_{b \in B - \varphi(C)} \right\|_p \right) \right\|_p,$$

where the infimum is over all matchings φ between the index sets A and B . Call this the p -Wasserstein distance between the persistence modules M and N .

We alert the reader that in Cohen-Steiner et al. (2010), the Wasserstein distance uses the ∞ -norm to measure distances in \mathbb{R}^2 . We use the 1-norm.

2.8 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 $M = M_0 \xrightarrow{f_1} M_1 \xleftarrow{f_2} M_2 \xrightarrow{f_3} \dots \xleftarrow{f_n} M_n =$

N . The number $n \geq 0$ is called the *length* of the zigzag. Note that by inserting identity maps, we can allow the morphisms to point in either direction.

2.9 Symmetric Lawvere metric

A *symmetric Lawvere metric* is a class \mathcal{C} together with a function d 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 $M, N, P \in \mathcal{C}$, $d(M, P) \leq d(M, N) + d(N, P)$. 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.

3 Weights and path metrics

In this section we use weights on morphisms in a category or weights on objects in an additive category to define a distance that we call a path metric.

3.1 Weights and metrics on categories

In this section we define weight, give examples of weights, define a metric on a category, and give an elementary property of such a metric.

Definition 3.1 A *weight*, w , on a class \mathcal{A} assigns $w(a) \in [0, \infty]$ to each $a \in \mathcal{A}$. A weight on a category is a weight on the class of all objects of the category.

Example 3.2 For any category we have the *zero weight* that assigns each object the weight 0. For any additive category we have the *one weight* that assigns all nonzero objects weight 1 and the zero object weight 0. For an abelian category let \mathcal{S} be the class of simple objects, whose only subobjects are 0 and themselves, together with 0. Define a weight on \mathcal{S} , called the *simple weight*, by $w(0) = 0$ and $w(S) = 1$ for all other $S \in \mathcal{S}$. For a field K and the category \mathbf{Vect}_K of K -vector spaces, we have a weight given by the dimension of the vector space. Call this the *dimension weight*. More generally, if R is an integral domain, then for the category \mathbf{Mod}_R of R -modules, we have a weight given by the rank of a module M , which equals the dimension of $M \otimes_R K$ where K is the field of fractions of R . Call this the *rank weight*.

Definition 3.3 Let \mathcal{C} be a class of objects in a category \mathbf{C} . 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 a category* \mathbf{C} is a metric on the class of all objects in \mathbf{C} .

Our definition does allow non-isomorphic objects M and N to have $d(M, N) = 0$. Let $M, M', N, N' \in \mathcal{C}$ with $M \cong M'$ and $N \cong N'$. It follows from the triangle inequality that $d(M, N) = d(M', N')$.

3.2 Path metric from a weight

We use a weight on a class of morphisms in a category to define a metric for that category which we will call the path metric. As a special case we use a weight on a class of objects in an additive category to define a metric on that category.

Let \mathbf{C} be a category together with a class, \mathcal{M} , of morphisms in \mathbf{C} and a weight w on \mathcal{M} .

Definition 3.4 Let γ be a zigzag in \mathbf{C} in which each morphism in the zigzag is in \mathcal{M} . Define the *cost* of γ , denoted $\text{cost}_w(\gamma)$, to be the sum of the weights of the morphisms in the zigzag. As a special case, the cost of the zigzag of length 0 is 0. Let $A, B \in \mathbf{C}$. Define the *path distance* by $d_w(A, B) = \inf_{\gamma} \text{cost}_w(\gamma)$, where the infimum is taken over all zigzags from A to B such that each morphism in the zigzag is in \mathcal{M} . If there are no such zigzags then let $d_w(A, B) = \infty$.

Lemma 3.5 *The path distance d_w is a symmetric Lawvere metric on \mathbf{C} (Sect. 2.9). If \mathcal{M} includes all isomorphisms in \mathbf{C} and the weight of each isomorphism is 0, then d_w is a metric on \mathbf{C} (Sect. 3.1), which we call the path metric.*

Proof First, for any object A , $d_w(A, A) = 0$ since there is a zigzag of length 0 from A to A , whose cost, by definition is 0. Next, $d_w(A, C) \leq d_w(A, B) + d_w(B, C)$ since we may concatenate a zigzag from A to B with a zigzag from B to C to obtain a zigzag from A to C whose cost is the sum of the costs of the two zigzags. Furthermore, $d_w(A, B) = d_w(B, A)$ since every zigzag has a reverse zigzag with the same cost.

For the second statement, for isomorphic objects A, B , consider the zigzag of length 1 given by $f : A \xrightarrow{\cong} B$, which has cost 0. Thus $d_w(A, B) = 0$. \square

Assumption 3.6 Let \mathbf{A} be an additive category. Let \mathcal{O} be a class of objects in \mathbf{A} . We will always assume that such a class contains 0 and that if $A \cong B$ and $A \in \mathcal{O}$ then $B \in \mathcal{O}$. Let w be a weight on \mathcal{O} . We will always assume that $w(0) = 0$ and if $A \cong B$ and $A, B \in \mathcal{O}$ then $w(B) = w(A)$.

Definition 3.7 Let \mathcal{O} be a class of objects in an additive category \mathbf{A} and let w be a weight on \mathcal{O} . See Assumption 3.6. Let \mathcal{M} be the class of morphisms in \mathbf{A} whose kernel and cokernel and both are in \mathcal{O} . Define a weight on \mathcal{M} , which we also denote w , by $w(f) = w(\ker f) + w(\text{coker } f)$. Note that it follows that \mathcal{M} contains all isomorphisms and that these have weight 0. Applying Definition 3.4, with zigzags of morphisms whose kernel and cokernel are in \mathcal{O} , we obtain a path distance d_w on \mathbf{A} . By Lemma 3.5, d_w is a metric, which call the path metric.

Since any morphism in an abelian category factors through its image (Sect. 2.2) we have the following.

Lemma 3.8 *Assume that \mathbf{A} is an abelian category. In Definition 3.7, if we restrict \mathcal{M} to morphisms having either zero kernel or zero cokernel then we obtain the same path metric.*

3.3 Exact weights and amplitudes

In this section we consider weights compatible with short exact sequences.

Let \mathbf{A} be an abelian category (or more generally a (Quillen) exact category) together with a class of objects \mathcal{O} containing 0 and a weight w on \mathcal{O} (see Assumption 3.6).

Definition 3.9 Say that the weight w on \mathcal{O} is *exact* if for each short exact sequence $0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$ in \mathbf{A} with $A, B, C \in \mathcal{O}$, $w(A) \leq w(B) + w(C)$, $w(B) \leq w(A) + w(C)$, and $w(C) \leq w(A) + w(B)$.

The following are examples of exact weights.

Example 3.10 Let $\mathbf{vect}_{\mathbf{K}}$ be the category of finite-dimensional vector spaces over K and K -linear maps. For $V \in \mathbf{vect}_{\mathbf{K}}$, let $w(V) = 0$ if $V = 0$, otherwise $w(V) = 1$ if $\dim(V)$ is even and $w(V) = 2$ if $\dim(V)$ is odd. Then w is an exact weight on $\mathbf{vect}_{\mathbf{K}}$.

Example 3.11 Let $\mathbf{Mod}_{\mathbf{R}}$ denote the category of right (or left) R -modules over a ring R . For $A \in \mathbf{Mod}_{\mathbf{R}}$, let $\text{pd}(A)$ denote the projective dimension of A . We claim that $w(A) = \text{pd}(A) + 1$ is an exact weight. Consider a short exact sequence $0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$ in $\mathbf{Mod}_{\mathbf{R}}$. Using the characterization of projective dimension using ext groups and the long exact sequence of ext groups, one obtains $\text{pd}(A) \leq \max(\text{pd}(B), \text{pd}(C))$, $\text{pd}(B) \leq \max(\text{pd}(A), \text{pd}(C))$, and $\text{pd}(C) \leq 1 + \max(\text{pd}(A), \text{pd}(B))$. It follows that w is exact. Similarly, if we replace projective dimension with injective dimension or flat dimension, we also obtain an exact weight.

Giunti et al. (2021) consider a stronger notion of exact weight on an abelian category which they call amplitude. We generalize their definition slightly to weights on \mathcal{O} .

Definition 3.12 Say that the weight α on \mathcal{O} is an *amplitude* if for each short exact sequence $0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$ in \mathbf{A} , if $A, B \in \mathcal{O}$ then $\alpha(A) \leq \alpha(B)$, if $B, C \in \mathcal{O}$ then $\alpha(C) \leq \alpha(B)$, and if $A, B, C \in \mathcal{O}$ then $\alpha(B) \leq \alpha(A) + \alpha(C)$. If, in addition, for each short exact sequence $0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$ with $A, B, C \in \mathcal{O}$, $\alpha(B) = \alpha(A) + \alpha(C)$ then the amplitude is called *additive*.

Example 3.13 The zero weight on an abelian category is an additive amplitude. The one weight on an abelian category is a non-additive amplitude. Since any short exact sequence of vector spaces splits, the dimension weight is an additive amplitude. Since localization is an exact functor, the rank weight is also an additive amplitude. For many other examples of amplitude, see Giunti et al. (2021). The exact weights in Examples 3.10 and 3.11 are not amplitudes. The simple weight extends to an additive amplitude on the class of semisimple objects. In the case of $\mathbf{Vect}_{\mathbf{K}}$ this produces the dimension weight.

Example 3.14 For the one weight w on an abelian category \mathbf{A} , the path metric d_w satisfies the following. For $A, B \in \mathbf{A}$, $d_w(A, B) = 0$ iff $A \cong B$, $d_w(A, B) = 1$ iff $A \not\cong B$ and there exists either an injection or a surjection between A and B , and otherwise $d_w(A, B) = 2$.

3.4 How to obtain weights with stronger properties

We will show that each weight has a canonical associated exact weight and the each exact weight has a canonical associated amplitude. Let \mathbf{A} be an abelian category with a class of objects \mathcal{O} including 0. First, we need the following lemma.

Lemma 3.15 *Let $\{\alpha_j\}_{j \in J}$ be a set of amplitudes on \mathcal{O} . For $A \in \mathcal{O}$, let $\alpha(A) = \sup_{j \in J} \alpha_j(A)$. Then α is an amplitude on \mathcal{O} .*

Proof To start, observe that $\alpha(0) = 0$. Next consider a short exact sequence $0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$. Assume $A, B, C \in \mathcal{O}$. Let $\varepsilon > 0$. Then by definition there is a $j \in J$ such that $\alpha_j(B) > \alpha(B) - \varepsilon$. It follows by definition and by assumption that $\alpha(A) + \alpha(C) \geq \alpha_j(A) + \alpha_j(C) \geq \alpha_j(B) > \alpha(B) - \varepsilon$. Therefore $\alpha(A) + \alpha(C) \geq \alpha(B)$. A similar argument shows that if $A, B \in \mathcal{O}$ then $\alpha(A) \leq \alpha(B)$ and that if $B, C \in \mathcal{O}$ then $\alpha(C) \leq \alpha(B)$. □

Similarly, we have the following.

Lemma 3.16 *Let $\{\alpha_j\}_{j \in J}$ be a set of exact weights on \mathcal{O} . For $A \in \mathcal{O}$, let $\alpha(A) = \sup_{j \in J} \alpha_j(A)$. Then α is an exact weight on \mathcal{O} .*

For two weights w, w' on \mathcal{O} say that $w \leq w'$ iff for all $A \in \mathcal{O}$, $w(A) \leq w'(A)$.

Definition 3.17 Let w be a weight on \mathcal{O} . We define the associated exact weight \underline{w} on \mathcal{O} to be the supremum of the exact weights upper-bounded by w . Note that this set of exact weights is nonempty because of the zero weight. We define the associated amplitude α_w on \mathcal{O} to be the supremum of the amplitudes upper-bounded by w .

Note that if w is an exact weight then $\underline{w} = w$ and if w is an amplitude, then $\alpha_w = w$.

Example 3.18 For the exact weight w in Example 3.10, the associated amplitude is the one weight (Example 3.2). For the exact weight w in Example 3.11, if \mathbf{A} has enough projectives, then for each $A \in \mathbf{A}$ there is a surjection $P \xrightarrow{f} A$ with P projective. From the short exact sequence $0 \rightarrow \ker f \rightarrow P \rightarrow A \rightarrow 0$, we obtain $\alpha_w(A) \leq \alpha_w(P) \leq w(P) = 1$. Thus the associated amplitude is also the one weight.

3.5 Weight from a metric and stable weights

In this section, we use a metric on a category to define a weight on that category. Let \mathbf{A} be a category together with a class of objects \mathcal{O} in \mathbf{A} (see Assumption 3.6).

Definition 3.19 Let d be a metric on \mathbf{A} (Sect. 3.1). For $A \in \mathbf{A}$ define $|d|(A) = d(A, 0)$. Then $|d|(0) = 0$ and if $A \cong B$ then $|d|(A) = |d|(B)$. Let $|d|_{\mathcal{O}}$ denote the restriction of $|d|$ to \mathcal{O} . Then $|d|_{\mathcal{O}}$ is a weight on \mathcal{O} (Assumption 3.6).

From a weight we obtain a path metric and from this path metric we obtain a weight.

Lemma 3.20 *For a weight w on \mathcal{O} , $|d_w|_{\mathcal{O}} \leq w$.*

Proof For $A \in \mathcal{O}$, the zigzag $A \rightarrow 0$ shows that $d_w(A, 0) \leq w(A)$. □

Thus, for a weight w on \mathcal{O} , we obtain a sequence of decreasing weights on \mathcal{O} , $w = w_1 \geq w_2 \geq w_3 \geq \dots$, with $w_{n+1} = |d_{w_n}|_{\mathcal{O}}$. Say that this sequence *stabilizes* if $w_{n+1} = w_n$ for some n .

Definition 3.21 Say the weight w on \mathcal{O} is *stable* if $|d_w|_{\mathcal{O}} = w$.

Lemma 3.22 For a metric d on \mathbf{A} , $|d|$ need not be an amplitude.

Proof Consider Example 3.10, where the weight w is not an amplitude and $|d_w| = w$. □

Lemma 3.23 For a weight w on \mathbf{A} , we may have $|d_w| \neq w$.

Proof Consider the following weight on $\mathbf{vect}_{\mathbb{K}}$. Let $w(0) = 0$, $w(V) = 1$ if $\dim(V) = 1$ and $w(V) = 3$ otherwise. Then $|d_w|(V) = 2$ if $\dim V = 2$. □

3.6 Bounds on path metrics

For a weight, we give an upper bound for its path metric. We define weights that give lower bounds for their path metrics. In Sect. 3.7 we will show that exact weights give such lower bounds.

Let \mathbf{A} be an additive category, together with a class, \mathcal{O} , of objects in \mathbf{A} , and a weight w on \mathcal{O} (see Assumption 3.6). There is an induced path metric d_w on \mathbf{A} (Definition 3.7).

Lemma 3.24 For all $A, B \in \mathbf{A}$, $d_w(A, B) \leq |d_w|(A) + |d_w|(B)$.

Proof By the triangle inequality, $d_w(A, B) \leq d_w(A, 0) + d_w(0, B) = |d_w|(A) + |d_w|(B)$. □

Combining Lemmas 3.24 and 3.20 we have the following.

Proposition 3.25 For all $A, B \in \mathcal{O}$, $d_w(A, B) \leq w(A) + w(B)$.

Definition 3.26 Say that the weight w *lower bounds its path metric* if for all $A, B \in \mathcal{O}$, $|w(A) - w(B)| \leq d_w(A, B)$.

3.7 Equivalent conditions on a weight

We conclude this section by showing that the three conditions on a weight that we have introduced are equivalent.

Theorem 3.27 Let \mathbf{A} be an additive category, together with a class of objects \mathcal{O} in \mathbf{A} and a weight w on \mathcal{O} (see Assumption 3.6). The weight w is *stable* (Definition 3.21) if and only if it *lower bounds its path metric* (Definition 3.26).

Proof Assume that w is stable. By the triangle inequality, for all $A, B \in \mathcal{O}$, $|d_w(A, 0) - d_w(B, 0)| \leq d_w(A, B)$. Since w is stable, we obtain $|w(A) - w(B)| \leq d_w(A, B)$.

Assume that w lower bounds its path metric. For all $A \in \mathcal{O}$, $w(A) = |w(A) - 0| = |w(A) - w(0)| \leq d_w(A, 0)$. By Lemma 3.20, $d_w(A, 0) \leq w(A)$. Thus, $d_w(A, 0) = w(A)$. □

Theorem 3.28 *Let \mathbf{A} be an additive category, together with a class of objects \mathcal{O} in \mathbf{A} and a weight w on \mathcal{O} . Assume that for all short exact sequences $0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$ in \mathbf{A} in which two of A, B, C are in \mathcal{O} then so is the third. The following three conditions on w are equivalent:*

- (1) w is exact (Definition 3.9);
- (2) w is stable; and
- (3) w lower bounds its path metric.

Proof We will show (1) iff (2). The remainder of the statement follows from Theorem 3.27.

First we show that (2) implies (1). Consider a short exact sequence $0 \rightarrow A \xrightarrow{f} B \xrightarrow{g} C \rightarrow 0$ with $A, B, C \in \mathcal{O}$. Then by assumption and the triangle inequality $w(A) = d_w(A, 0) \leq d_w(A, B) + d_w(B, 0)$. By assumption $d_w(B, 0) = w(B)$ and from the zigzag $A \xrightarrow{f} B$ we have that $d_w(A, B) \leq w(C)$. Thus $w(A) \leq w(C) + w(B)$. Similarly $w(B) = d_w(B, 0) \leq d_w(B, A) + d_w(A, 0) \leq w(C) + w(A)$ and $w(C) = d_w(C, 0) \leq d_w(C, B) + d_w(B, 0) \leq w(A) + w(B)$.

It remains to show that (1) implies (2). By Lemma 3.20, $|d_w|_{\mathcal{O}} \leq w$. We will obtain a contradiction to $|d_w|_{\mathcal{O}} < w$. Assume $|d_w|_{\mathcal{O}} < w$. By Definition 3.7 and Lemma 3.8, there is a zigzag γ consisting of morphisms with either zero kernel and cokernel in \mathcal{O} or zero cokernel and kernel in \mathcal{O} from some object $A \in \mathcal{O}$ to 0 such that $\text{cost}_w(\gamma) < w(A)$. The length of any such zigzag is a nonnegative integer. Take γ to be such a zigzag of minimal length. Let f be the first morphism of this zigzag, which is either of the form $A \xrightarrow{f} B$ or $A \xleftarrow{f} B$. Since f has either zero kernel and cokernel in \mathcal{O} or zero cokernel and kernel in \mathcal{O} , by our assumption on \mathcal{O} , $B \in \mathcal{O}$. Let γ' denote the remainder of the zigzag γ without the morphism f . Then γ' is a zigzag from B to 0 consisting of morphisms with either zero kernel and cokernel in \mathcal{O} or zero cokernel and kernel in \mathcal{O} . Since the length of γ' is less than the length of γ , by the minimality of γ , $\text{cost}_w(\gamma') = w(B)$. There are four cases to consider, depending on the direction of f and whether or not f has zero kernel or zero cokernel. For example, if $A \xleftarrow{f} B$ and f has zero cokernel, then we have the short exact sequence $0 \rightarrow \ker(f) \rightarrow B \rightarrow A \rightarrow 0$. Since w is exact, $w(A) \leq w(B) + w(\ker(f)) = \text{cost}_w(\gamma') + w(\ker(f)) = \text{cost}_w(\gamma)$, which is a contradiction. In the other cases, we also have a short exact sequence containing A, B , and either $\ker f$ or $\text{coker } f$. The same argument again gives us a contradiction. \square

4 Path metrics for persistence modules

In this section, we specialize the results of Sect. 3 to the case of persistence modules indexed by a small category whose set of objects comes equipped with a measure.

4.1 Indexing categories with measures

In Sect. 4.2, we will show that for an indexing category \mathbf{P} with a measure on its set of objects and a weight on an abelian category \mathbf{A} there is an induced weight on the category of persistence modules indexed by \mathbf{P} with values in \mathbf{A} .

Let \mathbf{P} be a small category whose set of objects P has a σ -algebra Ω and measure μ . The classical case of persistence modules is given by $P \subseteq \mathbb{Z}$ or $P \subseteq \mathbb{R}$ (assumed to be measurable) with morphisms \leq and the counting measure or the Lebesgue measure, respectively. The case of multi-parameter persistence modules is given by $P \subseteq \mathbb{Z}^d$ or $P \subseteq \mathbb{R}^d$ (assumed to be measurable) with the coordinate-wise/product partial order \leq and the counting measure or the Lebesgue measure, respectively.

4.2 Weights and path metrics for persistence modules

We now define an induced weight for persistence modules. Let \mathbf{P} be a small category whose set of objects P has a measure μ . Let \mathbf{A} be an abelian category together with a class of objects \mathcal{O} in \mathbf{A} and a weight w on \mathcal{O} (see Assumption 3.6).

Assume that we have a persistence module $M : \mathbf{P} \rightarrow \mathbf{A}$ such that for each $p \in P$, $M(p) \in \mathcal{O}$. Then we have a function $w(M) : P \rightarrow [0, \infty]$ given by $p \mapsto w(M(p))$. For example, if M is a persistence module with values in $\mathbf{Vect}_{\mathbb{K}}$ then $\dim(M)$ is the Hilbert function of M . If $w(M)$ is μ -integrable then we write $\mu(w(M))$ to denote the integral $\int_P w(M) d\mu$, which is also written as $\int_P w(M(p)) d\mu(p)$.

Definition 4.1 Consider the category of persistence modules indexed by \mathbf{P} with values in \mathbf{A} . Let $\mathcal{T}_{\mathcal{O},\mu,w}$ be the class of persistence modules M such that for all $p \in P$, $M(p) \in \mathcal{O}$ and such that $w(M)$ is μ -integrable. Then $(\mu \circ w)(M) = \mu(w(M))$ defines a weight $\mu \circ w$ on $\mathcal{T}_{\mathcal{O},\mu,w}$.

Lemma 4.2 *If w is an exact weight on \mathcal{O} then $\mu \circ w$ is an exact weight on $\mathcal{T}_{\mathcal{O},\mu,w}$.*

Proof Let 0 be the zero persistence module. Then $(\mu \circ w)(0) = \mu(w(0)) = \int_P 0 d\mu = 0$. Also, if $M \cong N$ then for all $p \in P$, $M(p) \cong N(p)$, so $w(M(p)) = w(N(p))$, and hence $(\mu \circ w)(M) = (\mu \circ w)(N)$.

Let $0 \rightarrow M \rightarrow N \rightarrow Q \rightarrow 0$ be a short exact sequence of persistence modules. Then for all $p \in P$, $0 \rightarrow M(p) \rightarrow N(p) \rightarrow Q(p) \rightarrow 0$ is a short exact sequence in \mathbf{A} . If $M, N, Q \in \mathcal{T}_{\mathcal{O},\mu,w}$ then for all $p \in P$, $M(p), N(p), Q(p) \in \mathcal{O}$. Since w is an exact weight on \mathcal{O} , $w(M(p)) \leq w(N(p)) + w(Q(p))$. Thus $\int_P w(M(p)) d\mu(p) \leq \int_P w(N(p)) d\mu(p) + \int_P w(Q(p)) d\mu(p)$. That is, $(\mu \circ w)(M) \leq (\mu \circ w)(N) + (\mu \circ w)(Q)$. The other cases are similar. \square

Similarly, we have the following.

Lemma 4.3 *If α is an amplitude on \mathcal{O} , then $\mu \circ \alpha$ is an amplitude on $\mathcal{T}_{\mathcal{O},\mu,w}$.*

Definition 4.4 Combining Definitions 3.7 and 4.1, we have a path metric $d_{\mu \circ w}$ on persistence modules indexed by \mathbf{P} with values in \mathbf{A} .

Lemma 4.5 *Let M, N be persistence modules indexed by \mathbf{P} with values in \mathbf{A} and let γ be a zigzag in $\mathcal{T}_{\mathcal{O},\mu,w}$ from M to N . Then $\text{cost}_{\mu \circ w}(\gamma) = \mu(\text{cost}_w(\gamma))$.*

Proof Consider a zigzag γ in $\mathcal{T}_{\mathcal{O},\mu,w}$ given by $M \xrightarrow{f_1} M_1 \xleftarrow{f_2} M_2 \xrightarrow{f_3} \dots \xleftarrow{f_n} N$. Then $\text{cost}_{\mu \circ w}(\gamma) = \sum_{j=1}^n ((\mu \circ w)(\ker f_j) + (\mu \circ w)(\text{coker } f_j)) = \sum_{j=1}^n (\int_P w(\ker f_j) d\mu + \int_P w(\text{coker } f_j) d\mu) = \int_P \sum_{j=1}^n (w(\ker f_j) + w(\text{coker } f_j)) d\mu = \int_P \text{cost}_w(\gamma) d\mu$. \square

4.3 Bounds for the path metric on persistence modules

We now provide an upper bound for the path metric induced by a weight and a lower bound on the path metric induced by an exact weight. Let \mathbf{P} be a small category whose set of objects P has a measure μ . Let \mathbf{A} be an abelian category together with a class of objects \mathcal{O} in \mathbf{A} and a weight w on \mathcal{O} (see Assumption 3.6).

Proposition 4.6 *For persistence modules M and N indexed by \mathbf{P} with values in \mathbf{A} , such that for all $p \in P$, $M(p), N(p) \in \mathcal{O}$, and $w(M)$ and $w(N)$ are μ -integrable, we have*

$$d_{\mu \circ w}(M, N) \leq \mu(w(M) + w(N)) = \int_P (w(M) + w(N)) d\mu.$$

Proof By Definition 4.1 and Proposition 3.25, $d_{\mu \circ w}(M, N) \leq (\mu \circ w)(M) + (\mu \circ w)(N) = \int_P w(M) d\mu + \int_P w(N) d\mu = \int_P (w(M) + w(N)) d\mu = \mu(w(M) + w(N))$. \square

Theorem 4.7 *Assume that for all short exact sequences $0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$ in \mathbf{A} , if two of A, B, C are in \mathcal{O} then so is the third, and that the weight w is exact. For persistence modules M, N indexed by \mathbf{P} with values in \mathbf{A} , such that for all $p \in P$, $M(p), N(p) \in \mathcal{O}$, and $w(M)$ and $w(N)$ are μ -integrable, we have*

$$\mu(|w(M) - w(N)|) = \int_P |w(M) - w(N)| d\mu \leq d_{\mu \circ w}(M, N).$$

Proof Consider a zigzag γ in $\mathcal{T}_{\mathcal{O},\mu,w}$ given by $M = M_0 \xrightarrow{f_1} M_1 \xleftarrow{f_2} M_2 \xrightarrow{f_3} \dots \xleftarrow{f_n} M_n = N$ such that each f_j has either zero kernel or zero cokernel. Then for all $p \in P$, $\gamma(p)$ is a zigzag in \mathcal{O} from $M(p)$ to $N(p)$. By Lemma 4.5, $\text{cost}_{\mu \circ w}(\gamma) = \mu(\text{cost}_w(\gamma))$. For each $p \in P$, by Definition 3.7 and Theorem 3.28, $\text{cost}_w(\gamma(p)) \geq d_w(M(p), N(p)) \geq |w(M(p)) - w(N(p))|$. Therefore $\text{cost}_{\mu \circ w}(\gamma) \geq \mu(|w(M) - w(N)|) = \int_P |w(M(p)) - w(N(p))| d\mu(p)$. Hence, by Lemma 3.8, $d_{\mu \circ w}(M, N) \geq \mu(|w(M) - w(N)|)$. \square

For example, for persistence modules M and N indexed by (P, μ) with values in \mathbf{Vect}_K such that $\dim M$ and $\dim N$ are μ -integrable, we have

$$\int_P |\dim M - \dim N| d\mu \leq d_{\mu \circ \dim}(M, N) \leq \int_P (\dim M + \dim N) d\mu. \tag{4.8}$$

4.4 Distance between interval modules

In this section we compute the path distance between interval modules indexed by a totally ordered set. Our interval modules are persistence modules indexed by (P, μ) , where P is a totally ordered set, and valued in \mathbf{Vect}_K .

It is a good exercise to check the following two lemmas (or see (Bubenik and Vergili 2018, Appendix A)).

Lemma 4.9 *Let I and J be interval modules. Then there is a nonzero map $f : I \rightarrow J$ if and only if the intervals intersect and for each $a \in I$ there exists $b \in J$ with $b \leq a$ and for each $b \in J$ there is an $a \in I$ with $b \leq a$.*

Lemma 4.10 *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 $\text{coker}(g) = J \setminus (I \cap J)$. (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 $\text{coker}(g)$.*

Proposition 4.11 *Let I, J be interval modules or the zero module, which we also denote by the empty set. Then $d_{\mu\text{odim}}(I, J) = \mu(I \Delta J)$, where $I \Delta J$ denotes the symmetric difference $(I \cup J) \setminus (I \cap J)$.*

Proof (\leq) If either I or J are zero, then we have a canonical zigzag $I \rightarrow 0$ or $0 \rightarrow J$. By Lemma 4.10 we have one of two canonical zigzags from I to J . In each of these cases the cost of this zigzag is $\mu(I \Delta J)$.

(\geq) By (4.8) $d_{\mu\text{odim}}(I, J) \geq \int |\dim I - \dim J| d\mu = \mu(I \Delta J)$. □

5 Wasserstein distances for Grothendieck categories

In this section we define p -Wasserstein distances for a Grothendieck category and show that it generalizes the usual definition. We also show that it satisfies a universal property.

5.1 The p -Wasserstein distance

Let \mathbf{A} be a Grothendieck category with a metric d (Sect. 3.1). Recall (Definition 2.3) that \mathbf{A}_\bullet is the full subcategory of objects isomorphic to direct sums of objects with local endomorphism rings. For $1 \leq p \leq \infty$, define the p -Wasserstein distance as follows.

Definition 5.1 Let $M, N \in \mathbf{A}_\ell$. Define

$$W_p(d)(M, N) = \inf \|\{d(M_a, N_a)\}_{a \in A}\|_p, \tag{5.2}$$

where the infimum is taken over all isomorphisms $M \cong \bigoplus_{a \in A} M_a$ and $N \cong \bigoplus_{a \in A} N_a$, where each M_a and N_a is either 0 or has a local endomorphism ring (and is thus indecomposable).

Lemma 5.3 *Let $M, N \in \mathbf{A}_\ell$. Assume $M \cong \bigoplus_{a \in A} M_a$ and $N \cong \bigoplus_{b \in B} N_b$, where each M_a and N_b has a local endomorphism ring. Then*

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

where the infimum is over all matchings: $C \subset A$ and $\varphi : C \rightarrow B$ is injective.

Proof By Theorem 2.2, the decompositions of M and N are unique up to isomorphism and reordering. Note that the direct sum in Definition 5.1 also allows zero objects. So the infimum in (5.2) is over all matchings of A and B , where the unmatched terms are matched with the zero object. \square

Proposition 5.4 $W_p(d)$ is a metric (Sect. 3.1) on \mathbf{A}_ℓ .

Proof By assumption, if $M \cong N$ then $d(M, N) = 0$. It follows that if $M \cong N$ then $W_p(d)(M, N) = 0$. Since d is symmetric, it follows that $W_p(d)$ is symmetric.

The proof of the triangle inequality uses Theorem 2.2. Let $M, N, P \in \mathbf{A}_\ell$. Let $\varepsilon > 0$. By including sufficiently many zero modules and relabeling, we may assume that $M \cong \bigoplus_{a \in A} M_a$, $N \cong \bigoplus_{a \in A} N_a$, $P \cong \bigoplus_{a \in A} P_a$, and that $W_p(d)(M, N) \geq \| \{d(M_a, N_a)\}_{a \in A} \|_p - \varepsilon$ and $W_p(d)(N, P) \geq \| \{d(N_a, P_a)\}_{a \in A} \|_p - \varepsilon$. Then

$$\begin{aligned} W_p(d)(M, P) &\leq \| \{d(M_k, P_k)\}_k \|_p \leq \| \{d(M_k, N_k) + d(N_k, P_k)\}_k \|_p \\ &\leq \| \{d(M_k, N_k)\}_k \|_p + \| \{d(N_k, P_k)\}_k \|_p \leq W_p(d)(M, N) + W_p(d)(N, P) + 2\varepsilon, \end{aligned}$$

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. The triangle inequality follows. \square

For example, if we have a measure space (P, μ) and a small category \mathbf{P} with set of objects P , we have the Grothendieck category $\mathbf{Vect}_{\mathbf{K}}^{\mathbf{P}}$ and metric $W_p(d_{\mu \circ \dim})$ on the subcategory $\mathbf{Vect}_{\mathbf{K}}^{\mathbf{P}}$ whose objects are isomorphic to direct sums of objects with local endomorphism rings.

5.2 The W_p Isometry Theorem

In this section we show that in the case of persistence modules indexed by $P \subseteq \mathbb{R}$ our definition of p -Wasserstein distance (Definition 5.1) agrees with the definition using persistence diagrams (Definition 2.7). Consider \mathbb{R} with the usual total order and let $P \subseteq \mathbb{R}$. For an interval I in P , let $P_{>I} = \{p \in P \mid \forall x \in I, x < p\}$. Let μ be a measure on P such that for all intervals I in P , $\mu(I) = \inf P_{>I} - \inf I$, where $\inf \emptyset = \infty$. For

example, we may take $P = \mathbb{R}$ or $P = [0, \infty)$ with the Lebesgue measure, or $P = \mathbb{Z}$ or $P = \mathbb{N}$ with the counting measure.

Recall (Sect. 2.7) that for an interval module I , $x(I) = (\inf I, \inf P_{>I})$ and that Δ denotes the diagonal in $[-\infty, \infty]^2$. Also, for $x, y \in [-\infty, \infty]^2$, $d(x, y) = \|x - y\|_1$.

Lemma 5.5 *Let I be an interval in P . Then $d(x(I), \Delta) = \mu(I)$.*

Proof $d(x(I), \Delta) = d((\inf I, \inf P_{>I}), \Delta) = \inf P_{>I} - \inf I = \mu(I)$. □

Lemma 5.6 *If I, J are intervals in P with $I \cap J \neq \emptyset$, then $d(x(I), x(J)) = \mu(I \Delta J)$, where $I \Delta J$ denotes the symmetric difference $(I \cup J) \setminus (I \cap J)$.*

Proof There are a number of cases to consider. However, in each case, $\mu(I \Delta J) = |\inf I - \inf J| + |\inf P_{>I} - \inf P_{>J}| = \|x(I) - x(J)\|_1 = d(x(I), x(J))$. □

Lemma 5.7 *If I and J are intervals in P with $I \cap J = \emptyset$, then $d(x(I), x(J)) \geq \mu(I) + \mu(J)$.*

Proof Without loss of generality, assume that $\inf I \leq \inf P_{>I} \leq \inf J \leq \inf P_{>J}$. Then $d(x(I), x(J)) = \inf P_{>J} - \inf P_{>I} + \inf J - \inf I \geq \inf P_{>J} - \inf J + \inf P_{>I} - \inf I = \mu(I) + \mu(J)$. □

Proposition 5.8 *Using Definition 2.7, for intervals I and J in P , $W_1(I, J) = \mu(I \Delta J)$.*

Proof There are only two 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 2.7 and Lemma 5.5,

$$\begin{aligned} W_1(I, J) &= \min(d(x(I), x(J)), d(x(I), \Delta) + d(\Delta, x(J))) \\ &= \min(d(x(I), x(J)), \mu(I) + \mu(J)). \end{aligned}$$

If $I \cap J \neq \emptyset$, then by Lemma 5.6, $d(x(I), x(J)) = \mu(I \Delta J) \leq \mu(I) + \mu(J)$, so $W_1(I, J) = \mu(I \Delta J)$. If $I \cap J = \emptyset$, then by Lemma 5.7 it follows that $W_1(I, J) = \mu(I) + \mu(J) = \mu(I \Delta J)$. □

Theorem 5.9 (W_p Isometry Theorem) *Let $P \subseteq \mathbb{R}$ with measure μ such that for each interval I in P , $\mu(I) = \inf P_{>I} - \inf I$. If $M, N \in \mathbf{Vect}_{\mathbf{K}}^P$ have a persistence diagram, then for $1 \leq p \leq \infty$,*

$$W_p(d_{\mu \circ \dim})(M, N) = \inf \|\{\mu(M_a \Delta N_a)\}_{a \in A}\|_p = W_p(M, N),$$

where on the left we are using Definition 5.1 and on the right we are using Definition 2.7, and the infimum is taken over all isomorphisms $M \cong \bigoplus_{a \in A} M_a$ and $N \cong \bigoplus_{a \in A} N_a$ where every M_a and N_a is either an interval module or is zero, which corresponds to the empty set.

Proof The first equality follows from Definition 5.1 and Proposition 4.11.

Assume $M \cong \bigoplus_{a \in A} I_a$ and $N \cong \bigoplus_{b \in B} I'_b$, where each I_a and I'_b is an interval module. By Definition 2.7 and Lemma 5.5,

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

where the infimum is over all matchings φ between A and B . By Lemma 5.7, this equals the infimum taken over matchings $\varphi : C \rightarrow B$ with the property that $I_c \cap I'_{\varphi(c)} \neq \emptyset$ for all $c \in C$ (where it could be that $C = \emptyset$). Thus, by Lemma 5.6,

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

Writing this more compactly we obtain the second equality. □

5.3 The universal property of $W_p(d)$

In this section we show that $W_p(d)$ may be characterized as the largest p -subadditive metric that is bounded by d on those objects with local endomorphism rings. Let \mathbf{A} be a Grothendieck category with metric d (Sect. 3.1). Let $1 \leq p \leq \infty$.

Definition 5.10 For $A, B \in \mathbf{A}$, let $d_p(A, B) = \min(d(A, B), \|(d(A, 0), d(0, B))\|_p)$.

One may check that d_p is a metric on \mathbf{A} [see Bubenik and Elchesen (2022, Lemma 3.13)].

Lemma 5.11 *Restricted to objects with local endomorphism rings and zero, $W_p(d)$ equals d_p .*

Proof Consider M, N with local endomorphism rings or being zero. By Definitions 5.1 and 5.10, $W_p(d)(M, N) = \min(d(M, N), \|(d(M, 0), d(0, N))\|_p) = d_p(M, N)$. □

Definition 5.12 Say that a metric d on \mathbf{A}_ℓ is p -subadditive if for any sets $\{M_a\}_{a \in A}$ and $\{N_a\}_{a \in A}$ of objects in \mathbf{A}_ℓ , $d(\bigoplus_{a \in A} M_a, \bigoplus_{a \in A} N_a) \leq \|\{d(M_a, N_a)\}_{a \in A}\|_p$.

Proposition 5.13 $W_p(d)$ is a p -subadditive metric on \mathbf{A}_ℓ .

Proof Consider $\bigoplus_{a \in A} M_a$ and $\bigoplus_{a \in A} N_a$, where $M_a, N_a \in \mathbf{A}_\ell$ for all $a \in A$.

For the left hand side, $W_p(d)(\bigoplus_{a \in A} M_a, \bigoplus_{a \in A} N_a) = \inf \|\{d(P_s, Q_s)\}_{s \in S}\|_p$, where $\bigoplus_{a \in A} M_a \cong \bigoplus_{s \in S} P_s$ and $\bigoplus_{a \in A} N_a \cong \bigoplus_{s \in S} Q_s$ with each P_s and Q_s either having a local endomorphism ring or being zero. For the right hand side, $\|\{W_p(d)(M_a, N_a)\}_{a \in A}\|_p = \inf \|\{d(P_{a,s}, Q_{a,s})\}_{a \in A, s \in B_a}\|_p$, where $M_a \cong \bigoplus_{s \in B_a} P_{a,s}$ and $N_a \cong \bigoplus_{s \in B_a} Q_{a,s}$ with each $P_{a,s}$ and $Q_{a,s}$ either having a local endomorphism ring or being zero. By Theorem 2.2 each term in the right hand side is a term in the left hand side. The result follows. □

Proposition 5.14 *Let d' be a p -subadditive metric on \mathbf{A}_ℓ that is bounded above by d on objects with local endomorphism rings and zero. Then $d' \leq W_p(d)$.*

Proof Let $M, N \in \mathbf{A}_\ell$. Consider Definition 5.1. For each pair of isomorphisms $M \cong \bigoplus_{a \in A} M_a$ and $N \cong \bigoplus_{a \in A} N_a$ where each M_a or N_a is either 0 or has a local endomorphism ring, since d' is p -subadditive, $d'(M, N) \leq \|\{d'(M_a, N_a)\}_{a \in A}\|_p$, which by assumption is bounded above by $\|\{d(M_a, N_a)\}_{a \in A}\|_p$. Therefore $d'(M, N) \leq W_p(d)(M, N)$. \square

Combining Lemma 5.11 and Propositions 5.13 and 5.14, we have the following.

Theorem 5.15 (Universal characterization of $W_p(d)$) *$W_p(d)$ is the largest p -subadditive metric on \mathbf{A}_ℓ that is bounded above by d on objects with local endomorphism rings and zero.*

6 Algebra and persistence modules

In this section we will prove that $W_1(d_{\mu \circ \dim})$ and $d_{\mu \circ \dim}$ are equal for certain persistence modules. Along the way, we will prove structure theorems for maps from an interval module and maps to an interval module and show that both monomorphisms and epimorphisms of persistence modules induce algebraic matchings of direct summands. Let $P \subseteq \mathbb{R}$. Let μ be a measure on P such that for all intervals I in P , $\mu(I) = \inf_{P > I} - \inf I$, where $P > I = \{p \in P \mid \forall x \in I, x < p\}$ and $\inf \emptyset = \infty$.

Throughout this section (with the exception of Definition 6.1), we will restrict \mathbf{Vect}_K^P to the full subcategory, \mathbf{Vect}_{ds}^P , whose objects are isomorphic to direct sums of interval modules. Recall that $\mu \circ \dim$ is a weight on the persistence modules whose Hilbert functions are integrable. It restricts to a weight on \mathbf{Vect}_{ds}^P . We obtain a corresponding path metric $d_{\mu \circ \dim}$ on \mathbf{Vect}_{ds}^P .

6.1 Change of bases

In this section we give a change-of-basis lemma that is a main technical ingredient in our proof of induced algebraic matchings and hence of our W_1 isometry theorem. To help with the arguments used in that proof, we give two examples that use this lemma.

Definition 6.1 Consider $M \in \mathbf{Vect}_K^P$. For each $a \in P$, let B_a be a basis for $M(a)$. Call $\{B_a\}_{a \in P}$ a *set of coherent bases* for M if for all $a \leq b \in P$, $M(a \leq b)$ restricts to a matching of B_a and B_b . That is, there is a subset $S \subseteq B_a$ such that $M(a \leq b)|_S$ is one-to-one and has its image in B_b and $M(a \leq b)|_{B_a \setminus S} = 0$.

We remark that a set of coherent bases for a persistence module is often visualized as a set of intervals called a *barcode*.

Notation 6.2 Following (Bubenik and Vergili 2018, Definition 9), for intervals $I, J \subseteq P$ or corresponding interval modules say that $I \leq J$ if for all $i \in I$ there exists $j \in J$ such that $i \leq j$ and if for all $j \in J$ there exists $i \in I$ such that $i \leq j$.

Lemma 6.3 (Change of basis lemma) *Let $M = I \oplus J$, where I, J are interval modules, $I \leq J$ and $I \cap J \neq \emptyset$. Let $\{e_c\}_{c \in I}$ and $\{f_c\}_{c \in J}$ denote sets of coherent bases for I and J , respectively. Consider $ke_c + \ell f_c$, where $c \in I \cap J$ and $k, \ell \in K \setminus \{0\}$. Then M has a set of coherent bases given by $\{ke_c\}_{c \in I \setminus J} \cup \{ke_c, ke_c + \ell f_c\}_{c \in I \cap J} \cup \{\ell f_c\}_{c \in J \setminus I}$.*

Proof It remains to show that the maps $M(c \leq d) : M(c) \rightarrow M(d)$ restrict to a matching of bases. If $I \setminus J \neq \emptyset$ then let $x \in I \setminus J$, let $y \in I \cap J$, and if $J \setminus I \neq \emptyset$ then let $z \in J \setminus I$. Then $M(x \leq y)(ke_x) = ke_y$, $M(y \leq z)(ke_y) = 0$, and $M(y \leq z)(ke_y + \ell f_y) = \ell f_z$. □

Example 6.4 Consider $f : M \rightarrow N$, where $N = N_1 \oplus N_2$, M, N_1, N_2 are interval modules, $N_1 \leq N_2 \leq M$, and $M \cap N_1 \neq \emptyset$. Let $\{e_c\}_{c \in M}$, $\{e'_c\}_{c \in N_1}$, and $\{e''_c\}_{c \in N_2}$ be coherent sets of bases for M, N_1, N_2 . Assume that $f(e_c) = ke'_c + \ell e''_c$ for some $c \in M \cap N_1$ where $k, \ell \neq 0$. It follows that $f(e_c) = ke'_c + \ell e''_c$ for all $c \in M \cap N_1$ and that $f(e_c) = \ell e''_c$ for all $c \in N_2 \setminus N_1$.

Apply Lemma 6.3 to write N as the internal direct sum $N_1 \oplus N'_2$, where N'_2 has a set of coherent bases given by $\{ke'_c + \ell e''_c\}_{c \in N_1 \cap N_2} \cup \{\ell e''_c\}_{c \in N_2 \setminus N_1}$. Let p_1, p'_2 denote the canonical maps to the direct summands in $N_1 \oplus N'_2$ and let i_1, i'_2 denote the canonical maps from the direct summands to $N_1 \oplus N'_2$. Then $f = i'_2 p'_2 f$. Since $i_1 p_1 + i'_2 p'_2 = 1_N$ and i_1 is a monomorphism it follows that $p_1 f = 0$.

Example 6.5 Consider $f : M \rightarrow N$ where $M = M_1 \oplus M_2$, M_1, M_2, N are interval modules $N \leq M_1 \leq M_2$ and $N \cap M_2 \neq \emptyset$. Let $\{e_c\}_{c \in N}$, $\{e'_c\}_{c \in M_1}$ and $\{e''_c\}_{c \in M_2}$ be sets of coherent bases for N, M_1, M_2 . Assume that $f(e'_c) = ke_c$ for all $c \in M_1 \cap N$, where $k \neq 0$ and $f(e''_c) = \ell e_c$ for all $c \in M_2 \cap N$, where $\ell \neq 0$.

Apply Lemma 6.3 to write M as the internal direct sum $M_1 \oplus M'_2$, where M'_2 has a set of coherent bases given by $\{e''_c - \ell k^{-1} e'_c\}_{c \in M_1 \cap M_2} \cup \{e''_c\}_{c \in M_2 \setminus M_1}$. Then $f i'_2 = 0$, where $i'_2 : M'_2 \rightarrow M_1 \oplus M'_2$ is the canonical map.

6.2 Structure theorems

In this section we give structure theorems for maps out of and into an interval module.

Notation 6.6 Given two intervals I and J , write $I \Subset J$ if $I \subset J$ and there exist $a, b \in J$ such that for all $i \in I, a < i < b$. We will also denote this by $J \supseteq I$.

Given a persistence module, $M = N \oplus \bigoplus_{j=1}^{\infty} M_j$, or $M = N \oplus \bigoplus_{j=1}^n M_j$, let $i_N : N \rightarrow M, p_N : M \rightarrow N$ denote the canonical maps. Similarly, for all j , let $i_j : M_j \rightarrow M$ and $p_j : M \rightarrow M_j$ denote the canonical maps. Recall Notations 6.2 and 6.6.

Theorem 6.7 (Structure theorem for maps from an interval module) *Let M be a direct sum of interval modules (with arbitrary indexing set) and let I be an interval module. Given a nonzero map $f : I \rightarrow M$, there exists an isomorphism $\theta : M \rightarrow N \oplus N'$ with*

- (1) $N' = \bigoplus_{j=1}^n M_j$ for some $n \geq 1$, or
- (2) $N' = \bigoplus_{j=1}^{\infty} M_j$,

such that $p_N \theta f = 0$ and for all j , M_j is an interval module with $M_j \leq I$, $M_j \cap I \neq \emptyset$, $p_j \theta f$ is nonzero, and the interval $M_j \cap I$ contains the interval $M_{j+1} \cap I$.

In the first case, $M_1 \supseteq M_2 \supseteq \dots \supseteq M_n$ and if I does not have a lower bound then $n = 1$. If $\inf I \in I$ then only the first case can occur. For the second case, $\lim_{j \rightarrow \infty} (\sup M_j) = \inf I$.

In both cases, $\ker f = I \setminus M_1$. In the first case,

$$\text{coker } f = N \oplus (M_n \setminus I) \oplus \bigoplus_{j=1}^{n-1} M_j \setminus ((M_j \setminus M_{j+1}) \cap I).$$

Proof Assume $M = \bigoplus_{\alpha \in A} M_\alpha$ where M_α is an interval module. If $p_\alpha f$ is nonzero for some $\alpha \in A$ then $M_\alpha \cap I \neq \emptyset$ and $M_\alpha \leq I$. Furthermore, there is a set of coherent bases $\{e_c\}_{c \in I}$ for I and a set of coherent bases $\{f_d\}_{d \in M_\alpha}$ for M_α .

For each $c \in I$, let $A_c = \{\alpha \in A \mid p_\alpha f(e_c) \neq 0\}$. By the definition of direct sum, $|A_c| < \infty$. If $c \leq d$ then $A_c \supseteq A_d$. Let $A' = \bigcup_{c \in I} A_c$. Since A' is a directed union of finite sets, A' is countable.

Since for all $\alpha \in A'$, $M_\alpha \leq I$, for each $\alpha, \beta \in A'$, $M_\alpha \cap M_\beta \neq \emptyset$. Order A' so that the right ends of the intervals are decreasing and if the right ends are equal then the left ends are increasing. That is, $\{M_\alpha\}_{\alpha \in A'} = \{M_j\}_{j=1}^\infty$ or $\{M_\alpha\}_{\alpha \in A'} = \{M_j\}_{j=1}^n$ such that for all j , $M_j \cap I \supset M_{j+1} \cap I$ and if $M_j \cap I = M_{j+1} \cap I$ then $M_j \setminus I \subset M_{j+1} \setminus I$. Note that $\ker f = I \setminus M_1$.

For all j either $M_{j+1} \leq M_j$ or $M_{j+1} \Subset M_j$. In the case that $\{M_\alpha\}_{\alpha \in A'} = \{M_j\}_{j=1}^n$, whenever $M_{j+1} \leq M_j$, we can apply Lemma 6.3 as in Example 6.4 so that we may remove M_{j+1} from our list. By induction, we have $M_1 \supseteq M_2 \supseteq \dots \supseteq M_{n'}$.

If I does not have a lower bound then $M_i \leq I$, $M_j \leq I$ and $M_i \Subset M_j$ leads to a contradiction.

For each $c \in I$, by the definition of direct sum, $p_j \theta f(c) \neq 0$ for only finitely many j . It follows that if $\inf I \in I$ then one has the case of only finitely many M_j and that if one has infinitely many M_j then $\lim_{j \rightarrow \infty} (\sup M_j) = \inf I$.

Finally, in the case that $\{M_\alpha\}_{\alpha \in A'} = \{M_j\}_{j=1}^n$, I has a set of coherent bases $\{e_c\}_{c \in I}$ and each M_j has a set of coherent bases $\{e_{j,c}\}_{c \in M_j}$ such that for $c \in (M_j \cap I) \setminus (M_{j+1} \cap I)$, $\theta f(e_c) = e_{1,c} + \dots + e_{j,c}$. It follows that $\text{coker } f$ is as claimed. \square

Corollary 6.8 Given a short exact sequence $0 \rightarrow I \rightarrow M \rightarrow N \rightarrow 0$ with I an interval module and M a finite direct sum of interval modules, it follows that

$$W_1(d_{\mu \text{odim}})(M, N) \leq \mu(I).$$

Proof Let f denote the given map $I \rightarrow M$. Apply Theorem 6.7 with $\ker f = 0$. We have $M \cong N' \oplus \bigoplus_{j=1}^n M_j$, where each M_j is an interval module with $M_j \leq I$, and for all j , $M_j \cap I \neq \emptyset$ and $M_j \cap I \supset M_{j+1} \cap I$. Furthermore,

$$N \cong N' \oplus (M_n \setminus I) \oplus \bigoplus_{j=1}^{n-1} M_j \setminus ((M_j \setminus M_{j+1}) \cap I).$$

It follows that

$$W_1(d_{\mu\text{odim}})(M, N) \leq \mu(M_n \cap I) + \sum_{j=1}^{n-1} \mu((M_j \setminus M_{j+1}) \cap I) = \mu(M_1 \cap I) = \mu(I).$$

□

In the dual case we have the following.

Theorem 6.9 (Structure theorem for maps to an interval module) *Let M be a direct sum of interval modules and let I be an interval module. Given a nonzero map $f : M \rightarrow I$, there exists an isomorphism $\theta : N \oplus \bigoplus_{\alpha \in A} M_\alpha \rightarrow M$ such that $f\theta i_N = 0$ and for all $\alpha \in A$, $I \leq M_\alpha$, $M_\alpha \cap I \neq \emptyset$, and $f\theta i_\alpha$ is nonzero. It follows that $\text{coker } f = I \setminus \bigcup_{\alpha \in A} M_\alpha$.*

If A is finite then $\bigoplus_{\alpha \in A} M_\alpha \cong \bigoplus_{j=1}^n M_j$ for some $n \geq 1$, where $M_1 \ni M_2 \ni \dots \ni M_n$, and if I does not have an upper bound then $n = 1$. Furthermore $\text{coker } f = I \setminus M_1$ and

$$\ker f = N \oplus (M_n \setminus I) \oplus \bigoplus_{j=1}^{n-1} M_j \setminus ((M_j \setminus M_{j+1}) \cap I).$$

Proof Assume $M = \bigoplus_{\alpha \in B} M_\alpha$ where M_α is an interval module. Let $A = \{\alpha \in B \mid f i_\alpha \text{ is nonzero}\}$. Let $N = \bigoplus_{\alpha \in B \setminus A} M_\alpha$. For all $\alpha \in A$, $M_\alpha \cap I \neq \emptyset$ and $I \leq M_\alpha$.

Now assume that A is finite. Order the elements of $\{M_\alpha\}_{\alpha \in A}$ so that their left ends are increasing and if the left ends are equal then the right ends are increasing. That is, for some $n \geq 1$, we have $\{M_j\}_{j=1}^n$ where $M_1 \cap I \supset \dots \supset M_n \cap I$, and if $M_j \cap I = M_{j+1} \cap I$ then $M_j \setminus I \subset M_{j+1} \setminus I$. For all j either $M_j \leq M_{j+1}$ or $M_j \ni M_{j+1}$. Whenever $M_j \leq M_{j+1}$ apply Lemma 6.3 as in Example 6.5 so that we may remove M_{j+1} from our list. By induction, we obtain $M_1 \ni M_2 \ni \dots \ni M_n'$. If I does not have an upper bound then $I \leq M_i$, $I \leq M_j$ and $M_i \subsetneq M_j$ leads to a contradiction. □

Corollary 6.10 *Given a short exact sequence $0 \rightarrow M \rightarrow N \rightarrow I \rightarrow 0$, where I is an interval module and N is a finite direct sum of interval modules, it follows that $W_1(d_{\mu\text{odim}})(M, N) \leq \mu(I)$.*

Proof Let f denote the given map $N \rightarrow I$. Apply Theorem 6.9 with $\text{coker } f = 0$. We have

$N \cong N' \oplus \bigoplus_{j=1}^n M_j$, where each M_j is an interval module with $I \leq M_j$, and $M_1 \ni M_2 \ni \dots \ni M_n$. Furthermore,

$$M \cong N' \oplus (M_n \setminus I) \oplus \bigoplus_{j=1}^{n-1} M_j \setminus ((M_j \setminus M_{j+1}) \cap I).$$

It follows that

$$W_1(d_{\mu\text{odim}})(M, N) \leq \mu(M_n \cap I) + \sum_{j=1}^{n-1} \mu((M_j \setminus M_{j+1}) \cap I) = \mu(M_1 \cap I) = \mu(I).$$

□

6.3 Induced algebraic matching

In this section we show that for monomorphisms and epimorphisms of persistence modules there is an induced algebraic matching of interval modules.

Say that two intervals I and J have the *same right end* if $\sup I = \sup J$ and $\sup I \in I$ iff $\sup J \in J$.

Theorem 6.11 (Induced algebraic matching for monomorphisms) *Let $f : M \rightarrow N$ be a monomorphism between persistence modules with direct-sum decompositions into finitely many interval modules. Then there are internal direct sum decompositions $M = \bigoplus_{a \in A} M_a$ and $N = \bigoplus_{a \in A} N_a$ where each M_a is either an interval module or zero and each N_a is an interval module such that following hold. For all $a \in A$, if M_a is nonzero then M_a and N_a have the same right end, $p'_a f i_a : M_a \rightarrow N_a$ is a monomorphism, where $i_a : M_a \rightarrow M$ and $p'_a : N \rightarrow N_a$ are the canonical maps, for all other interval modules N_b with the same right end as M_a , $p'_b f i_a = 0$ and for all other interval modules M_b with the same right end as N_a , $p'_a f i_b = 0$.*

Proof Let $M = \bigoplus_{k=1}^m M_k$ and $N = \bigoplus_{j=1}^n N_j$. The map f determines and is determined by the maps $f_{j,k} := p'_j f i_k$, where $i_k : M_k \rightarrow M$ and $p'_j : N \rightarrow N_j$ are the canonical maps. Our proof is by a matrix reduction argument. Since f is a monomorphism, for each M_k there exists an N_j with the same right end such that $M_k \subseteq N_j$ and $f_{j,k}$ is nonzero (see Lemma 4.9 and Lemma 4.10(1)).

Partition the intervals in $\{M_k\}_{k=1}^m$ and $\{N_j\}_{j=1}^n$ into subsets with the same right end. Use this partition to order the $\{M_k\}$ and $\{N_j\}$. For the $\{M_k\}$ and $\{N_j\}$ with the same right end, order them by reverse-inclusion and inclusion, respectively.

Consider one of the blocks $\{M_k\}, \{N_j\}$ with the same right end. Choose k_1 so that M_{k_1} is a largest interval. Let N_{j_1} be a smallest one in the block with f_{j_1,k_1} nonzero. Apply Lemma 6.3 iteratively to N_{j_1} and the other N_j in the block for which f_{j,k_1} is nonzero (see Example 6.4). We obtain a basis for N such that f_{j_1,k_1} is nonzero and f_{j,k_1} is zero for the other N_j in the block. Reorder the N_j in the block so that N_{j_1} is first. Next, apply Lemma 6.3 iteratively to M_{k_1} and the other M_k in the block for which $f_{j_1,k}$ is nonzero (see Example 6.5). We obtain a basis for M such that f_{j_1,k_1} is nonzero and $f_{j_1,k}$ is zero for the other M_k in the block.

Now consider a next largest M_{k_2} in the block. Since f is a monomorphism, there is a smallest N_{j_2} with $j_2 \neq j_1$ such that f_{j_2,k_2} is nonzero. Again apply Lemma 6.3 iteratively to obtain a basis for N such that f_{j_2,k_2} is nonzero and f_{j,k_2} is zero for the N_j in the block with $j \neq j_2$. Reorder the N_j in the block so that N_{j_2} is second. Also, apply Lemma 6.3 iteratively to obtain a basis for M such that f_{j_2,k_2} is nonzero and $f_{j_2,k}$ is zero for the M_k in the block with $k \neq k_2$. Continue in the same way for the remainder of the M_k in the block. Repeat for each of the blocks.

For each M_k , let N_k be the corresponding direct summand of N obtained by the above procedure. For the remaining N_j , let $M_j = 0$. □

Corollary 6.12 *Let $f : M \rightarrow N$ be a monomorphism between persistence modules with direct-sum decompositions into finitely many interval modules. Then $W_1(d_{\mu \circ \dim})(M, N) \leq \int_P \dim(\text{coker } f) d\mu$.*

Proof By Theorem 6.11, $M = \bigoplus_a M_a$ and $N = \bigoplus_a N_a$ where each M_a is an interval module or zero and each N_a is an interval module, and $f_a := p'_a f i_a$ is a monomorphism. Note that M_a and N_a have the same right ends and that $\text{coker } f_a = N_a \setminus M_a$. We remark that there may be $b \neq a$ such that $p'_b f i_a$ is nonzero (see Theorem 6.7).

By the rank-nullity theorem, $\int_P \dim(\text{coker } f) d\mu = \int_P (\dim N - \dim M) d\mu = \sum_a \int_P (\dim N_a - \dim M_a) d\mu = \sum_a \int_P \dim(N_a \setminus M_a) d\mu = \sum_a \int_P \dim(\text{coker } f_a) d\mu = \sum_a d_{\mu \circ \dim}(M_a, N_a)$. Therefore $W_1(d_{\mu \circ \dim})(M, N) \leq \int_P \dim(\text{coker } f) d\mu$. □

The following is the Matlis dual (Miller 2020a, Section 2.5) of Theorem 6.11, and the result follows by Matlis duality. However, we give an independent, elementary proof. Say that two intervals I and J have the same left end if $\inf I = \inf J$ and $\inf I \in I$ iff $\inf J \in J$.

Theorem 6.13 (Induced algebraic matching for epimorphisms) *Let $f : M \rightarrow N$ be an epimorphism between persistence modules with direct-sum decompositions into finitely many interval modules. Then there are internal direct sum decompositions $M = \bigoplus_{a \in A} M_a$ and $N = \bigoplus_{a \in A} N_a$ where each M_a is an interval module and each N_a is either an interval module or zero such that the following hold. For all $a \in A$, if N_a is nonzero then M_a and N_a have the same left end, $p'_a f i_a : M_a \rightarrow N_a$ is an epimorphism, where $i_a : M_a \rightarrow M$ and $p'_a : N \rightarrow N_a$ are the canonical maps, for all other interval modules M_b with the same left end as N_a , $p'_a f i_b = 0$ and for all other interval modules N_b with the same left end as M_a , $p'_b f i_a = 0$.*

Proof Let $M = \bigoplus_{k=1}^m M_k$ and $N = \bigoplus_{j=1}^n N_j$. The map f determines and is determined by the maps $f_{j,k} := p'_j f i_k$, where $i_k : M_k \rightarrow M$ and $p'_j : N \rightarrow N_j$ are the canonical maps. Our proof is by a matrix reduction argument. Since f is an epimorphism, for each N_j there exists an M_k with the same left end such that $N_j \subseteq M_k$ and $f_{j,k}$ is nonzero (see Lemma 4.9 and Lemma 4.10(1)).

Partition the intervals in $\{M_k\}_{k=1}^m$ and $\{N_j\}_{j=1}^n$ into subsets with the same left end. Use this partition to order the $\{M_k\}$ and $\{N_j\}$. For the $\{M_k\}$ and $\{N_j\}$ with the same left end, order them by inclusion and reverse-inclusion, respectively.

Consider one of the blocks $\{M_k\}, \{N_j\}$ with the same left end. Choose j_1 so that N_{j_1} is a largest interval. Let M_{k_1} be a smallest interval in the block with $f_{j_1,k}$ nonzero. Apply Lemma 6.3 iteratively to M_{k_1} and the other M_k in the block for which $f_{j_1,k}$ is nonzero (see Example 6.5). We obtain a basis for M such that f_{j_1,k_1} is nonzero and $f_{j_1,k}$ is zero for the other M_k in the block. Reorder the M_k in the block so that M_{k_1} is first. Next, apply Lemma 6.3 iteratively to N_{j_1} and the other N_j in the block for which f_{j,k_1} is nonzero (see Example 6.4). We obtain a basis for N such that f_{j_1,k_1} is nonzero and f_{j,k_1} is zero for the other N_j in the block.

Now consider a next largest N_{j_2} in the block. Since f is an epimorphism, there is a smallest M_{k_2} with $k_2 \neq k_1$ such that f_{j_2, k_2} is nonzero. Again apply Lemma 6.3 iteratively to obtain a basis for M such that f_{j_2, k_2} is nonzero and $f_{j_2, k}$ is zero for the M_k in the block with $k \neq k_2$. Reorder the M_k in the block so that M_{k_2} is second. Also, apply Lemma 6.3 iteratively to obtain a basis for N such that f_{j_2, k_2} is nonzero and f_{j, k_2} is zero for the N_j in the block with $j \neq j_2$. Continue in the same way for the remainder of the N_j in the block. Repeat for each of the blocks.

For each N_j , let M_j be the corresponding direct summand of M obtained by the above procedure. For the remaining M_k , let $N_k = 0$. □

Corollary 6.14 *Let $f : M \rightarrow N$ be an epimorphism between persistence modules with direct-sum decompositions into finitely many interval modules. Then $W_1(d_{\mu \circ \dim})(M, N) \leq \int_P \dim(\ker f) d\mu$.*

Proof By Theorem 6.13, $M = \bigoplus_a M_a$ and $N = \bigoplus_a N_a$ where each M_a is an interval module and each N_a is an interval module or zero, and $f_a := p'_a f i_a$ is an epimorphism. Note that M_a and N_a have the same left ends and that $\ker f_a = M_a \setminus N_a$. We remark that there may be $b \neq a$ such that $p'_a f i_b$ is nonzero (see Theorem 6.9).

By the rank-nullity theorem, $\int_P \dim(\ker f) d\mu = \int_P (\dim M - \dim N) d\mu = \sum_a \int_P (\dim M_a - \dim N_a) d\mu = \sum_a \int_P \dim(M_a \setminus N_a) d\mu = \sum_a \int_P \dim(\ker f_a) d\mu = \sum_a d_{\mu \circ \dim}(M_a, N_a)$. Therefore $W_1(d_{\mu \circ \dim})(M, N) \leq \int_P \dim(\ker f) d\mu$. □

6.4 The W_1 isometry theorem

In this section we prove a W_1 isometry theorem, first in the finite case and then in the general case. The main ingredients are the induced algebraic matching theorems of the previous section.

Proposition 6.15 *Let $M, N \in \mathbf{Vect}_{ds}^P$. Then $d_{\mu \circ \dim}(M, N) \leq W_1(d_{\mu \circ \dim})(M, N)$.*

Proof We need to show that $d_{\mu \circ \dim}(M, N) \leq \inf \|\{d_{\mu \circ \dim}(M_a, N_a)\}_{a \in A}\|_1$, where the infimum is taken over all isomorphisms $M \cong \bigoplus_{a \in A} M_a$ and $N \cong \bigoplus_{a \in A} N_a$, where each M_a and N_a is either 0 or an interval module.

Let $M \cong \bigoplus_{a \in A} M_a$ and $N \cong \bigoplus_{a \in A} N_a$, where each M_a and N_a is either 0 or an interval module. For each $a \in A$, since M_a and N_a are either zero or an interval module, there is a zigzag γ_a of interval modules from M_a to N_a of length at most two with cost $d_{\mu \circ \dim}(M_a, N_a)$. Add identity maps to these zigzags so that they are all of the form $\cdot \rightarrow \cdot \leftarrow \cdot \rightarrow \cdot \leftarrow \cdot$. By taking the direct sum of the maps in these zigzags, we obtain a zigzag in \mathbf{Vect}_{ds}^P from M to N . Since the kernel and cokernel of a direct sum is the direct sum of the kernels and cokernels, respectively, the cost of this zigzag equals the sum of the costs of the zigzags γ_a . The result follows. □

Say that a persistence module M has finite total persistence if $\dim(M)$ is integrable, that is $\int_P \dim(M) d\mu < \infty$.

Remark 6.16 This condition can be weakened substantially using primary decomposition (Thomas 2019; Miller 2020b).

Theorem 6.17 (W_1 isometry theorem) *Let $M, N \in \mathbf{Vect}_{\mathbf{ds}}^{\mathbf{P}}$ such that each has finite total persistence. Then $W_1(d_{\mu\text{odim}})(M, N) = d_{\mu\text{odim}}(M, N)$.*

Proof For simplicity, denote $d_{\mu\text{odim}}$ by d . By Proposition 6.15, we have that $W_1(d)(M, N) \geq d(M, N)$. So, it remains to show that $W_1(d)(M, N) \leq d(M, N)$.

Let $\varepsilon > 0$. By definition, there exists a zigzag γ from M to N given by

$$M = M_0 \xrightarrow{f_1} M_1 \xleftarrow{f_2} M_2 \xrightarrow{f_3} \dots \xleftarrow{f_n} M_n = N$$

such that $\text{cost}_{\mu\text{odim}}(\gamma) < d(M, N) + \frac{\varepsilon}{2}$. It follows that $d(M_{i-1}, M_i) < \infty$ for all $i = 1, \dots, n$. If M_{i-1} has finite total persistence and M_i does not then $d(M_{i-1}, M_i) = \infty$. Thus we may assume that each M_i has finite total persistence.

By the triangle inequality,

$$W_1(d)(M, N) \leq \sum_{i=1}^n W_1(d)(M_{i-1}, M_i). \tag{6.18}$$

Let $1 \leq i \leq n$. By assumption, we have $M_i \cong \bigoplus_{j=1}^{\infty} I_{i,j}$, where $I_{i,j}$ is an interval module or zero. Since M_i has finite total persistence, we may choose N_i such that

$$(\mu \circ \text{dim})\left(\bigoplus_{j=N_i+1}^{\infty} I_{i,j}\right) < \frac{\varepsilon}{8n}. \tag{6.19}$$

Let M'_i denote $\bigoplus_{j=1}^{N_i} I_{i,j}$ and let M''_i denote $\bigoplus_{j=N_i+1}^{\infty} I_{i,j}$. Let $\iota_i : M'_i \rightarrow M_i$ and $\pi_i : M_i \rightarrow M'_i$ denote the canonical inclusion and projection maps.

By the triangle inequality,

$$\begin{aligned} W_1(d)(M_{i-1}, M_i) &\leq W_1(d)(M_{i-1}, M'_{i-1}) + W_1(d)(M'_{i-1}, M'_i) + W_1(d)(M'_i, M_i) \\ &< W_1(d)(M'_{i-1}, M'_i) + \frac{\varepsilon}{4n}. \end{aligned} \tag{6.20}$$

Consider the case $f_i : M_{i-1} \rightarrow M_i$. Let $f'_i : M'_{i-1} \rightarrow M'_i$ be given by $f'_i = \pi_i \circ f_i \circ \iota_{i-1}$. Since f'_i factors through its image, by the triangle inequality and Corollaries 6.12 and 6.14,

$$W_1(d)(M'_{i-1}, M'_i) \leq (\mu \circ \text{dim})(\ker f'_i) + (\mu \circ \text{dim})(\text{coker } f'_i). \tag{6.21}$$

Now

$$\begin{aligned} (\mu \circ \text{dim})(\ker f'_i) &\leq (\mu \circ \text{dim})(\ker(\pi_i \circ f_i)) \\ &\leq (\mu \circ \text{dim})(\ker f_i) + (\mu \circ \text{dim})(M''_i) \end{aligned} \tag{6.22}$$

and

$$\begin{aligned}
 (\mu \circ \dim)(\text{coker } f'_i) &\leq (\mu \circ \dim)(\text{coker}(f_i \circ l_{i-1})) \\
 &\leq (\mu \circ \dim)(\text{coker } f_i) + (\mu \circ \dim)(M''_{i-1}).
 \end{aligned}
 \tag{6.23}$$

Combining (6.21), (6.22), (6.23), and (6.19) we have,

$$W_1(d)(M'_{i-1}, M'_i) < (\mu \circ \dim)(\ker f_i) + (\mu \circ \dim)(\text{coker } f_i) + \frac{\varepsilon}{4n}. \tag{6.24}$$

The other case, $f_i : M_i \rightarrow M_{i-1}$ is similar and we obtain the same inequality as (6.24). Combining (6.18), (6.20), and (6.24), we have

$$W_1(d)(M, N) < \text{cost}_{\mu \circ \dim}(\gamma) + \frac{\varepsilon}{2} < d(M, N) + \varepsilon.$$

Therefore $W_1(d)(M, N) \leq d(M, N)$. □

7 Applications

We end by applying our distances to a few simple examples.

7.1 Multiparameter persistence modules

In this section we consider three examples of two-parameter persistence modules and the distances between them.

Example 7.1 Consider the 1-dimensional simplicial complex K at the top of Fig. 1. Let $P = \{0, 1, 2, 3, 4\}^2 \subset \mathbb{Z}^2$ with the usual coordinate-wise partial order and the counting measure μ . Let X be the P -filtration of K given by the vertices a, b, c appearing at $(0, 2), (1, 1), (2, 0)$, respectively, and the edge e appearing at $(3, 2)$ and $(2, 4)$ and the edge f appearing at $(2, 3)$ and $(4, 2)$. See the bottom left of Fig. 1. Let Y be the P -filtration of K given by the vertices a, b, c appearing at $(0, 2), (1, 1), (2, 0)$, respectively, and the edge e appearing at $(2, 3)$ and $(4, 2)$ and the edge f appearing at $(3, 2)$ and $(2, 4)$. See the bottom right of Fig. 1. Note that the two-parameter persistence modules $H_0(X)$ and $H_0(Y)$ have identical dimension vectors.

Now consider $Z := X \cap Y$ and $W := X \cup Y$. Z differs from X and Y in that it has no edges at the indices highlighted in Fig. 1. W differs from X and Y in that it has both edges at the indices highlighted in Fig. 1. The inclusions $Z \xrightarrow{i} X \xrightarrow{k} W$ and $Z \xrightarrow{j} Y \xrightarrow{\ell} W$ induce two zigzags from $H_0(X)$ to $H_0(Y)$.

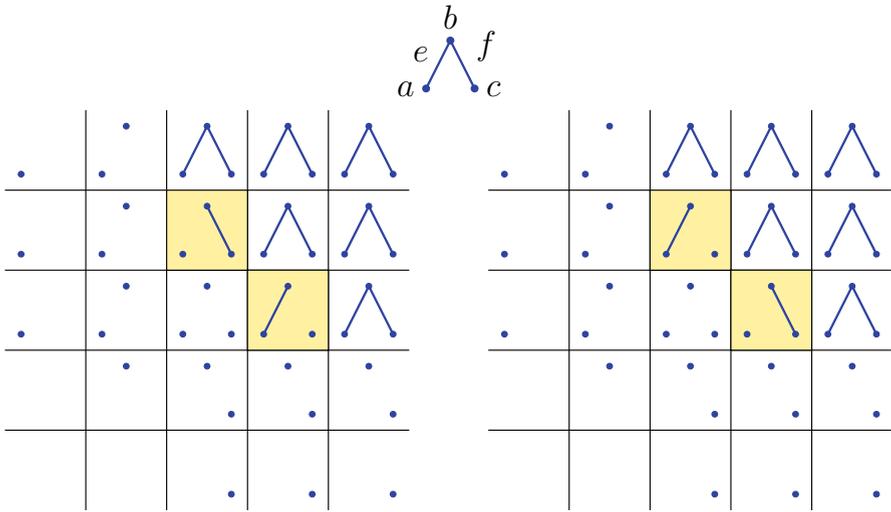
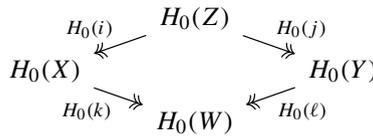


Fig. 1 A one dimensional simplicial complex K (top) and a pair of two-parameter filtrations, X (bottom left) and Y (bottom right). The differences between X and Y are highlighted



Let γ denote the top zigzag and let γ' denote the bottom zigzag. We have $\text{cost}_{\mu\text{odim}}(\gamma) = \sum_p \dim \ker H_0(i) + \sum_p \dim \ker H_0(j) = 2 + 2 = 4$ and $\text{cost}_{\mu\text{odim}}(\gamma') = \sum_p \dim \ker H_0(k) + \sum_p \dim \ker H_0(\ell) = 2 + 2 = 4$. In either case, we have $d_{\mu\text{odim}}(H_0(X), H_0(Y)) \leq 4$.

Since $H_0(X)$ and $H_0(Y)$ have identical dimension vectors, along any zigzag from $H_0(X)$ to $H_0(Y)$ any change in the dimension vector must be later undone. Thus, $d_{\mu\text{odim}}(H_0(X), H_0(Y))$ is even. Since $H_0(X)$ is not isomorphic to $H_0(Y)$, $d_{\mu\text{odim}}(H_0(X), H_0(Y)) \neq 0$. It remains to show that $d_{\mu\text{odim}}(H_0(X), H_0(Y)) \neq 2$. Since $H_0(X)$ and $H_0(Y)$ have identical dimension vectors, this can only happen if there exists a zigzag of length two from $H_0(X)$ to $H_0(Y)$ with middle vector space M where there exists a unique $p \in P$ where $\dim M(p)$ differs from $\dim H_0(X)(p) = \dim H_0(Y)(p)$ by one and for all $q \in P$ with $q \neq p$, $\dim M(q) = H_0(X)(q) = \dim H_0(Y)(q)$. However, because of the two highlighted indices in Fig. 1, there is no such M . Therefore $d_{\mu\text{odim}}(H_0(X), H_0(Y)) = 4$.

Example 7.2 Consider the simplicial complex K at the top of Fig. 2. Let $P = [0, 5]^2 \subset \mathbb{R}^2$ with the usual coordinate-wise partial order and the Lebesgue measure μ . Let $t \in [0, 1]$. Let X_t be the P -filtration of K given by the vertices a, b, c appearing at $(2, 0), (1, 1), (t, 2)$, respectively, and the edge e appearing at $(4, 3)$ and the edge f appearing at $(3, 4)$. For $t < 1$, see the bottom left of Fig. 2, and for $t = 1$, see the bottom right of Fig. 2.

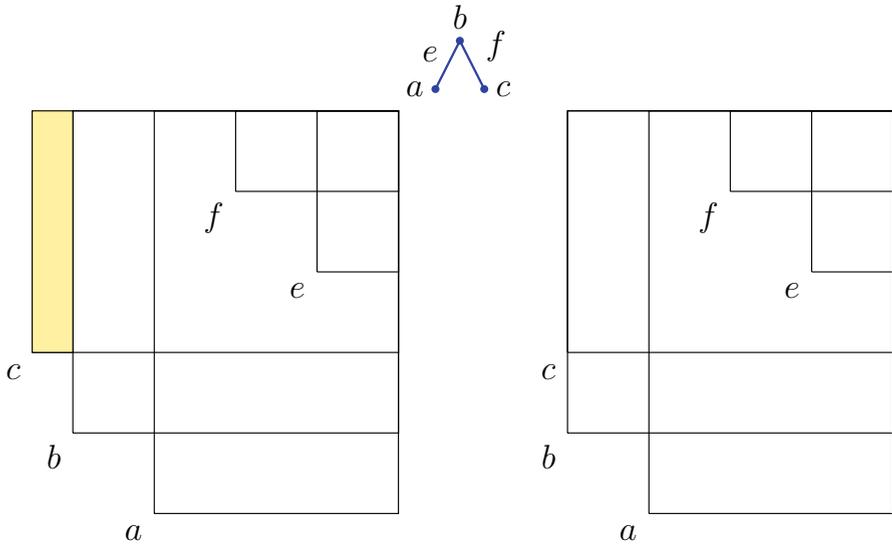


Fig. 2 A one dimensional simplicial complex K (top) and a pair of two-parameter filtrations, X_t (bottom left) and X_1 (bottom right). The difference between X_t and X_1 is highlighted

Consider the two-parameter persistence modules $M_t := H_0(X_t)$ and $M_1 := H_0(X_1)$. The inclusion $i : X_1 \hookrightarrow X_t$ induces a monomorphism $H_0(i) : M_1 \hookrightarrow M_t$. Thus, by Definition 4.4, $d_{\mu\text{odim}}(M_t, M_1) \leq \int_P \dim(\text{coker } H_0(i)) d\mu = 3(1 - t)$. By (4.8), we also have that $d_{\mu\text{odim}}(M_t, M_1) \geq \int_P (\dim M_t - \dim M_1) d\mu = 3(1 - t)$. Therefore $d_{\mu\text{odim}}(M_t, M_1) = 3(1 - t)$. Note that as $t \rightarrow 1$, $d_{\mu\text{odim}}(M_t, M_1) \rightarrow 0$. So, in this example the metric $d_{\mu\text{odim}}$ behaves continuously, as we would like.

Now consider the metrics $W_p(d_{\mu\text{odim}})$, where $1 \leq p \leq \infty$. Let $[x]$ denote the homology class represented by x . For $t < 1$, the persistence module M_t is indecomposable. However, $M_1 \cong A \oplus B$, where A is generated by $[a]$ and $[b]$ and B is generated by $[c] - [b]$. By (4.8), we have that $d_{\mu\text{odim}}(M_t, A) \geq \int_P \dim M_t d\mu - \int_P \dim A d\mu \geq 39 - 29 = 10$ and $d_{\mu\text{odim}}(M_t, B) \geq \int_P \dim M_t d\mu - \int_P \dim B d\mu \geq 39 - 10 = 29$. We also have that $d_{\mu\text{odim}}(0, A) = \int_P \dim A d\mu = 29$, and $d_{\mu\text{odim}}(0, B) = \int_P \dim B d\mu = 10$. Therefore for all $1 \leq p \leq \infty$, $W_p(d_{\mu\text{odim}})(M_t, M_1) \geq \|(10, 10)\|_p \geq 10$, even as $t \rightarrow 1$.

Since indecomposability is unstable, the metrics $W_p(d_{\mu\text{odim}})$ are also unstable. Thus the metric $d_{\mu\text{odim}}$ seems to be a better choice for multiparameter persistence modules than the metrics $W_p(d_{\mu\text{odim}})$.

Example 7.3 Consider the two-parameter persistence modules M , N , and Q which are one-dimensional in the left, middle, and right subsets of the plane in Fig. 3, respectively, and are zero elsewhere.

We have a short exact sequence $0 \rightarrow M \rightarrow N \rightarrow Q \rightarrow 0$. Let μ denote the Lebesgue measure on \mathbb{R}^2 . In the path metric, $d_{\mu\text{odim}}(M, N)$ equals the area of the triangle in the right of Fig. 3. However $W_1(d_{\mu\text{odim}})(M, N)$ equals the area of the

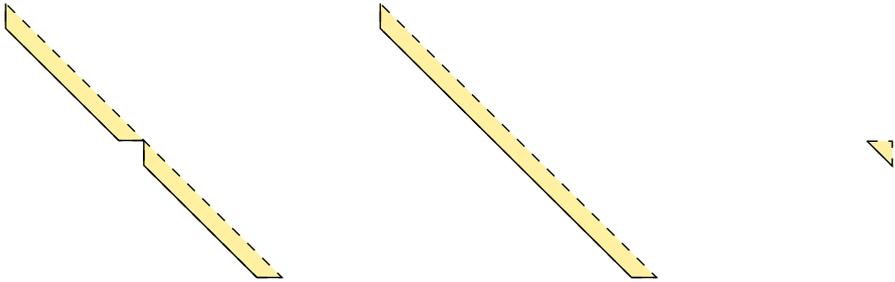


Fig. 3 Middle: a region in the plane whose boundary is a trapezoid. Left: a subset of this region obtained by removing the triangular subregion on the right

trapezoid in the middle of Fig. 3. Thus, M and N are close in the path metric and distant in the Wasserstein metric.

Which metric is more appropriate may depend on the application. For example, let $X = D_1^2 \sqcup D_2^2$ be the disjoint union of two discs. Consider three bifiltrations on X . In the first, the boundary of the first disc, ∂D_1^2 , appears on the solid anti-diagonal line in the middle of Fig. 3 and the remainder of X appears on the dashed line in middle of Fig. 3. Call this bifiltration X_1 . In the second, ∂D_1^2 appears on the two solid anti-diagonal lines in the left of Fig. 3 and the remainder of X appears on the dashed line in left of Fig. 3. Call this bifiltration X_2 . In the third, ∂D_1^2 appears on the left solid anti-diagonal line in the left of Fig. 3, ∂D_2^2 appears on the right solid anti-diagonal line in the left of Fig. 3, and all of X appears on the dashed line in left of Fig. 3. Call this bifiltration X_3 . Then $H_1(X_1) = N$, $H_1(X_2) = M$, and $H_1(X_3) = M$. For X_1 and X_2 , $d_{\mu \circ \dim}(M, N)$ seems to give a better answer for their proximity, but for X_1 and X_3 , $W_1(d_{\mu \circ \dim})(M, N)$ seems to give a better answer for their proximity.

7.2 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 ,

$$\begin{aligned} L &= K \rightarrow K \rightarrow K \leftarrow K \leftarrow K \\ M &= K \rightarrow K \rightarrow K \leftarrow 0 \leftarrow 0 \\ N &= 0 \rightarrow 0 \rightarrow K \leftarrow K \leftarrow K \end{aligned}$$

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,

$$\bullet \rightarrow \bullet \rightarrow \bullet \leftarrow \bullet \leftarrow \bullet \tag{7.4}$$

or modules over the corresponding path algebra, or functors from the category (7.4) 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 (Gabriel 1972). However, we will show that our distances for this quiver behave differently than for the corresponding ordered quiver

$$\bullet \rightarrow \bullet \rightarrow \bullet \rightarrow \bullet \rightarrow \bullet.$$

As we did for persistence modules, we consider the set of objects in the indexing category to be a subset of the integers with the counting measure μ . We then have the corresponding metrics $d_{\mu \circ \dim}$ and $W_p(d_{\mu \circ \dim})$. However, unlike for persistence modules, the metrics $W_1(d_{\mu \circ \dim})$ and $d_{\mu \circ \dim}$ are not equal. Indeed, there is a surjective map $M \oplus N \rightarrow L$ whose kernel has measure one and so $d_{\mu \circ \dim}(M \oplus N, L) = 1$. However, for $W_1(d_{\mu \circ \dim})$ we need to match indecomposables (see Definition 5.1), so $W_1(d_{\mu \circ \dim})(M \oplus N, L) = d_{\mu \circ \dim}(M, L) + d_{\mu \circ \dim}(N, 0) = 2 + 3 = 5$. Which of these metrics is most appropriate will depend on the application.

Acknowledgements The authors would like to thank the referees whose many comments substantially improved the paper and also to José Vélez-Marulanda for corrections to an earlier draft. The first author would like to acknowledge that this research was supported by the NSF-Simons Southeast Center for Mathematics and Biology (SCMB) through the grants National Science Foundation DMS1764406 and Simons Foundation/SFARI 594594, and that this material is based upon work supported by, or in part by, the Army Research Laboratory and the Army Research Office under contract/Grant No. W911NF-18-1-0307.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Bauer, U., Botnan, M.B., Oppermann, S., Steen, J.: Cotorsion torsion triples and the representation theory of filtered hierarchical clustering. *Adv. Math.* **369**, 107171 (2020)
- Bauer, U., Lesnick, M.: Induced matchings and the algebraic stability of persistence barcodes. *J. Comput. Geom.* **6**(2), 162–191 (2015)
- Blumberg, A.J., Lesnick, M.: Universality of the homotopy interleaving distance. [arXiv:1705.01690](https://arxiv.org/abs/1705.01690) [math.AT] (2017)
- Botnan, M.B., Crawley-Boevey, W.: Decomposition of persistence modules. *Proc. Am. Math. Soc.* **148**(11), 4581–4596 (2020)
- Botnan, M.B., Curry, J., Munch, E.: A relative theory of interleavings. [arXiv:2004.14286](https://arxiv.org/abs/2004.14286) [math.CT] (2020)
- Bubenik, P., de Silva, V., Scott, J.: Metrics for generalized persistence modules. *Found. Comput. Math.* **15**(6), 1501–1531 (2015)
- Bubenik, P., de Silva, V., Scott, J.: Interleaving and Gromov–Hausdorff distance. [arXiv:1707.06288](https://arxiv.org/abs/1707.06288) [math.CT] (2017)
- Bubenik, P., Elchese, A.: Universality of persistence diagrams and the bottleneck and Wasserstein distances. *Comput. Geom.* **105–106**, 101882 (2022)
- Bubenik, P., Elchese, A.: Virtual persistence diagrams, signed measures, Wasserstein distances, and Banach spaces. *J. Appl. Comput. Topol.* (2022). <https://doi.org/10.1007/s41468-022-00091-9>
- Bubenik, P., Milićević, N.: Homological algebra for persistence modules. *Found. Comput. Math.* **21**(5), 1233–1278 (2021)
- Bubenik, P., Scott, J.A.: Categorification of persistent homology. *Discrete Comput. Geom.* **51**(3), 600–627 (2014)
- Bubenik, P., Vergili, T.: Topological spaces of persistence modules and their properties. *J. Appl. Comput. Topol.* **2**(3–4), 233–269 (2018)
- Bucur, I., Deleanu, A.: Introduction to The Theory of Categories and Functors. With the Collaboration of Peter J. Hilton and Nicolae Popescu. Pure and Applied Mathematics, vol. XIX. Wiley, London (1968)

- Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L.J., Oudot, S.Y.: Proximity of persistence modules and their diagrams. In: Proceedings of the 25th Annual Symposium on Computational Geometry, SCG '09, pp. 237–246, New York, NY, USA. ACM (2009)
- Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. *Discrete Comput. Geom.* **37**(1), 103–120 (2007)
- Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Mileyko, Y.: Lipschitz functions have L_p -stable persistence. *Found. Comput. Math.* **10**(2), 127–139 (2010)
- Crawley-Boevey, W.: Decomposition of pointwise finite-dimensional persistence modules. *J. Algebra Appl.* **14**(5), 1550066 (2015)
- de Silva, V., Munch, E., Stefanou, A.: Theory of interleavings on categories with a flow. *Theory Appl. Categ.* **33**, Paper No. 21, 583–607 (2018)
- de Silva, V., Munch, E., Patel, A.: Categorical Reeb graphs. *Discrete Comput. Geom.* **55**(4), 854–906 (2016)
- Divol, V., Lacombe, T.: Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *J. Appl. Comput. Topol.* **5**(1), 1–53 (2021)
- Elchesen, A., Mémoli, F.: The reflection distance between zigzag persistence modules. *J. Appl. Comput. Topol.* **3**(3), 185–219 (2019)
- Gabriel, P.: Unzerlegbare Darstellungen. I. *Manuscripta Math.* **6**, 71–103; correction, *ibid.* **6** (1972), 309 (1972)
- Giunti, B., Nolan, J.S., Otter, N., Waas, L.: Amplitudes on abelian categories. [arXiv:2107.09036](https://arxiv.org/abs/2107.09036) [math.AT] (2021)
- Harker, S., Kramár, M., Levanger, R., Mischaikow, K.: A comparison framework for interleaved persistence modules. *J. Appl. Comput. Topol.* **3**(1–2), 85–118 (2019)
- Harrington, H.A., Otter, N., Schenck, H., Tillmann, U.: Stratifying Multiparameter Persistent Homology. *SIAM J. Appl. Algebra Geom.* **3**(3), 439–471 (2019)
- Krause, H.: Krull–Schmidt categories and projective covers. *Expos. Math.* **33**(4), 535–549 (2015)
- Lesnick, M.: The theory of the interleaving distance on multidimensional persistence modules. *Found. Comput. Math.* **15**(3), 613–650 (2015)
- Lesnick, M., Wright, M.: Interactive visualization of 2-d persistence modules. [arXiv:1512.00180](https://arxiv.org/abs/1512.00180) [math.AT] (2015)
- McCleary, A., Patel, A.: Bottleneck stability for generalized persistence diagrams. In: Proceedings of the American Mathematical Society, p. 1 (2019)
- Miller, E.: Modules over posets: commutative and homological algebra (2019)
- Miller, E.: Essential graded algebra over polynomial rings with real exponents (2020a)
- Miller, E.: Primary decomposition over partially ordered groups (2020b)
- Morozov, D., Beketayev, K., Weber, G.H.: Interleaving distance between merge trees. In: Proceedings of TopoInVis (2013)
- Munch, E., Stefanou, A.: The ℓ^∞ -cophenetic metric for phylogenetic trees as an interleaving distance. In: Research in Data Science, volume 17 of Association for Women in Mathematics Series, pp. 109–127. Springer, Cham (2019)
- Pareigis, B.: Categories and functors. Translated from the German. Pure and Applied Mathematics, Vol. 39. Academic Press, New York (1970)
- Patel, A.: Generalized persistence diagrams. *J. Appl. Comput. Topol.* **1**(3–4), 397–419 (2018)
- Popescu, N.: Abelian Categories with Applications to Rings and Modules. Academic Press, London (1973)
- Scolamiero, M., Chachólski, W., Lundman, A., Ramanujam, R., Öberg, S.: Multidimensional persistence and noise. *Found. Comput. Math.* **17**(6), 1367–1406 (2017)
- Skraba, P., Turner, K.: Wasserstein stability for persistence diagrams [arXiv:2006.16824](https://arxiv.org/abs/2006.16824) [math.AT] (2020)
- Stenström, B.: Rings of Quotients. Springer, New York (1975)
- Thomas, A.: Invariants and Metrics for Multiparameter Persistent Homology. PhD thesis, Duke University (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.