WASSERSTEIN STABILITY FOR PERSISTENCE DIAGRAMS

PRIMOZ SKRABA AND KATHARINE TURNER

Abstract. The stability of persistence diagrams is among the most important results in applied and computational topology. Most results in the literature phrase stability in terms of the *bottleneck distance* between diagrams and the ∞-norm of perturbations. This has two main implications: it makes the space of persistence diagrams rather pathological and it is often provides very pessimistic bounds with respect to outliers. In this paper, we provide new stability results with respect to the p-Wasserstein distance between persistence diagrams. This includes an elementary proof for the setting of functions on sufficiently finite spaces in terms of the p-norm of the perturbations, along with an algebraic framework for p-Wasserstein distance which extends the results to wider class of modules. We also provide apply the results to a wide range of applications in topological data analysis (TDA) including topological summaries, persistence transforms and the special but important case of Vietoris-Rips complexes.

Contents

1. Introduction

Persistent (co)homology has been the subject of extensive study in applied topology. Roughly speaking it is a homology theory for filtrations or filtered spaces. A landmark result in applied topology is that persistent (co)homology and more importantly persistence diagram are stable with respect to perturbations of the input filtration. The classical result states:

Theorem 1.1 ([18]). *Let X be a triangulable space with continuous tame functions $f, g : X \to \mathbb{R}$. Then the persistence diagrams $\mathrm{Dgm}(f)$ and $\mathrm{Dgm}(g)$ for their sublevel set filtrations satisfy*

$$d_B(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq ||f - g||_\infty$$

where $d_B(\cdot)$ represents the bottleneck distance.

This result has been generalized to algebraic [2] and categorical settings [7], with recent work strongly aimed at multiparameter and more general settings, particularly where classical notions of persistence diagrams do not exist. Here we study the p-Wasserstein stability of persistence diagrams for $1 \leq p \leq \infty$. This has been far less studied, with existing results almost exclusively in terms of the classical results relating *interleaving distances* between filtrations and the ∞-Wasserstein distance, i.e. bottleneck distance. Upper bounds on the p-Wasserstein distances are less common and often rely on bottleneck stability resulting in pessimistic bounds. Furthermore, there is often a requirement for p to be sufficiently large. p-Wasserstein distances for small values of p, i.e. $p = 1, 2$ rather than $p = \infty$, are important for a number of reasons. From the applications side, the 2-Wasserstein distance on persistence diagrams has been found to be much more effective than bottleneck distance.

One of the main difficulties in establishing a p-Wasserstein bound is that we cannot use an *interleaving*, which is a key tool in persistence stability results. Indeed, at first there does not seem to be a natural algebraic description of Wasserstein distance in the same way as there is for bottleneck distance. Here we take a fundamentally different approach to proving p-Wasserstein stability which, at its core, focuses on a cellular p-Wasserstein stability theorem. The proof exploits the local correspondences between coordinates of the points in the persistence diagram with critical cells in a filtration over a cellular complex. This cellular p-Wasserstein stability theorem then can be modified and applied to a variety of settings to prove a range of stability theorems. The stability of linear representations of persistent homology are usually stated as upper bounds in terms of the 1-Wasserstein distance, for which the pre-existing stability results for p-Wasserstein distances cannot be applied. An important corollary of this paper is that we also get stability results for these linear representations.

Using the cellular stability theorem as inspiration, we also develop the algebraic theory to deal with the p-Wasserstein distance for persistence modules. The description of the distance is different to the notion of an interleaving which is closer to the single morphism characterization from [2]. Our generalized description recovers this result in the special case of bottleneck distance.

Using this theory, we prove the stability results in a more general setting, albeit with more technically involved proofs compared to the cellular stability proof. We also show a Minkowski-type of result connecting what we call the norm of a persistence modules, also called total persistence, in a short exact sequence[1]. The upper bound follows naturally from our study of the Wasserstein distance, but surprisingly, there is a lower bound as well, giving a geometric result for extensions of persistence modules for all $p \geq 1$.

In summary, the main contributions of this paper are:

(1) A cellular p-Wasserstein stability theorem, which also provides a simplified proof of bottleneck stability of diagrams for finite complexes.
(2) We apply the above theorem to produce stability theorems for a number of applications, including grey-scale images, persistent homology transforms, and Vietoris-Rips filtrations. As the 2 - Wasserstein distance is widely used in applications, this addresses a significant gap in the applied topology literature.
(3) We give an algebraic formulation of the p-Wasserstein distance and prove stability which applies to pointwise finite dimensional modules with an additional technical condition,
(4) Upper and lower bounds relating the p-Wasserstein distance of diagrams of modules in a short exact sequence,
(5) Sufficient algebraic conditions which ensure that our algebraic definition yields a pseudo-distance without depending on the structure of persistence modules.

The paper can be thought of as split into two main parts. After introducing the relevant preliminaries, the first part of the paper is focused on results which are immediately relevant for applications in topological data analysis. We present a number of examples which illustrate the inherent instabilities in persistence diagrams, followed by general cellular stability theorem. We then discuss applications to the important cases of image analysis, persistent homology transforms, Vietoris-Rips filtrations, and other topological summaries. The results are focused on explicit bounds for finite cases, which are the main interest in applications.

The second part of the paper provides an algebraic perspective on Wasserstein stability. This includes a formulation which we show is equivalent to the Wasserstein distance for persistence diagrams but which applies to a wider class of modules. The main idea can be thought of the algebraic generalization of the cellular stability proof. This section is not required for the applications and so can be skipped by readers

[1] We note that this is not a norm in the strict sense as persistence modules do not have a vector space structure.

more interested in the practical implications of the results. Likewise, the algebraically minded reader need not go through the applications of the cellular stability theorem.

2. Preliminaries

As described above, we distinguish between settings where the underlying objects are finite and more general settings where the objects may be infinite but still sufficiently nice. This is due to the fact that the finite cases illustrate the ideas and are often sufficient for many applications. In finite settings, we restrict ourselves to functions over a finite CW-complex denoted by K. We generally do not require any additional structure (e.g. simplicial, cubical, etc.). Exceptions are for the analysis of particular applications, e.g. cubical complexes for images (Section 5.1) and the simplicial structure of Vietoris-Rips complexes (Section 5.3). In the algebraic section we consider more generally persistence modules.

Definition 2.1. *A persistence module $\mathcal{F}$ is a collection of vector spaces $\{F_\alpha\}_{\alpha\in\mathbb{R}}$ along with induced maps $\psi_\alpha^\beta : F_\alpha \to F_\beta$ for all $\alpha \leq \beta$ such that ψ_α^α is the identity for all α and $\psi_\alpha^\beta \circ \psi_\beta^\gamma = \psi_\alpha^\gamma$ whenever $\alpha < \beta < \gamma$. If F_α is finite dimensional for all α, then we say $\mathcal{F}$ is pointwise finite dimensional (or p.f.d.).*

One of the most common ways persistence modules arise is via filtrations of finite CW-complexes, especially those associated to functions. Without loss of generality, when considering functions we restrict ourselves to sublevel sets: for $f : K \to \mathbb{R}$, the corresponding sublevel set filtration $\{K_\alpha\}_{\alpha\in\mathbb{R}}$ with

$$K_\alpha = \{\sigma | f(\sigma) \leq \alpha\}.$$

From the definition, it is clear we only consider functions which are piecewise constant on the interior of cells, that is

$$f(\sigma) = \sup_{x\in\sigma} f(x).$$

We further will require that our functions are *monotone* which means that if τ is a face of σ then $f(\tau) \leq f(\sigma)$. This monotone assumption greatly simplifies the exposition as this ensures that all sublevel sets are (closed) CW-complexes. This is the most common setting in applications of persistence. This is more restrictive than the definition in [24], which only includes the condition that the space monotonically non-decreasing and importantly excludes piecewise linear (PL) functions. However, in most cases one can find a piecewise constant function which results in an isomorphic persistence module to the module arising from the PL function. One could generalize the results in Section 4 to a more general setting such as constructible functions, but the increase in technical complications yields relatively little gain, in light of the algebraic framework in Section 7, where we work directly with persistence modules.

Applying the homology functor over a field to the filtration, we obtain the corresponding persistence module, denoted $\{\mathrm{H}_k(K_\alpha)\}_{\alpha\in\mathbb{R}}$ with the maps $\mathrm{H}_k(K_\alpha) \to \mathrm{H}_k(K_\beta)$, induced by the inclusions $K_\alpha \hookrightarrow K_\beta$ for all $\alpha \leq \beta$. For a piecewise constant filtration over a finite CW-complex, the resulting persistence module is pointwise finite dimensional. We restrict ourselves to p.f.d modules (with an additional technical condition – Definition 2.9) due to the following result.

Theorem 2.2 ([21] Theorem 1.1)**.** *A p.f.d. persistence module admits an* interval decomposition*. That is, the module can be decomposed into rank one summands:*

$$\bigoplus_x \mathbb{I}\{\mathbf{b}(x), \mathbf{d}(x)\}$$

which are unique up to isomorphism. This is referred to as a ***persistence barcode*** *and we refer to the elements as bars or intervals.*

By considering the each bar in the persistence barcode as a point in $\mathbb{R}^2$ with the first coordinate $\mathbf{b}(x)$ and second coordinate $\mathbf{d}(x)$, we obtain the **persistence diagram**. We refer to $\mathbf{b}(x)$ as the birth time and $\mathbf{d}(x)$ as the death time.

All the results in this paper apply to the four types of intervals [13], i.e. open-open – $(\mathbf{b}, \mathbf{d})$, open-closed– $(\mathbf{b}, \mathbf{d}]$, closed-open – $[\mathbf{b}, \mathbf{d})$, and closed-closed – $[\mathbf{b}, \mathbf{d}]$ which may appear in the decomposition of persistence modules, hence our choice of notation. Observe that the corresponding persistence diagrams are the same with respect to Wasserstein distance, i.e. the Wasserstein distance between intervals of different types is 0. We note however that the results do not apply as-is for zig-zag modules as this would already require modifying the definition of persistence module (Definition 2.1).

Remark 2.3. *For p.f.d. modules, the decomposition may contain an uncountable number of intervals. However, there can only be countably many intervals where* $\mathbf{b} < \mathbf{d}$*. As in the case of different interval types, removing intervals where* $\mathbf{b} = \mathbf{d}$*, also called an emphemeral submodule* [12]*, induces a Wasserstein distance of 0. Hence, we may assume the decomposition has countably many intervals.*

Definition 2.4. *Let* $\mathrm{Dgm}_k(\mathcal{F})$ *denote the k-dimensional persistence diagram of persistence module* $\mathcal{F}$ *respectively. Taking the grading over dimension, we denote*

$$\mathrm{Dgm}(\mathcal{F}) = \bigoplus_k \mathrm{Dgm}_k(\mathcal{F}). \tag{1}$$

As a notational convenience, we index over the points in the diagram wherever possible:

$$x \in \mathrm{Dgm}(\mathcal{F}) \leftrightarrow (\mathbf{b}(x), \mathbf{d}(x)) \in \mathbb{R}^2$$

Our main focus is the Wasserstein distance between diagrams.

Definition 2.5. *Given two diagrams,* $\mathrm{Dgm}_k(\mathcal{F})$ *and* $\mathrm{Dgm}_k(\mathcal{G})$*, the* (p,q)*-Wasserstein distance is*

$$W_{p,q}(\mathrm{Dgm}_k(\mathcal{F}), \mathrm{Dgm}_k(\mathcal{G})) = \inf_{\mathbf{M}} \left(\sum_{x \in \mathrm{Dgm}_k(K(f))} ||x - \mathbf{M}(x)||_q^p \right)^{\frac{1}{p}}$$

where $\mathbf{M} : \mathrm{Dgm}_k(K(f)) \to \mathrm{Dgm}_k(K'(g))$ *represents a bijection, possibly to or from points on the diagonal. The total* (p,q)*-Wasserstein distance is defined as*

$$W_{p,q}(\mathrm{Dgm}(\mathcal{F}), \mathrm{Dgm}(\mathcal{G})) = \left(\sum_k \left(W_{p,q}(\mathrm{Dgm}_k(\mathcal{F}), \mathrm{Dgm}_k(\mathcal{G})\right)^p \right)^{\frac{1}{p}}$$

For all fixed p, all $W_{p,q}$ are bi-Lipschitz equivalent. Hence, we focus on the case $p = q$, which we denote by W_p. Taking the limit $p \to \infty$ recovers the bottleneck distance

$$W_\infty(\mathrm{Dgm}(\mathcal{F}), \mathrm{Dgm}(\mathcal{G})) = \inf_{\mathbf{M}} \sup_k \sup_{x \in \mathrm{Dgm}_k(\mathcal{F})} ||x - \mathbf{M}(x)||_\infty \tag{2}$$

It is worth commenting on the relative strength of stability results for different p. The following statements illustrate that bottleneck stability is the weakest form of stability. We first note the following lemma[2] whose proof can be found in Appendix A.

Lemma 2.6. *For any* $p' \leq p$*, given persistence diagrams* $\mathrm{Dgm}(\mathcal{F})$ *and* $\mathrm{Dgm}(\mathcal{G})$*,*

$$W_p(\mathrm{Dgm}(\mathcal{F}), \mathrm{Dgm}(\mathcal{G})) \leq W_{p'}(\mathrm{Dgm}(\mathcal{F}), \mathrm{Dgm}(\mathcal{G})).$$

Thus when bounding the p-Wasserstein distance from above, the smaller p is, the stronger a stability result. An important quantity for understanding the Wasserstein distance more abstractly is *the norm of a persistence module*:

Definition 2.7. *Let the p-norm of a p.f.d. module* $\mathcal{F}$ *be the p-th root of the sum of the p-th power of the lengths of the bars, i.e.*

$$||\mathcal{F}||_p = \left(\sum_{x \in \mathrm{Dgm}(\mathcal{F})} \ell(x)^p \right)^{\frac{1}{p}}$$

where $\ell(x) = \mathbf{d}(x) - \mathbf{b}(x)$*, i.e. the length of a bar.*

This also called the *total persistence* and is a natural quantity going back to [25], where it was observed that the squared total persistence or as we refer to it, the squared 2-norm is precisely the running time for the incremental algorithm for computing a persistence diagram. By analogy, we refer to this as the norm of a persistence module/diagram, although we do not assume the properties of a norm.

The remainder of this section consists of algebraic preliminaries for the category of persistence modules. The reader who is either familiar with this material or is interested primarily in the applications of Wasserstein stability, may skip the remainder of this section as these concepts are used in Section 7.

Unsurprisingly, when studying the stability of persistence modules it is critical to understand the morphisms between persistence modules. The following facts are standard and are included for completeness.

[2]This is a standard result but the proof in the appendix is included for completeness for the reader.

Scholium 2.8. *Given a morphism between persistence modules, the kernel, image, and cokernel are well-defined and are persistence modules themselves. This was studied extensively in* [2]*, where the following facts were proven.*

- *A monic morphism between persistence modules, denoted by $\mathcal{A} \overset{f}{\hookrightarrow} \mathcal{B}$, induces an injective multiset map between the death times of $\mathcal{A}$ and the death times of $\mathcal{B}$. This induces an injective set map denoted $f_\star$ from the indecomposables of $\mathcal{A}$ to the indecomposables of $\mathcal{B}$.*
- *An epic morphism between persistence modules, denoted by $\mathcal{A} \overset{g}{\twoheadrightarrow} \mathcal{B}$, induces an injective multiset map between the birth times of $\mathcal{B}$ and the birth times of $\mathcal{A}$. This induces an injective set map denoted $g^\star$ from the indecomposables of $\mathcal{B}$ to the indecomposables of $\mathcal{A}$.*

We make extensive use of these induced matchings. *Finally, we note that standard constructions such as the pullback and pushout are well-defined. Furthermore, a prior application of these properties including a discussion of short exact sequences of persistence modules, can be found in* [28].

An important assumption is that we assume that all modules have a common parameterization – in many cases, e.g. Vietoris-Rips filtrations, this is a natural assumption and it avoids problems which arising in comparing persistence modules and diagrams in more general settings. An important point in this paper is that all the morphisms we consider, unless specifically stated, are ungraded with respect to this parameterization. Consider a morphism between persistence modules, $f : \mathcal{A} \to \mathcal{B}$ and let $\mathcal{A}_\alpha$ denote the vector space at α. For any $[x] \in \mathcal{A}_\alpha$

$$[x] \mapsto f([x]) \in \mathcal{B}_\alpha.$$

Note that interleaving maps *do not* satisfy this condition. For further discussion of this notion in the context of bottleneck distance, see [28].

While interval decompositions exist for p.f.d. persistence modules, these can still not be sufficiently well-behaved for our purposes. As an intermediate step in our proofs, we will often consider finitely generated modules. So we want a construction for approximating p.f.d. modules by finitely generated modules subject to the following technical condition of bounded p-energy.

Definition 2.9. *If $\mathcal{F}$ is p.f.d. persistence module, we say $\mathcal{F}$ has bounded p-energy if*

$$\sum_{x \in \mathrm{Dgm}(\mathcal{F})} \ell(x)^p < \infty$$

Observe that the p.f.d. condition guarantees that there are a countable number of intervals. We focus most of our discussion on finite intervals. Infinite intervals are technically simpler as they correspond to freely generated summands and so the condition that there only a countable number is sufficient.

Our goal is to approximate the submodule consisting only of finite intervals with a finitely generated module. In Section 7, we will make use of the following construction.

Lemma 2.10. *Let $\mathcal{F}$ be a p.f.d. module with bounded p-energy. For any $\varepsilon > 0$, there exists a finitely generated module $\mathcal{F}'$ with morphisms $i_{\mathcal{F}} : \mathcal{F}' \hookrightarrow \mathcal{F}$ and $q_{\mathcal{F}} : \mathcal{F} \twoheadrightarrow \mathcal{F}'$ such that*

$$W_p(\mathrm{Dgm}(\mathcal{F}), \mathrm{Dgm}(\mathcal{F}')) \leq \varepsilon, \qquad ||\operatorname{coker} i_{\mathcal{F}}||_p \leq \varepsilon, \qquad ||\ker q_{\mathcal{F}}||_p \leq \varepsilon.$$

Proof. The proof is constructive – without loss of generality, assume that there are a countable number of intervals. Sort the intervals in the decomposition of $\mathcal{F}$ by decreasing length, i.e. $i \leq j$ implies $(\mathbf{d}_i - \mathbf{b}_i) \geq (\mathbf{d}_j - \mathbf{b}_j)$. Since $\mathcal{F}$ has bounded p-energy, the sum of the p-lengths is finite and there exists a constant K, such that

$$\sum_{i=K}^{\infty} (\mathbf{d}_i - \mathbf{b}_i)^p < \varepsilon^p$$

Define

$$\mathcal{F}' = \bigoplus_{i=1}^{K-1} \{\mathbf{b}_i, \mathbf{d}_i\}$$

Note that the interval type is equivalent to the corresponding interval in $\mathcal{F}$. To bound the Wasserstein distance, consider the transport plan of sending all intervals with index $i > k$ to the diagonal. By assumption, this has a cost less ε. The morphisms $i_{\mathcal{F}}$ and $q_{\mathcal{F}}$ are the obvious morphisms matching the index, with $q_{\mathcal{F}}$ mapping all intervals after index k to 0. The bound on norms then again follows by assumption. □

We conclude this section with the following remark.

Remark 2.11. *The central notion of the paper, the Wasserstein distance is defined in several different settings, so we often omit the term Wasserstein. For example, we refer to the above simply as the* diagram distance *(Definition 2.5), the distance between points embedded in $\mathbb{R}^d$ as the* point set distance *(Definition 5.6), and the algebraic notion as the* module distance *(Definition 7.7).*

Furthermore, there are several different notions of points we consider: points in the persistence diagrams, elements of a point set in $\mathbb{R}^d$, and it will be useful for exposition to consider the vertices of a Vietoris-Rips complex as points. To minimize confusion, we restrict the term points to refer to persistence diagrams, preferring the term vertices for the more geometric notions[3]*. For a complete list of notation, see Appendix C.*

3. Existing stability results and their limitations

As already mentioned, the classical stability results all involve the bottleneck distance between persistence diagrams. A complete overview of these results is beyond the scope of this paper and we direct the reader to [14] for the stability of geometric constructions and [8] for the categorical foundations of ∞-Wasserstein stability. This should not be considered as a complete list as there is a large body of work on stability which we do not address review here.

3.1. Lipschitz functions on compact manifolds.

The most relevant related work to the results presented in this paper can be found in [19]. To the best of our knowledge, this paper contains the main existing stability result for bounding the $(p \neq \infty)$-Wasserstein distance between two persistence diagrams. It is for the setting of sub-level set filtrations of Lipschitz functions.

Theorem 3.1 (Wasserstein Stability Theorem [19])**.** *Let X be a triangulable, compact metric space that implies bounded degree-k total persistence, for $k \geq 1$ and let $f, g : X \to \mathbb{R}$ be two tame Lipschitz functions. Then*

$$W_p(f, g) \leq C^{1/p} \|f - g\|_\infty^{1-\frac{p}{k}}$$

for all $p \geq k$, where $C = C_X \max\{Lip(f)^k, Lip(g)^k\}$ and C_X is a constant dependent on X.

To put our results into context, it is worthwhile understanding the limitations of this theorem. We will find lower bounds on C_X and k, restricting ourselves to the case where X is a compact d-dimensional manifold. An important aspect is the bounded degree-k total persistence which will force this stability result to only hold for sufficiently large p.

Definition 3.2. *A metric space X* implies bounded degree k-total persistence *if there exists a constant C_X that depends only on X such that*

$$||\mathrm{Dgm}(X(f))||_k^k < C_X$$

for every tame function f with Lipschitz constant $Lip(f) \leq 1$.

To construct a counterexample for functions over manifolds we will use a function which is the sum of functions with supports over disjoint balls of small radius. Slight modifications would also provide counterexamples for stratified spaces.

Lemma 3.3. *Given an d-dimensional compact Riemannian manifold X and $r > 0$ small enough, there exists a packing of $\left\lfloor \frac{\mathrm{vol}(X)}{\kappa \omega_d 2^d r^d} \right\rfloor$ disjoint balls of radius r in X, where ω_d is the volume of the d-dimensional Euclidean ball and κ is a constant which depends on the infimum of the scalar curvature of X.*

Proof. We note that since we restrict ourselves to compact manifolds, the curvature must be bounded. Lemma 3.3 in [4] establishes a lower bound on the volume of a ball of radius r. The result then follows from standard arguments involving packing and covering numbers. a packing bound. $\square$

Lemma 3.4. *Let X an d-dimensional compact manifold. If X has bounded degree-k total persistence then $k \geq d$.*

Proof. We will prove this via a counterexample of the contrapositive. Let $P = \{p_1, \dots p_N\}$ be the centers of a packing of $N = \left\lfloor \frac{\mathrm{vol}(X)}{\kappa \omega_d 2^d r^d} \right\rfloor$ disjoint balls of radius r in X (such a packing exists by Lemma 3.3). Set

[3]While this has the drawback of resulting in references to vertices in a point set, we feel this is a good compromise

$T_{r,p}$ to be a teepee shaped function about p with height r, with $T_{r,p}(x) = \max\{r - d(x,p), 0\}$. We then consider functions $f_r = \sum_{i=1}^N T_{r,p_i}$(see Figure 1). Observe that f is 1-Lipschitz. Then,

$$||\mathrm{Dgm}(X(f))||_k^k = \sum_{i=1}^{N} r^k = \left\lfloor \frac{\mathrm{vol}(X)}{\kappa \omega_d 2^d r^d} \right\rfloor r^k = O(r^{d-k})$$

For this to be uniformly bounded for all small enough $r > 0$, we require $d \geq k$. □

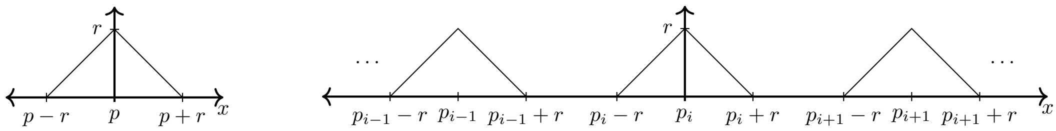


FIGURE 1. (Left) The teepee function and (right) sum of teepee functions from Lemma 3.4.

The same example used in the above lemma, coupled with g the zero function provides a lower bound on C_X. We see that C_X grows linearly as a function of the volume of X.

Other papers that prove Wasserstein stability results are few and far between. In [15], Chen and Edelsbrunner consider non-Lipschitz functions on non-compact spaces, using scale-space diffusion. They focus on convergence properties as opposed to stability but also attain some Wasserstein stability results. Crucially, just as in the Lipschitz case, this p-Wasserstein stability only holds for $p > d$ where d is the dimension of the domain. The condition $p > d$ also appears in stability results for Čech filtrations, or equivalently distance filtrations, for point clouds in $\mathbb{R}^d$.

3.2. **Erroneous appeals to previous p-Wasserstein stability results.** Unfortunately, the Lipschitz Wasserstein stability theorem in [19] appears to be one of the most misunderstood and miscited results within the field of topological data analysis. Common errors include using a small p (often 1 or 2) for high dimensional data, assertions that the persistence diagrams depend Lipschitz-continuously on data and applying the theorem to Vietoris-Rips filtrations. Luckily, many of the erroneous applications can now be covered by the stability results in this paper. Rather than discuss individual examples, in Section 6, we examine the consequences for topological summaries, which are increasingly the most common way to apply persistence diagrams to data.

4. Cellular Wasserstein Stability

We begin with a result mirroring the classical stability theorem by bounding the differences at the chain level, which will induce an upper bound on the p-Wasserstein distance between the corresponding diagrams. As stated in Section 2, we remind the reader that K is be a finite CW-complex and $f : K \to \mathbb{R}$ is a monotone function on the complex so all sublevel sets are subcomplexes. The persistence module associated to the sublevel set filtration of such a function, denoted $\mathrm{Dgm}(K(f))$, is p.f.d. Since we work with a fixed complex, we shorten the notation to $\mathrm{Dgm}(f)$. We begin by defining an L^p norm between of a function on K.

Definition 4.1. *The L^p norm of a function $f : K \to \mathbb{R}$ is given by*

$$\|f\|_p^p = \sum_{\sigma \in K} |f(\sigma)|^p.$$

This induces a distance between functions.

Definition 4.2. *The L^p distance between two monotone functions $f, g : K \to \mathbb{R}$ is given by*

$$\|f - g\|_p^p = \sum_{\sigma \in K} |f(\sigma) - g(\sigma)|^p.$$

Note that this notion of the L^p distance between functions is analogous to the L^p distance for functions over discrete sets where the sum here is over the discrete set of cells. It is different to distances involving integration of functions over the space.

Question 4.3. *What conditions are required on the space and the function to relate the cellular p-norm to the more common functional p-norm, e.g. the integral of f^p over the space?*

Remark 4.4. *Given that we restrict ourselves to piecewise constant monotone functions, the only additional condition we require is that the underlying complex is finite. We have chosen to present the results in this way, so that it is clear that it applies to common settings including simplicial and cubical complexes.*

The main idea in the proof of cellular Wasserstein stability is to bound the Wasserstein distance by considering a straightline homotopy between f and g. We split the straightline homotopy into finitely many sub-intervals where a local result will hold, and then collect together the summands for the final desired inequality. By focusing on small enough sub-intervals we can exploit a consistent correspondence between the coordinates of the points in the persistence diagram with critical cells in the filtration. Though we phrase the proof in a different way, this idea first appeared in [20] to show a bottleneck stability result for vineyards. Indeed the idea of tracking of points in the persistence diagram is fundamental to the definition of vineyards. We refer the reader to [20] for more details and an algorthmic perspective on tracking critical simplices.

Lemma 4.5. *Let $f : K \to \mathbb{R}$ be monotone functions on a finite CW complex K. There exists an injective map*

$$\Omega_f : \mathrm{Dgm}(f) \to K \times (K \cup \{\emptyset\})$$

such that for $x \in \mathrm{Dgm}(f)$ with $\mathbf{d}(x)$ finite has $\Omega_f(x) = (\sigma, \tau)$ for some (σ, τ) such that $f(\sigma) = \mathbf{b}(x)$ and $f(\tau) = \mathbf{d}(x)$, and for $\mathbf{d}(x)$ is infinite then $\Omega_f(x) = (\sigma, \emptyset)$ for some σ with $f(\sigma) = \mathbf{b}(x)$.

Proof. Extend the partial order induced by f to a total order such that it remains a filtration. Note that each cell either creates a new class or bounds an existing class. Hence, each cell may be assigned to a unique pair and hence a point in the persistence diagram, possibly on the diagonal. This inverse of this assignment defines an injective map as required. □

We know that Ω_f will be unique when f is injective but in general Ω_f is not unique and depends on the choice of extension to a total order. Note that we do not use any properties of the choice of the extension other than that it defines Ω_f. For a more in-depth discussion, see [39][Lemma 3.11]. This has also been extensively used to perform gradient descent over persistence diagram [35, 16], with a mathematical framework proposed in [33].

Definition 4.6. *A* critical pair *is a pair of cells in the image of Ω_f (where with an abuse of notation one of these cells may be $\emptyset$) such that $f(\tau) - f(\sigma) > 0$. We say a cell is* critical *if it part of a critical pair. If we restrict Ω_f to the k-dimensional diagram, we refer to those cells as k-critical.*

We first consider the easy case: where the ordering of cells does not change.

Lemma 4.7. *Let $f_t : K \to \mathbb{R}$, $t \in [a, a']$ be a continuous family of monotone functions over a CW complex K such that for all $a < s < a'$ the order (potentially with equality) of the function values of the cells remains the same. Then*

$$W_p(\mathrm{Dgm}(f_a), \mathrm{Dgm}(f_{a'})) \leq ||f_a - f_{a'}||_p.$$

If we fix a homology dimension k then

$$W_p(\mathrm{Dgm}_k(f_a), \mathrm{Dgm}_k(f_{a'}))^p \leq \sum_{\dim(\sigma)\in\{k,k+1\}} |f_a(\sigma) - f_{a'}(\sigma)|^p.$$

Proof. Fix $c \in (a, a')$. For any two cells σ_1 and σ_2, if there exists an $s \in (a, a')$ such that $f_s(\sigma_1) = f_s(\sigma_2)$ then $f_t(\sigma_1) = f_t(\sigma_2)$ for all $t \in (a, a')$. Furthermore $f_c(\sigma_1) < f_c(\sigma_2)$ implies that $f_t(\sigma_1) < f_t(\sigma_2)$ for all $t \in (a, a')$.

Denote the off-diagonal points by $x \in \mathrm{Dgm}(f_c)$, which we also write as as $\{(\mathbf{b}(x)^c, \mathbf{d}(x)^c)\}$. Recall that persistence diagrams are multisets. As such we would use multiple indices whenever the location of points in persistence diagram coincide. Let Ω_{f_c} be as defined in Lemma 4.5 and hence assigns cells in K to each of the $\mathbf{b}(x)^c$ and the $\mathbf{d}(x)^c$ that correspond to the critical cells of the persistent homology by sublevel sets of f_c. Denote these cells by $\sigma(\mathbf{b}(x)^c)$ and $\sigma(\mathbf{d}(x)^c)$ (with $\sigma(\mathbf{d}(x)^c) = \emptyset$ if $\mathbf{d}(x)^c$ is infinite), i.e. $\Omega_{f_c}(\mathbf{b}(x)^c, \mathbf{d}(x)^c) = (\sigma(\mathbf{b}(x)^c), \sigma(\mathbf{d}(x)^c))$.

As the order of the cells in the sublevel set filtration is consistent over $t \in (a, a')$, Ω is constant and the only difference is that times are reparameterised. The persistent homology classes in $\mathrm{Dgm}(f_c)$ denoted by $(\mathbf{b}(x)^c, \mathbf{d}(x)^c) = (f_c(\sigma(\mathbf{b}(x)^c)), f_c(\sigma(\mathbf{d}(x)^c)))$ live in the persistent diagram for f_t as $(f_t(\sigma(\mathbf{b}(x)^c), f_t(\sigma(\mathbf{d}(x)^c)))$. Thus, the off-diagonal points $x \in \mathrm{Dgm}(f_t)$ are

$$\{(f_t(\sigma(\mathbf{b}(x)^c)), f_t(\sigma(\mathbf{d}(x)^c))) : x \in \mathrm{Dgm}(f_c)\}.$$

Note that here we assign $f_t(\emptyset) = \infty$ for all t.

The f_t are continuous with respect to t and so bottleneck stability implies that

$$\mathrm{Dgm}(f_a) = \{(f_a(\sigma(\mathbf{b}(x)^c)), f_a\sigma(\mathbf{d}(x)^c)) : x \in \mathrm{Dgm}(f_c)\}$$
$$\mathrm{Dgm}(f_{a'}) = \{(f_{a'}(\sigma(\mathbf{b}(x)^c)), f_{a'}(\sigma(\mathbf{d}(x)^c))) : x \in \mathrm{Dgm}(f_c)\}$$

where we are potentially moving points into the diagonal as $t \to a$ or $t \to a'$. This labelling determines a matching ϕ between the diagrams $\mathrm{Dgm}(f_a)$ and $\mathrm{Dgm}(f_{a'})$ with

$$\phi((f_a(\sigma(\mathbf{b}(x)^c)), f_a(\sigma(\mathbf{d}(x)^c))) = (f_{a'}(\sigma(\mathbf{b}(x)^c)), f_{a'}(\sigma(\mathbf{d}(x)^c))).$$

The p-th power of the cost of this matching ϕ is bounded by

$$\begin{aligned}
&\sum_x \|(f_a(\sigma(\mathbf{b}(x)^c)), f_a(\sigma(\mathbf{d}(x)^c))) - (f_{a'}(\sigma(\mathbf{b}(x)^c)), f_{a'}(\sigma(\mathbf{d}(x)^c)))\|_p^p \\
&\quad = \sum_x |f_a(\sigma(\mathbf{b}(x)^c)) - f_{a'}(\sigma(\mathbf{b}(x)^c))|^p + \sum_{\mathbf{d}(x)\neq\infty} |f_a(\sigma(\mathbf{d}(x)^c)) - f_{a'}(\sigma(\mathbf{d}(x)^c))|^p \\
&\quad \leq \sum_{\sigma\in K} |f_a(\sigma) - f_{a'}(\sigma)|^p = \|f_a - f_{a'}\|_p^p
\end{aligned}$$

Note that we are using the facts that the distance to the diagonal is bounded by the distance to any specific point on the diagonal, and that every cell appears at most once in the middle sum. Since the p-Wasserstein distance is the smallest possible matching cost, we conclude that

$$W_p(\mathrm{Dgm}(f_a), \mathrm{Dgm}(f_{a'}))^p \leq \|f_a - f_{a'}\|_p^p.$$

If cells have coinciding function values over the interval (a, a') then we observe that changes in homology at that function value must be caused by one of the set of cells. However, as the bound on the p-Wasserstein distance only uses the function values which are equal for all cells in the set over the entire interval, the distance is independent of the choice of critical cell.

The proof for when we restrict to homology dimension k follows from the observation that only addition of the k-cells and the $k+1$-cells can affect homology in dimension k. □

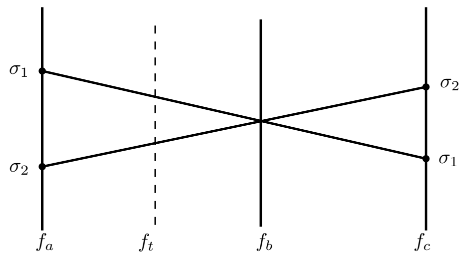


FIGURE 2. A linear interpolation between functions f_a and f_c can be subdivided into intervals where the ordering does not change in the interior of the interval. If the underlying space is a finite CW complex, the number of such intervals is finite.

To complete the proof we observe that any straightline homotopy may be divided into intervals where the ordering does not change, see Figure 2. Since our underlying space is a finite CW complex, the number of such intervals must also be finite. This implies one of our main theorems:

Theorem 4.8 (Cellular Wasserstein Stability Theorem). *Let $f, g : K \to \mathbb{R}$ be monotone functions. Then*

$$W_p(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \|f - g\|_p.$$

If we fix a homology dimension k then

$$W_p(\mathrm{Dgm}_k(f), \mathrm{Dgm}_k(g))^p \leq \sum_{\dim(\sigma)\in\{k,k+1\}} |f(\sigma) - g(\sigma)|^p.$$

Proof. Let $f_t : K \to \mathbb{R}$ be the linear interpolation between f and g as t varies. That is, for $t \in [0, 1]$ and $\sigma \in K$, let $f_t(\sigma) = (1 - t)f(\sigma) + tg(\sigma)$. Observe that f_t is monotone for all t and that for $0 \leq a \leq a' \leq 1$ we have $\|f_a - f_{a'}\|_p = |a' - a|\|f - g\|_p$.

Each of the functions $t \mapsto f_t(\sigma)$ is linear which implies that $f_t(\sigma) = f_t(\hat{\sigma})$ for two or more values of t if and only if $f(\sigma) = f(\hat{\sigma})$ and $g(\sigma) = g(\hat{\sigma})$, in which case $f_t(\sigma) = f_t(\hat{\sigma})$ for all $t \in [0, 1]$).

There are only finitely many values $t = a_1, a_2, \ldots a_n$ in $(0, 1)$, sorted in increasing value, where there exists $\sigma, \hat{\sigma}$ with $f_t(\sigma) = f_t(\hat{\sigma})$ but $f(\sigma) \neq f(\hat{\sigma})$. Set $a_0 = 0, a_{n+1} = 1$. In each of the intervals (a_i, a_{i+1}) the order of the function values is consistent and so we can apply Lemma 4.7.

$$\begin{aligned} W_p(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) &\leq \sum_{i=0}^{n} W_p(\mathrm{Dgm}(f_{a_i}), \mathrm{Dgm}(f_{a_{i+1}})) \\ &\leq \sum_{i=0}^{n} \|f_{a_i} - f_{a_{i+1}}\|_p = \sum_{i=0}^{n} (a_{i+1} - a_i)\|f - g\|_p = \|f - g\|_p \end{aligned}$$

The proof for when we restrict to homology dimension k is highly analogous, using the corresponding bound in Lemma 4.7 restricting to homology dimension k. □

5. Applications

We present some applications of the results of the cellular Wasserstein stability theorem. Sublevel set filtrations of grayscale images and persistent homology transforms of different geometric embeddings of the same simplicial complex are both cases which involve height functions determined by vertex values. We will prove Lipschitz stability in terms of the l_p norms over the set of vertices, where the Lipschitz constants are bounded by the number of cells in the links of each vertex. We also will prove some immediate corollaries for stability of Rips filtrations.

5.1. **Stability of the sublevel set filtrations of grayscale images.** Our first application is for the stability of grayscale images. The natural application is to two dimensional images, however we will state our results for more general d-dimensional images. An image is a real-valued piecewise constant function where each pixel/voxel is assigned a value. There are two main methods in the literature for creating a filtration of cubical complexes from a grayscale image.

Method 1. We can create a cubical complexes from a 2D image where each pixel corresponds to a 2-dimensional cubical cell. The edges correspond to sides of the pixels, and vertices to the corners. This construction naturally extends to higher dimensional images. There is a natural sublevel set filtration induced on the complex: the image defines values for the maximal dimensional cells (i.e. pixels/voxels) and the function values for lower dimensional cells are given as the minimum value over all cofaces.

Method 2. We can also consider the dual of the cubical complex in Method 1, which is again a cubical complex. In a 2D image we have a vertex for each pixel and an edge for each pair of neighbouring pixels (not including diagonals), and 2-cells where four pixels intersect. This construction naturally extends to higher dimensional images. We can build a filtration on this cubical complex by setting the values on the vertices as those of the pixel/voxel values provided, and setting the function values for higher dimensional cells as the maximum value over all faces.

It is worth noting that the sublevel set filtrations for these two methods can result in substantially different persistent homology. This difference stems from whether diagonally neighbouring pixels are considered connected. However, applying Theorem 4 separately to both methods obtains stability for both methods individually.

Theorem 5.1. *Let f and g be the grayscale functions for two images. Let $\hat{f}$ and $\hat{g}$ be the corresponding monotone functions on cubical complexes generated by either Method 1 or 2 (both $\hat{f}$ and $\hat{g}$ using the same method). Then we have the stability result*

$$W_p(\mathrm{Dgm}(\hat{f}), \mathrm{Dgm}(\hat{g})) \leq \left(\sum_{i=0}^{d} 2^{d-i} \binom{d}{i} \right) ||f - g||_p$$

Proof. Let us suppose we are using Method 1 for constructing our persistence diagrams. As the underlying space is a cubical complex, changing the function value of a maximal cell can affect all of the lower dimensional cells it contains. Each d-dimensional hypercube contains $2^{d-k}\binom{d}{k}$ k-dimensional hypercubes on its boundary. Summing up over all dimensions yields a bound on how many cell-values change when we change the value of a pixel. Applying Theorem 4 yields the result.

The proof for Method 2 is similar. Changing the function value of a vertex can affect all of the higher dimensional cofaces. There are at most $2^k\binom{d}{k}$ possibly affected k-dimensional cells and applying Theorem 4 completes the proof. $\square$

5.2. **Stability of persistent homology transforms.** The study of persistent homology transforms are a relatively recent development in the persistent homology literature [41, 26, 22] with applications to statistical shape analysis. Given an embedded shape $M \subset \mathbb{R}^n$, every unit vector v corresponds to a height function in direction v,

$$h_v : M \to \mathbb{R}$$
$$h_v : x \mapsto \langle x, v\rangle.$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. The resulting k-dimensional persistence diagram computed by filtering M by the sub-level sets of h_v, is denoted $\mathrm{Dgm}_k(h_v^M)$. This diagram records geometric information from the perspective of direction v. As v changes, the persistent homology classes track geometric features in M. The key insight behind the persistent homology transform (PHT) is we do not need to choose a specific direction; by considering the persistent homology from every direction, we obtain a surprising amount of information. The most general setting of the PHT is for constructible sets which are compact definable sets. We denote the set of constructible subsets of $\mathbb{R}^d$ by $\mathrm{CS}(\mathbb{R}^d)$. This includes all compact piece-wise linear or semi-algebraic sets.

Definition 5.2. *The* Persistent Homology Transform PHT *of a constructible set $M \in CS(\mathbb{R}^d)$ is the map* $\mathrm{PHT}(M) : S^{d-1} \to \mathrm{Dgm}^d$ *that sends a direction to the set of persistent diagrams gotten by filtering M in the direction of v:*

$$\mathrm{PHT}(M) : v \mapsto \big(\mathrm{Dgm}_0(h_v^M), \mathrm{Dgm}_1(h_v^M), \ldots, \mathrm{Dgm}_{d-1}(h_v^M)\big)$$

where $h_v^M : M \to \mathbb{R}$, $h_v^M(x) = \langle x, v\rangle$ is the height function on M in direction v. Letting the set M vary gives us the map

$$\mathrm{PHT} : CS(\mathbb{R}^d) \to C^0(S^{d-1}, \mathrm{Dgm}^d),$$

where $C^0(S^{d-1}, \mathrm{Dgm}^d)$ is the set of continuous functions from S^{d-1} to Dgm^d, the latter being equipped with some Wasserstein p-distance.

The persistent homology transform is a complete descriptor of constructible sets; for $M_1, M_2 \subset \mathbb{R}^d$, $\mathrm{PHT}(M_1) = \mathrm{PHT}(M_2)$ implies $M_1 = M_2$ as subsets of $\mathbb{R}^d$. This was originally proved in [41] for piecewise linear compact subsets in $\mathbb{R}^2$ and $\mathbb{R}^3$, and then the more general proof was given in [22] and independently in [26]. Here we restrict ourselves to different embeddings of the same simplicial complex where the embeddings are each determined linearly by the placement of the vertices. We call such shapes the geometric vertex embedding of a finite simplicial complex and notably they are always a constructible set.

Definition 5.3. *Let K be a finite simplicial complex with vertex set V. For $f : V \to \mathbb{R}^d$ we can define a piece-wise linear extension of $f : K \to \mathbb{R}^d$ by setting $f(\sum a_i v_i) = \sum a_i f(v_i)$. We call $f : K \to \mathbb{R}^d$ a* geometric vertex embedding *of K if $f(K)$ is a geometric realization of K (i.e. no self-intersections).*

We can define a metric on the space of persistent homology transforms by considering the appropriate integrals of Wasserstein distances in each direction. We obtain a different distance for each $p \in [1, \infty]$

Definition 5.4. *For $p \in [1, \infty)$, and constructible sets $M_1, M_2 \subset \mathbb{R}^d$ we can define a p-PHT distance between M_1, M_2 by*

$$d_p^{\mathrm{PHT}}(M_1, M_2) = \left(\int_{S^{d-1}} W_p(\mathrm{Dgm}(h_v^{M_1}), \mathrm{Dgm}(h_v^{M_2}))^p \, dv \right)^{1/p}.$$

We can use the cellular Wasserstein stability result to prove a stability theorem for the persistent homology transforms of different vertex embeddings of the same simplicial complex.

Theorem 5.5. *Fix a simplicial complex K with vertex set V. Let C_K be the maximum number of simplicies any one vertex of K is a member of. Let $C_{p,d} = 2\omega_{d-2} \int_0^{\frac{\pi}{2}} \cos^p(\theta) \sin^{d-2}(\theta)\, d\theta$ where ω_{d-2} is the area of the unit sphere S^{d-2}. Then for $f, g : K \to \mathbb{R}^d$ be different geometric vertex embeddings we have*

$$d_p^{PHT}(f(K), g(K)) \leq \left(C_K C_{p,d} \sum_{v \in V} \|f(v) - g(v)\|_2^p \right)^{1/p}.$$

Proof. Define functions $k_w^f : K \to \mathbb{R}$ by setting

$$k_w^f([v_0, \dots v_n]) = \max\{h_w(f(v_0), h_w(f(v_1)), \dots h_w(f(v_n))\},$$

and $k_w^g : K \to \mathbb{R}$ analogously. As discussed in [22], the sublevel set filtrations of k_w^f and $h_w^{f(K)}$ have the same persistent homology. Similarly, k_w^g and $h_w^{g(K)}$ give the same sub-level set persistent homology. By Theorem 4.8, we know that

$$W_p(\mathrm{Dgm}(k_w^f), \mathrm{Dgm}(k_w^g))^p \le \sum_{\Delta \in K} |k_w^f(\Delta) - k_w^g(\Delta)|^p.$$

For any finite set X,

$$\left| \max_{x \in X} f(x) - \max_{y \in X} g(y) \right| \le \max_{x \in X} |f(x) - g(x)|$$

which implies

$$\sum_{\sigma \in K} \left|k_w^f(\sigma) - k_w^g(\sigma)\right|^p \le \sum_{\sigma \in K} \max_{v \in \sigma} \left\{ \left|k_w^f(v) - k_w^g(v)\right|^p \right\} \le C_K \sum_{v \in V} \left|k_w^f(v) - k_w^g(v)\right|^p.$$

But $k_w^f(v) = \langle w, f(v) \rangle$ and $k_w^g(v) = \langle w, g(v) \rangle$ which implies

$$\sum_{\sigma \in K} \left|k_w^f(\sigma) - k_w^g(\sigma)\right|^p \le C_K \sum_{v \in V} \left|\langle w, f(v) - g(v) \rangle\right|^p.$$

$$\begin{aligned}
d_p^{PHT}(f(K), g(K))^p &= \int_{S^{d-1}} W_p(\mathrm{Dgm}(h_w^f), \mathrm{Dgm}(h_w^g))^p \, dw \\
&\le \int_{S^{d-1}} C_K \sum_{v \in V} |\langle w, f(v) - g(v) \rangle|^p \, dw \\
&\le C_K \sum_{v \in V} \int_{S^{d-1}} |\langle w, f(v) - g(v) \rangle|^p \, dw \\
&= C_K \sum_{v \in V} \|f(v) - g(v)\|_2^p \int_{S^{d-1}} |\langle w, e_1 \rangle|^p \, dw \\
&= C_K \sum_{v \in V} \|f(v) - g(v)\|_2^p \, 2 \int_0^{\frac{\pi}{2}} \cos^p(\theta) \omega_{d-2} \sin^{d-2}(\theta) \, d\theta \\
&= C_K C_{p,d} \sum_{v \in V} \|f(v) - g(v)\|_2^p
\end{aligned}$$

□

In particular $C_{p,3} = \frac{4\pi}{p+1}$, $C_{p,2} \le 2$ for all p, and $C_{1,d} = \frac{2\omega_{d-2}}{d-1}$.

5.3. **Stability results for Rips complexes.** The goal of this section is to bound the change in $\mathrm{Dgm}(\mathcal{R}(\mathcal{P}))$ as the underlying point set $\mathcal{P}$ changes, so we first find the appropriate distance between point sets. We will first state the definition of the Wasserstein distances between measures. This views each point cloud as a sum of point masses. In order for this distance to be defined we require that the point sets have same cardinality.

Definition 5.6. *Let $\mathcal{P}_0$ and $\mathcal{P}_1$ be two finite point sets in $\mathbb{R}^d$ and assume $|\mathcal{P}_0| = |\mathcal{P}_1|$. Define the* point set Wasserstein distance *between them as*

$$W_p^{point\ set}(\mathcal{P}_0, \mathcal{P}_1) = \inf_\phi \left(\sum_{v \in \mathcal{P}_0} ||v - \phi(v)||^p \right)^{\frac{1}{p}}.$$

where ϕ is a bijection.

Since we are dealing with finite sets this definition is equivalent to the classical Wasserstein distance between the measures μ_0 and μ_1 where $\mu_i = \sum_{x \in \mathcal{P}_i} \delta_x$.

Before stating our stability results let us first recall some basic definitions.

Definition 5.7. *Given a point cloud $\mathcal{P} \subset \mathbb{R}^d$, the Vietoris-Rips complex is the simplicial complex $\mathcal{R}_\delta(\mathcal{P})$ where a k-simplex is a subsets of $k+1$ points $\{v_1, \dots, v_{k+1}\}$ such that $||v_i - v_j||_2 \le \delta$ for all $i, j = 1, \dots, k+1$.*

We implicitly use the identification of the vertices of $\mathcal{R}(\mathcal{P})$ and the points of $\mathcal{P}$. By varying δ, we obtain a filtration.

Definition 5.8. *The Vietoris-Rips filtration (or simply Rips filtration) of a point set $\mathcal{P}$ is the filtration $\{\mathcal{R}_\delta(\mathcal{P})\}$ induced by ranging δ from 0 to ∞. The corresponding persistence diagram is denoted* $\mathrm{Dgm}(\mathcal{R}(\mathcal{P}))$.

Theorem 5.9. *Fix $M > 0$. For all $p \geq 1$, for all k, and all point clouds $\mathcal{P}_0, \mathcal{P}_1$ with $|\mathcal{P}_0|, |\mathcal{P}_1| = M$ we have*

$$W_p(\mathrm{Dgm}_k(\mathcal{R}(\mathcal{P}_0)), \mathrm{Dgm}_k(\mathcal{R}(\mathcal{P}_1))) \leq 2\binom{M-1}{k}^{1/p} W_p^{\textit{point set}}(\mathcal{P}_0, \mathcal{P}_1)$$

where $\mathrm{Dgm}_k(\mathcal{R}(\mathcal{P}_0))$ and $\mathrm{Dgm}_k(\mathcal{R}(\mathcal{P}_1))$ are the k-dimensional persistence diagrams for the Vietoris-Rips filtration on the point sets $\mathcal{P}_0$ and $\mathcal{P}_1$ respectively. Furthermore,

$$W_p(\mathrm{Dgm}(\mathcal{R}(\mathcal{P}_0)), \mathrm{Dgm}(\mathcal{R}(\mathcal{P}_1))) \leq 2^{M/p+1} W_p^{\textit{point set}}(\mathcal{P}_0, \mathcal{P}_1).$$

Proof. Let $\phi : \mathcal{P}_0 \to \mathcal{P}_1$ be a bijection which achieves the minimum of

$$W_p(\mathcal{P}_0, \mathcal{P}_1) = \inf_\phi \left(\sum_{v \in \mathcal{P}_0} ||v - \phi(v)||^p \right)^{\frac{1}{p}}.$$

Relabel the points in $\mathcal{P}_0 = \{x_1, \dots x_M\}$ and $\mathcal{P}_1 = \{y_1, \dots y_M\}$ so that $\phi(x_i) = y_i$. Let K be the complete simplicial complex on M vertices $\{v_1, \dots v_M\}$. Define a functions $f, g : K \to \mathbb{R}$ by $f([v_{i_0}, v_{i_1}, \dots v_{i_k}])$ the time when $[x_{i_0}, x_{i_1}, \dots x_{i_k}]$ is included in $\mathcal{R}(\mathcal{P}_0)$ and $g([v_{i_0}, v_{i_1}, \dots v_{i_k}])$ the time when $[y_{i_0}, y_{i_1}, \dots y_{i_k}]$ is included in $\mathcal{R}(\mathcal{P}_1)$.

Suppose for now that $k \geq 1$. Then

$$\begin{aligned}|f([v_{i_0}, v_{i_1}, \dots v_{i_k}]) - g([v_{i_0}, v_{i_1}, \dots v_{i_k}])| &= |(\max_{j,l}\{\|x_{i_j} - x_{i_l}\|\} - \max_{j,l}\{\|y_{i_j} - y_{i_l}\|\}| \\ &\leq \max_{j,l} |\|x_{i_j} - x_{i_l}\| - \|y_{i_j} - y_{i_l}\||.\end{aligned}$$

By the triangle inequality $|\|x_{i_j} - x_{i_l}\| - \|y_{i_j} - y_{i_l}\|| < \|x_{i_j} - y_{i_j}\| + \|x_{i_l} - y_{i_l}\|$. This implies $|f([v_{i_0}, v_{i_1}, \dots v_{i_k}]) - g([v_{i_0}, v_{i_1}, \dots v_{i_k}])| \leq \max_{j \neq l} \|x_{i_j} - y_{i_j}\| + \|x_{i_l} - y_{i_l}\| \leq 2 \max_j \|x_{i_j} - y_{i_j}\|$.

Since K is the complete simplicial complex over M vertices, each edge $[v_i, v_j]$ appears in $\binom{M-2}{k-1}$ k-simplices (we only need to decide which extra $k-1$ vertices to include).

Using the cellular stability theorem,

$$\begin{aligned}&W_p(\mathrm{Dgm}_k(\mathcal{R}(\mathcal{P}_0)), \mathrm{Dgm}_k(\mathcal{R}(\mathcal{P}_1)))^p \\ &\quad\leq \sum_{[v_{i_0}, \dots v_{i_k}]} |f([v_{i_0}, \dots v_{i_k}]) - g([v_{i_0}, \dots v_{i_k}])|^p + \sum_{[v_{i_0}, \dots v_{i_{k+1}}]} |f([v_{i_0}, \dots v_{i_k}]) - g([v_{i_0}, \dots v_{i_{k+1}}])|^p \\ &\quad\leq \sum_i \binom{M-2}{k-1} 2^p \|x_i - y_i\|^p + \sum_i \binom{M-2}{k} 2^p \|x_i - y_i\|^p \\ &\quad\leq 2^p \binom{M-1}{k} W_p^{\text{point set}}(\mathcal{P}_0, \mathcal{P}_1)^p\end{aligned}$$

For $k = 0$ the calculations are even easier as the vertex values are all 0.

$$\begin{aligned}W_p(\mathrm{Dgm}_0(\mathcal{R}(\mathcal{P}_0)), \mathrm{Dgm}_0(\mathcal{R}(\mathcal{P}_1)))^p &\leq \sum_{i<j} |f([v_i, v_j]) - g([v_i, v_j])|^p \\ &= \sum_{i<j} |\|x_i - x_j\| - \|y_i - y_j\||^p \\ &\leq \sum_{i<j} (2\|x_i - y_i\|)^p \\ &= 2^p W_p^{\text{point set}}(\mathcal{P}_0, \mathcal{P}_1)^p\end{aligned}$$

To prove the second part, we again use the cellular stability theorem to compute

$$\begin{aligned}
W_p(\mathrm{Dgm}_k(\mathcal{R}(\mathcal{P}_0)), \mathrm{Dgm}_k(\mathcal{R}(\mathcal{P}_1)))^p &\leq \sum_{k=1}^{M} \sum_{[v_{i_0}, v_{i_1}, \dots v_{i_k}]} |f([v_{i_0}, v_{i_1}, \dots v_{i_k}]) - g([v_{i_0}, v_{i_1}, \dots v_{i_k}])|^p \\
&\leq \sum_{k=0}^{M} \binom{M}{k} 2^p \sum_i \|x_i - y_i\|^p \\
&= 2^p 2^M W_p^{\text{point set}}(\mathcal{P}_0, \mathcal{P}_1)^p.
\end{aligned}$$

□

6. Consequences for topological summaries

Stability results for topological summary statistics computed from persistent homology bound the distance between the summaries from above in terms of the p-Wasserstein distance of the corresponding persistence diagrams, most often using 1-Wasserstein distance between the input persistence diagrams. This provides the weakest upper bound on the distance between topological summaries as $W_1(X,Y) \geq W_p(X,Y)$ for all $p \geq 1$. This would not be a problem in of itself, except that when the distance between persistence diagrams is upper bounded by some geometric measure of difference of the input, e.g. Hausdorff distance, the bottleneck distance i.e. ∞-Wasserstein distance, is almost exclusively used. In this case, the ∞-Wasserstein distance provides the loosest lower bound with respect to the distortion measure of the input. We thus can not combine the resulting bounds to provide further stability. However, as we provide upper bounds on the 1-Wasserstein distance between the diagrams, we obtain immediate corollaries for stability results of these topological summary statistics in terms of the input data. In this section, we present some of the positive results that follow.

Corollary 6.1. *Suppose that (T, d) is a metric space of topological summaries such that*

$$d(T(X), T(Y)) \leq C_T W_1(X, Y)$$

for all persistence diagrams X, Y. If f, g are monotone functions over cellular complex K, with $T(f)$ and $T(g)$ the corresponding topological summaries for the sub-level filtrations of f and g respectively then

$$d(T(f), T(g)) \leq C_T \|f - g\|_{K,1}.$$

The proof follows directly from the earlier stability results in this paper. It can be directly applied to a number of topological summaries already in the literature where the condition of $d(T(X), T(Y)) \leq C_T W_1(X, Y)$ for all persistence diagrams X, Y has already been established. This includes

(1) sliced Wasserstein kernel, $C_T = 1$, see [11]
(2) persistent images, $C_T = 1$, see [1]
(3) persistent scale space, $C_T = 1$, see [36, 32]
(4) weighted Betti curves see [31, 42, 17],
(5) learned/optimized representations [29, 30],
(6) persistent homology rank function [37], $C_T = 1$ (see Corollary 6.5).

Related results for topological summaries constructed via grey scale images, the persistent homology transform, and Rips complexes follow from the theorems in Sections 5.

In the rest of this section, we examine in more detail Lipschitz stability as relates to linear representations of persistence diagrams, providing neccessary conditions. We also consider persistence landscapes which are one of the most common forms of non-linear representations. We prove negative Lipschitz stability results for all L^p function norms of persistence landscapes where $p < \infty$.

6.1. Linear representations of persistence diagrams.

A growingly common form of topological summary statistic are linear representations of persistence diagrams. Examples of linear representations include persistence images, persistent rank functions and weighted Betti curves. We view persistence diagrams as measures over the plane (and call these persistence measures) and then have a function from the plane to some Banach space. The resulting linear representation is the integral of these functions over the persistence measure.

Definition 6.2. *Let $\mathcal{B}$ be a Banach space. A* linear representation *is a function $\Phi : \mathcal{D} \to \mathcal{B}$ such that $\Phi(\mu) = \int_{\mathbb{R}^{2+}} f(x) d\mu(x)$ for some $f : \mathbb{R}^{2+} \to \mathcal{B}$. Here we view persistence diagram X as a measure $\mu_X = \sum_{x \in X} \delta_x$.*

As these topological summaries lie in Banach spaces, often even Hilbert spaces, the number of statistical methods available for analysis increases. Often these constructions of linear representations are justified as maintaining relevant persistence homology information because of stability with respect to 1-Wasserstein distances of the original persistence diagrams. Lipschitz stability with respect to 1-Wasserstein distance of diagrams has been shown for a number of linear representations, see for example persistence scale space kernel [36] and persistence images [17]. Related theoretic bounds for distances between general linear representations are in [23], where they prove conditions when linear representations are continuous with respect to Wasserstein distances and provide a Lipschitz bound for the supremenum norm between linear representations in terms of the 1-Wasserstein distance. hence, in this section, we focus on Lipschitz stability for linear representations into general Banach spaces.

For completeness we recall the necessary and sufficient conditions for Lipschitz stability. Note that all the L_q metrics over $\mathbb{R}^{2+} \cup \Delta$ are bi-Lipschitz equivalent up to a slight change in constant. For the sake of clarity we will restrict the case of the L_1 metric on $\mathbb{R}^{2+} \cup \Delta$.

Theorem 6.3. *The non-trivial linear representation $\Phi : \mathcal{D} \to \mathcal{B}$ is Lipschitz continuous with respect to W_p with constant C if and only if $f : \mathbb{R}^{2+} \cup \Delta \to \mathcal{B}$ is Lipschitz continuous with constant C and $p = 1$.*

Proof. Let us first assume that $\Phi : \mathcal{D} \to \mathcal{B}$ is Lipschitz continuous with respect to W_p with constant C. Let $x \in \mathbb{R}^{2+}$ with $f(x)$ non-trivial. Set X to be the persistence diagram consisting of k copies of x, and Y the persistence diagram containing no off-diagonal points. Now

$$W_p(X, Y) = (k\|x - \Delta\|_p^p)^{1/p} = k^{1/p}\|x - \Delta\|_p.$$

In contrast $\|\Phi(X) - \Phi(Y)\|_{\mathcal{B}} = \|\Phi(X)\|_{\mathcal{B}} = \|k \cdot f(x)\|_{\mathcal{B}} = k\|f(x)\|_{\mathcal{B}}$.

By assumption we have

$$k\|f(x)\|_{\mathcal{B}} \le Ck^{1/p}\|x - \Delta\|_p$$

for all k which clearly creates a contradiction if $p > 1$.

Let $x, y \in \mathbb{R}^{2+} \cup \Delta$ and set X and Y to be the persistence diagrams containing only the diagonal alongside x and y respectively. We have

$$\begin{aligned} \|f(x) - f(y)\|_{\mathcal{B}} &= \|\Phi(x) - \Phi(y)\|_{\mathcal{B}} \\ &\le CW_1(X, Y) \\ &\le C\|x - y\|_1 \end{aligned}$$

where the first inequality follows by assumption and the second because $\phi(x) = y$ determines a matching (which may not necessarily be optimal).

To prove the other direction, suppose $\|f(x) - f(y)\| \le C\|x - y\|_1$ for all $x, y \in \mathbb{R}^{2+} \cup \Delta$. Let $X, Y \in \mathcal{D}$ and let $\mathbf{M}$ be a correspondence between them.

$$\begin{aligned} \|\Phi(X) - \Phi(Y)\| &= \left\| \sum_{x \in X} f(x) - \sum_{y \in Y} f(y) \right\| = \left\| \sum_{x \in X} f(x) - f(\mathbf{M}(x)) \right\| \\ &\le \sum_{x \in X} \|f(x) - f(\mathbf{M}(x))\| \le \sum_{x \in X} C \left\| x - \mathbf{M}(x) \right\|_1 \end{aligned}$$

This holds for all matchings $\mathbf{M}$ and hence $\|\Phi(X) - \Phi(Y)\| \le CW_1(X, Y)$. ☐

Definition 6.4. *We define the k-th dimensional persistent homology rank function corresponding to the filtration K to be*

$$\begin{aligned} \beta_k(K) &: \mathbb{R}^{2+} \to \mathbb{Z} \\ (a, b) &\mapsto \mathbf{rk}\,\mathrm{im}(\mathrm{H}_k(K_a) \to \mathrm{H}_k(K_b)) \end{aligned}$$

where K_a is the filtration at a. Persistent homology rank functions lie in the space of real valued functions over $\mathbb{R}^{2+}$. Given a weighting function ϕ over $\mathbb{R}^{2+}$ we can define an L^q function distance function by

$$d_q(f, h) = \left(\int_{x<y} |f - h|^q \phi(y - x)\, dx\, dy \right)^{\frac{1}{q}} \tag{3}$$

Following [37] *we use $\phi(t) = e^{-t}$.*

Corollary 6.5. *Rank functions with L^q weighted metric are Lipschitz continuous with respect to the p-Wasserstein distances between diagrams if and only if $q = p = 1$. In this case, the Lipschitz constant is* 1.

Proof. We can see that rank functions are a linear representation. Let $\mathcal{B}_q$ be the Banach space of functions over $\mathbb{R}^{2+}$ with norm $\|\beta\|_q^q = \int_{\mathbb{R}^{2+}} |\beta(x,y)|^q \phi(y-x) dxdy$. Define $f : \mathbb{R}^{2+} \to \mathcal{B}$ by $f(a,b) = 1_{\{(x,y): a \leq x \leq y \leq b\}}$. Then we can observe that for any diagram X we have $\beta(X) = \sum_{x \in X} f(x)$.

Since rank functions are linear representations we can apply Theorem 6.3. We automatically get the requirement that $p = 1$. We will next show that we will also need $q = 1$ through a counterexample.

Let $x_1 \leq x_2 \leq y$. Then

$$\|f(x_1, y) - f(x_2, y)\|_q^q = \int_{x_1}^{x_2} \int_t^y e^{t-s} ds\, dt = (x_2 - x_1) - e^{x_2 - y} + e^{x_1 - y}$$

which goes to $x_2 - x_1$ as y goes to ∞. For

$$\|f(x_1, y_1) - f(x_2, y_2)\|_q \leq C\|(x_1, y) - (x_2, y)\|_1 = C(x_2 - x_1)$$

for all $x_1 < x_2 < y$ we need $(x_2 - x_1) < C^q(x_1 - x_2)^q$ for all x_1, x_2 which will only hold if $q = 1$.

All that remains to be shown is that for $f : \mathbb{R}^{2+} \to \mathcal{B}$ by $f(a,b) = 1_{\{(x,y): a \leq x \leq y \leq b\}}$ we have

$$\|f(x_1, y_1) - f(x_2, y_2)\|_1 \leq |x_1 - x_2| + |y_1 - y_2|.$$

Without loss of generality assume $x_1 \leq x_2$. If $x_2 \leq y_1$ then

$$\|f(x_1, y_1) - f(x_2, y_2)\|_1 \leq \|f(x_1, y_1) - f(x_2, y_1)\|_1 + \|f(x_2, y_1) - f(x_2, y_2)\|_1.$$

Using the integrals above we see that $\|f(x_1, y_1) - f(x_2, y_1)\|_1 \leq |x_2 - x_1|$ and analogously that $\|f(x_2, y_1) - f(x_2, y_2)\|_1 \leq |y_2 - y_1|$. Together they imply that $\|f(x_1, y_1) - f(x_2, y_2)\|_1 \leq \|(x_1, y_1) - (x_2, y_2)\|_1$.

If $x_2 > y_1$ then the supports of $f(x_1, y_1)$ and $f(x_2, y_2)$ are disjoint. Routine calculations show that $\|f(x, y)\|_1 \leq |y - x|$. In this scenario, $\|(x_1, y_1) - (x_2, y_2)\| \geq |y_1 - x_1| + |y_2 - x_2|$ and hence $\|f(x_1, y_1) - f(x_2, y_2)\|_1 \leq \|(x_1, y_1) - (x_2, y_2)\|_1$. □

6.2. **Persistence Landscapes are not Lipschitz stable.** Landscapes [6] were among the first functionals proposed for persistence diagrams and remain among the most popular in practice.

Definition 6.6. *The* persistence landscape *of persistence module M is the function $\lambda : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ defined by*

$$\lambda(k, t)(M) = \sup\{h \geq 0 \mid \mathbf{rk}(M(t - h \leq t + h)) \geq k).$$

We call $\lambda(k, \cdot)(M)$ the k-th persistence landscape.

The L^q distance between persistence landscapes is defined as the sum over k of the L^q distances between the k-th persistence landscape. Let $pl_k(f)$ denote the k-th persistence landscape for sublevel persistence diagram for f.

Unlike the linear functionals in the previous subsection, there is no Lipschitz nor even Hölder stability with respect to the p-Wasserstein distances of their corresponding persistence diagrams.

Theorem 6.7. *Let $(\mathcal{D}, W_p)$ denote the space of persistence diagrams with the W_p metric and let (PL, L_q) denote the space of persistence landscapes with the L^q metric. For all $q \in [1, \infty)$, the function pl : $(\mathcal{D}, W_p) \to (PL, L^q)$ which sends each persistence diagram to its corresponding persistence landscape is* not *Hölder continuous.*

Proof. Let X and Y be the persistence diagrams with one off-diagonal point at $(0, a)$ and $(0, a - r)$ respectively, where $r \ll a$. The first persistence landscapes for $pl(X)$ and $pl(Y)$ are both a triangle function. These are centred at $a/2$ and $(a - r)/2$ respectively. We can compute that $pl(X) - pl(Y)$ is a trapezium shape:

$$(pl(X) - pl(Y))(t) \begin{cases} = t \text{ for } 2t \in [(a-r)/2, a/2] \\ = r \text{ for } t \in [a/2, a - r] \\ = a - t \text{ for } t \in [a - r, a] \\ = 0 \text{ otherwise} \end{cases}$$

When $a \gg r$, the contribution of the integral over $[a/2, a - r]$ will dominate the L^q distance between $pl(X)$ and $pl(Y)$. The function distance is bounded below by

$$\|pl(X) - pl(Y)\|_q > \left(\int_{a/2, a-r} r^q \, dt \right)^{1/q} = r(a/2 - r)^{1/q}$$

We also know that for $r \ll a$ the optimal matching between X and Y sends the point at $(0, a)$ to $(0, a-r)$ and hence $W_p(X,Y) = r$ for all $p \in [1,\infty]$. For a Hölder stability result to hold we would need there to be $\alpha, C > 0$ such that $\|pl(X) - pl(Y)\|_q \leq CW_p(X,Y)^\alpha$ for all $X, Y \in \mathcal{D}$. This would imply

$$r(a/2 - r)^{1/q} \leq Cr^\alpha \qquad \forall a \gg r.$$

By setting r small and a large we can make the left hand side arbitrarily large and the right hand side arbitrarily small which provides a contradiction regardless of the choice of q, C and α. This means there cannot be any Hölder continuity when $q \neq \infty$. □

Porism 6.8. *Let M be a simplicial complex containing at least one edge. Let (X, L^p) denote the space of monotone functions over M with the L^p metric. For all $p, q \in [1,\infty)$, the function $PL : (X, d_{L^p}) \to (pl, L^q)$ which sends each function to the persistence landscape of its sublevel set filtration is* not *Hölder continuous.*

Proof. We prove by creating an example of a pair of function that produce the persistence diagrams in Theorem 6.7. Fix an edge $[x_1, x_2]$ in M. Set $f([x_1]) = 0$, $f([x_2]) = 0$, $g([x_1, x_2]) = a - r$ and $g(\tau) = a$ for all other cells $\tau \in M$. Note that $\|f - g\|_p = r$ for all $r \in [0, a]$. The persistence diagram of the sublevel set filtrations of f and g are the X and Y used in the proof of Theorem 6.7. The remainder of the proof are the same inequalities as before. □

Remark 6.9. *It is worth noting that* [6] *does have a limited version of Wasserstein stability using* [19]. *This corollary states that for X a triangulable, compact metric space that implies bounded degree-k total persistence for some real number $k \geq 1$, and f, g two tame Lipschitz functions we have*

$$\|PL(f) - PL(g)\|_p \leq C \|f - g\|_\infty^{\frac{p-k}{p}}$$

for all $p \geq k$, where

$$C = C_{X,k}\|f\|_\infty(Lip(f)^k + Lip(g)^k) + C_{X,k+1}\frac{1}{p+1}(Lip(f)^{k+1} + Lip(g)^{k+1}).$$

See Section 3 for some limitations in terms of k and $C_{X,k}$.

7. Algebraic Wasserstein Stability

In this section we give an approach to Wasserstein stability at the algebraic level. For bottleneck distance this has yielded important insights, and this new approach opens the possibility to studying Wasserstein stability without an underlying chain complex as well as relaxing the finiteness conditions. There are two significant obstructions. A key tool in understanding bottleneck stability is the concept of interleaving between persistence modules. The corresponding interleaving distance naturally corresponds to a sup-norm as morphisms must be defined over the whole space, and also the representation of an interleaving by a single morphism [2] cannot be used since its effect on Wasserstein distance cannot be controlled. Instead, we construct an alternative intermediate object which interpolates between the two persistence modules without reference to the underlying chain complex and filtration. The section is organized into four parts: in the first part, we give our definition of algebraic distance and show that it is equivalent to the diagram distance. We then prove an additional bound on extensions of persistence modules which may be of independent interest, we rederive the cellular stability result in the algebraic setting, i.e. two functions over one chain complex (not necessarily finite). Finally, we discuss how the algebraic distance could be defined in other contexts. For convenience we recall the following set up from Section 2.

- We are concerned with the category of p.f.d. persistence modules with a common parameterization.
- We only consider ungraded morphisms between persistence modules, i.e. given a morphism $f : \mathcal{A} \to \mathcal{B}$, for any $x \in \mathcal{A}_\alpha$, $x \mapsto f(x) \in \mathcal{B}_\alpha$, where x denotes a basis element.
- For readability, we mention freely generated submodules, i.e. infinite intervals, only where necessary to avoid confusion. As p.f.d. modules have decompositions, the modules may be split into summands consisting of either only finite or only infinite intervals, these two cases can be dealt with separately, or the resulting distance is infinite and the desired results hold trivially.

Throughout this section we make extensive use of the algebraic properties of persistence modules as well as standard constructions such as short exact sequences. For a more in-depth description of the algebraic structure of persistence modules see [2] and for a discussion of short exact sequences of persistence modules, see [28]. Turning to the Wasserstein distance, recall that the definition for persistence diagrams is defined as

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) = \inf_{\mathbf{M}} \left(\sum_{x \in \mathrm{Dgm}(\mathcal{A})} |\mathbf{b}(x) - \mathbf{b}(\mathbf{M}(x))|^p + |\mathbf{d}(x) - \mathbf{d}(\mathbf{M}(x))|^p) \right)^{\frac{1}{p}}, \tag{4}$$

Each matching gives an *transport plan*, moving points from $\mathrm{Dgm}(\mathcal{A})$ to $\mathrm{Dgm}(\mathcal{B})$ or to the diagonal (or from the diagonal to $\mathrm{Dgm}(\mathcal{B})$). To avoid confusion, we will refer to the above as the diagram or matching distance, in contrast to the algebraic distance. A matching that achieves the infimum is called an *optimal transport plan*.

Remark 7.1. *Note that a distance of zero does not imply isomorphism of modules, as points on the diagonal do not contribute to the norm as they are zero length. Furthermore, an infinite distance is also possible. Hence, the p-Wasserstein distance is an extended pseudo-distance on the set of persistence modules. It can be made into an extended distance by considering the observable category of persistence modules* [12]. *We refer the reader to* [9] *for a more complete account of different variations of persistence modules. For readability, we refer to W_p simply as a distance.*

Our goal is to define an algebraic distance which yields the same distance in the case of interval decompositions. Our approach is similar in spirit to the single morphism characterization of interleaving from [2]. Unfortunately, a matching is generally not realizable as an ungraded morphism. This is overcome in [2] by employing a shift operator which ensures that a morphism exists. This approach cannot be applied in our setting as the shift operator will incur a transport cost which is proportional to the number of points in the diagram. So rather than look for a single morphism, we look for an intermediate object with morphisms to the modules we are comparing.

Definition 7.2. *A* interpolating object *of two persistence modules $\mathcal{A}$ and $\mathcal{B}$ is a triple $(\mathcal{C}, \varphi, \psi)$ in a diagram*

$$\begin{tikzcd} \mathcal{C} \arrow[r, "\varphi"] \arrow[d, "\psi"] & \mathcal{A} \\ \mathcal{B} & \end{tikzcd}$$

This idea has appeared in the literature several times as classical interleaving can be interpreted as a span (e.g. [13]). There exists a dual notion with morphisms from $\mathcal{A}$ and $\mathcal{B}$ to a third module (cospan).

Definition 7.3. *We define the zero module, denoted* $\mathbf{0}$, *as a module which has no intervals of positive length in its decomposition.*

Remark 7.4. *Note that* $\mathbf{0}$ *is not unique up to isomorphism, however, all such modules are at a Wasserstein distance of 0.*

We recall the definition of the p-norm of a persistence module (Definition 2.7).

$$||\mathcal{A}||_p = \left(\sum_{x \in \mathrm{Dgm}(\mathcal{A})} \ell(x)^p \right)^{\frac{1}{p}}$$

This is directly related to the p-Wasserstein diagram distance to the trivial or $\mathbf{0}$ persistence module. The transport distance to the diagonal matches to the midpoint of each bar for all $p \geq 1$. This introduces a multiplicative factor when relating the two.

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathbf{0}) = 2^{\frac{1-p}{p}} ||\mathcal{A}||_p$$

We will phrase the results in terms of $||\mathcal{A}||_p$ to avoid extra constants in the expressions.

Remark 7.5. *In the above definition, if a module has any infinite intervals then they have an infinite norm. Our statements still hold in these cases, as it may happen modules with infinite norms may be have a finite distance.*

We now define the main quantity which we study in section.

Definition 7.6. *Let the $(\mathcal{C}, \varphi, \psi)$ be an interpolating object between two persistence modules $\mathcal{A}$ and $\mathcal{B}$. Given a p-norm on modules, we define the* ***transport cost of the interpolating object*** *to be*

$$c(\varphi, \psi) = ||\ker\varphi \oplus \operatorname{coker}\varphi \oplus \ker\psi \oplus \operatorname{coker}\psi||_p$$

We may now define the algebraic Wasserstein distance:

Definition 7.7. *The algebraic p-Wasserstein distance between persistence modules is*

$$W_p^{\mathrm{alg}}(\mathcal{A},\mathcal{B}) = \inf_{(\mathcal{C},\varphi,\psi)} c(\varphi,\psi)$$

Our main result here will be to show that in the case of p.f.d. modules, the algebraic and diagram Wasserstein distances agree. It is non-trivial that the above actually defines a distance. In the case of p.f.d. modules, the equivalence we prove implies this result. However, in Section 7.3, we provide sufficient algebraic conditions for $W_p^{\mathrm{alg}}(\mathcal{A},\mathcal{B})$ to define a distance.

We begin with the following qualitative result – note that this holds for arbitrary p.f.d. modules with finite p-norm.

Lemma 7.8. *If $\mathcal{A} \hookrightarrow \mathcal{B}$ and dually $\mathcal{B} \twoheadrightarrow \mathcal{A}$ implies*

$$||\mathcal{A}||_p \leq ||\mathcal{B}||_p$$

Proof. Recall that $\mathcal{A} \overset{f}{\hookrightarrow} \mathcal{B}$ implies that there is an injective set map $f_\star$ from the intervals of $\mathcal{A}$ to the intervals of $\mathcal{B}$.

$$\begin{aligned}
||\mathcal{A}||_p^p &= \sum_{x\in \mathrm{Dgm}(\mathcal{A})} (\mathbf{d}(x) - \mathbf{b}(x)))^p \\
&\leq \sum_{x\in \mathrm{Dgm}(\mathcal{A})} (\mathbf{d}(f_\star(x)) - \mathbf{b}(f_\star(x)))^p \\
&\leq \sum_{x\in \mathrm{im}\, f_\star} (\mathbf{d}(f_\star(x)) - \mathbf{b}(f_\star(x)))^p + \sum_{x'\in \mathcal{B}-\mathrm{im}\, f_\star} (\mathbf{d}(x') - \mathbf{b}(x'))^p \\
&= ||\mathcal{B}||_p^p
\end{aligned}$$

Every interval in $\mathcal{A}$ maps to an interval which is has the same death time but must have an equal or earlier birth time (the first inequality). This follows from the fact that it f is a homomorphism of persistence modules. Hence, each term in the summation for $\mathcal{A}$ is dominated by the corresponding term in $\mathcal{B}$, implying the inequality. The proof for $\mathcal{B} \twoheadrightarrow \mathcal{A}$ is similar as there is an injective set map from the intervals $\mathcal{A}$ to the intervals of $\mathcal{B}$. ☐

To obtain quantitative bounds, we use the fact that we can consider *short exact sequences as transport plans* – which allows us to avoid mapping back to the underlying space, i.e. a finite CW complex. We begin by defining two interpolations from a persistence module $\mathcal{A}$ to $\mathbf{0}$: one which sends the births to the deaths and the other sends deaths to births. Let $\mathcal{A}$ be p.f.d. persistence module with the decomposition $\bigoplus_x \mathbb{I}\{\mathbf{b}(x), \mathbf{d}(x)\}$. We omit reference to the interval types as valid choices can easily be determined.

Definition 7.9. *For $t \in [0,1]$, define*

$$\mathcal{A}_t = \bigoplus_{x\in \mathrm{Dgm}(\mathcal{A})} \mathbb{I}\{\mathbf{b}(x), t\mathbf{b}(x) + (1-t)(\mathbf{d}(x))\}$$

We refer this to as the death-birth interpolation.

Definition 7.10. *For $t \in [0,1]$, define*

$$\mathcal{A}_t = \bigoplus_{x\in \mathrm{Dgm}(\mathcal{A})} \mathbb{I}\{(1-t)\mathbf{b}(x) + t\mathbf{d}(x), \mathbf{d}(x)\}$$

We refer this to as the birth-death interpolation.

In both cases, $\mathcal{A}_0 \cong \mathcal{A}$ by construction and the diagram distance between $\mathcal{A}_1$ and $\mathbf{0}$ is 0 since $\mathcal{A}_1$ only has points on the diagonal (ephemeral classes). We abuse notation and write $\mathcal{A}_1 \cong \mathbf{0}$. In the death-birth interpolation, there is a map $\mathcal{A}_s \twoheadrightarrow \mathcal{A}_t$ for $s < t$ and in the birth-death interpolation, there is a map $\mathcal{A}_t \hookrightarrow \mathcal{A}_s$ for $s < t$.

Observation 7.11. *Constructing a death-birth or birth-death interpolation requires that the module has finite norm. Hence, we cannot interpolate an infinite interval to 0. While somewhat counter-intuitive, a simple calculation shows that the distance between two freely generated modules, i.e. with infinite interval summands can be finite. However, in this case, the module which we interpolate to 0 must consist of only finite intervals.*

We first show that the above interpolation can be pushed forward or pulled back via monomorphisms and epimorphisms respectively. These are based on standard constructions [43] and their properties. Some of the statements will be obvious to experts, but we include for completeness.

Lemma 7.12. *Given a monomorphism $\varphi : \mathcal{A} \hookrightarrow \mathcal{B}$ and the death-birth interpolation on $\mathcal{A}$, for any $t \in [0, 1]$, there exists $\mathcal{B}_t$ such that the following commutative diagram exists*

$$
\begin{tikzcd}
\mathcal{A} \arrow[r, hook] \arrow[d, two heads] & \mathcal{B} \arrow[d, two heads] \\
\mathcal{A}_t \arrow[r, hook] \arrow[d, two heads] & \mathcal{B}_t \arrow[d, two heads] \\
0 \arrow[r, hook] & \operatorname{coker} \varphi
\end{tikzcd}
$$

Furthermore, if there exists a short exact sequence

$$0 \to \mathcal{A} \hookrightarrow \mathcal{B} \twoheadrightarrow \mathcal{C} \to 0,$$

there exists a short exact sequence

$$0 \to \mathcal{A}_t \hookrightarrow \mathcal{B}_t \twoheadrightarrow \mathcal{C} \to 0$$

Proof. First, we define $\mathcal{B}_t$ via the pushout

$$
\begin{tikzcd}
\mathcal{A} \arrow[r, hook, "\varphi"] \arrow[d, two heads, "\psi"] & \mathcal{B} \arrow[d, two heads, "j"] \\
\mathcal{A}_t \arrow[r, hook, "i"] & \mathcal{B}_t
\end{tikzcd}
$$

We observe the outer rectangle commutes since $\varphi(\mathcal{A}) \to \operatorname{coker} \varphi$ is trivial by definition. The upper square commutes by the construction of the pushout and the lower square commutes by the universality of the colimit. Since pushouts preserve epimorphisms, j is epic. To show that i is injective, and hence, monic, or any element $\alpha \in \mathcal{A}_t$, there exists a $\gamma \in \mathcal{A}$ such that $\psi(\gamma) = \alpha$ and since φ is monic, $\varphi(\gamma) \neq 0$. Hence for each α, the image of i is $(\alpha \oplus \varphi(\gamma))/(\alpha \sim \varphi(\gamma)) \neq 0$. Through a diagram chase, it can directly be verified that for any two non-trivial elements in $\mathcal{A}_t$, $\alpha \neq \alpha'$, they do not map to the same equivalence class, showing that i is injective. To prove the second part of the lemma, we extend the pushout diagram to $\mathcal{C}$.

$$
\begin{tikzcd}
\mathcal{A} \arrow[r, hook, "\varphi"] \arrow[d, two heads, "\psi"] & \mathcal{B} \arrow[d, two heads, "j"] \arrow[ddr, two heads, bend left, "\zeta"] & \\
\mathcal{A}_t \arrow[r, hook, "i"] \arrow[drr, bend right, "0"'] & \mathcal{B}_t \arrow[dr, dashed, two heads, "h"] & \\
 & & \mathcal{C}
\end{tikzcd}
$$

Since $\zeta \circ \varphi = 0$ by exactness, the outer diagram commutes and by the universality of push outs h exists and is unique. The fact that h is epic follows from $\zeta = h \circ j$, and ζ is epic by assumption. The inclusion $\operatorname{im} i \subseteq \ker h$ follows by commutativity ($h \circ i = 0$). To show $\ker h \subseteq \operatorname{im} i$, choose a lift $[x]$ of an element in $\ker h$ through j. By commutativity $\zeta([x]) = 0$, hence by exactness $[x] \in \operatorname{im} \phi$. By commutativity, $j([x])$ must be contained in $\operatorname{im} i$, completing the proof. $\square$

Lemma 7.13. *Given an epimorphism $\varphi : \mathcal{A} \twoheadrightarrow \mathcal{B}$ and the birth-death interpolation on $\mathcal{B}$, for any $t \in [0, 1]$, there exists a $\mathcal{A}_t$ such that the following commutative diagram exists*

$$
\begin{tikzcd}
\ker \varphi \arrow[r, two heads] \arrow[d, hook] & 0 \arrow[d, hook] \\
\mathcal{A}_t \arrow[r, two heads] \arrow[d, hook] & \mathcal{B}_t \arrow[d, hook] \\
\mathcal{A} \arrow[r, two heads] & \mathcal{B}
\end{tikzcd}
$$

Furthermore, if there exists a short exact sequence

$$0 \to \mathcal{C} \hookrightarrow \mathcal{A} \twoheadrightarrow \mathcal{B} \to 0,$$

there exists a short exact sequence

$$0 \to \mathcal{C} \hookrightarrow \mathcal{A}_t \twoheadrightarrow \mathcal{B}_t \to 0$$

Proof. First, we define $\mathcal{A}_t$ via the pullback

$$\begin{tikzcd} \mathcal{A}_t \arrow[r, two heads, "i"] \arrow[d, hook, "j"] & \mathcal{B}_t \arrow[d, hook, "\psi"] \\ \mathcal{A} \arrow[r, two heads, "\varphi"] & \mathcal{B} \end{tikzcd}$$

We observe the outer rectangle commutes since $\ker \varphi \to \mathcal{A}$ is trivial by definition. The lower square commutes by the construction of the pullback, and the upper square commutes by the universality of the pullback. Since pullbacks preserve monomorphisms, j is a monomorphism. To show that i is epic, we show that it is a surjection. For any element $\alpha \in \mathcal{B}_t$, $\psi(\alpha)$ is non-trivial. Since φ is epic, there exists $\gamma \in \mathcal{A}$ such that $\varphi(\gamma) = \psi(\alpha)$. Hence, there exists a non-trivial element $(\gamma, \alpha) \in \mathcal{A}_t$, hence the map is surjective. The proof of the second part of the lemma proceeds as similarly to the proof of Lemma 7.12. Consider the diagram

$$\begin{tikzcd} C \arrow[dr, dashed, "h"] \arrow[drr, bend left, "0"] \arrow[ddr, bend right, "\zeta"'] & & \\ & \mathcal{A}_t \arrow[r, two heads, "i"] \arrow[d, hook, "j"] & \mathcal{B}_t \arrow[d, hook, "\psi"] \\ & \mathcal{A} \arrow[r, two heads, "\varphi"] & \mathcal{B} \end{tikzcd}$$

By exactness $\varphi \circ \zeta = 0$, hence the outer diagram commutes and by the universality of pullbacks, h exists and is unique. Furthermore, by commutativity, $\zeta = j \circ h$, so h is monic. $\operatorname{im} h \subseteq \ker i$ follows from the fact that commutativity implies that $i \circ h = 0$. To show $\ker i \subseteq \operatorname{im} h$, consider $x \in \ker i$. Since j is monic, $j(x) \neq 0$, and $\varphi \circ j(x) = 0$ by commutativity. Hence $j(x) \in \operatorname{im} \zeta$ and so we conclude $x \in \operatorname{im} h$. □

Based on these interpolations, we mirror the proofs in Section 4 to relate the algebraic and matching distances. We use the fact that the norm of the module is equal to the corresponding distance induced by interpolating the module to the **0** module. In the proofs, we make extensive use of the fact that monomorphisms map death times to death times and epimorphisms map birth times to birth times [2, Theorem 4.2]. Hence, there always exists a matching between indecomposables and this matching is unique if the death times (respectively birth times) are unique.

Lemma 7.14. *Let $\mathcal{A}$ be a module consisting of a single non-zero summand $\{b, d\}$. Given a monomorphism $\mathcal{A} \overset{\varphi}{\hookrightarrow} \mathcal{B}$ and a death-birth interpolation or dually $\mathcal{B} \overset{\psi}{\twoheadrightarrow} \mathcal{A}$ and a birth-death interpolation. There exists a decomposition of $[0,1]$ into a finite number of disjoint open intervals where*

(1) the closure of the intervals cover $[0,1]$,

(2) For any $t \in [0,1]$, φ_t (or ψ_t) within each interval the induced matchings $(\varphi_t)_\star$ (or $(\psi_t^\star)$ are constant.

Proof. We first prove the case of a monomorphism – choosing the interval as a basis for $\mathcal{A}$, we apply φ and obtain the image which is again a bar $\{b, d\}$. This is bar is a linear combination of generators in $\mathcal{B}$ (the basis we use in $\mathcal{B}$). The main claim is that this linear combination of generators is finite. This follows from the p.f.d. assumption. For any $s \in (b, d)$, there corresponding vector space is finite dimensional. This follows since all birth times must occur at or before b by the properties of persistence module homomorphisms and all the death times at or before d but after b. If this is not the case the morphism does not commute with the internal morphisms of the module. Since all births must come before the first death, it follows that if linear combination were infinite there would exist an $s \in (b, d)$ where $\mathcal{B}_s$ would be infinite dimensional contradicting the p.f.d. assumption. To complete the proof, we observe that as we change the death time, only this linear combination can be affected (one can directly verify this in the construction), which implies that φ_t is only non-constant when d has equality with a finite number of the death times.

The proof for the epimorphism is similar. We claim that there is a finite number of bars which map to $a \in \mathcal{A}$. Formally we say that the space generated by the set $\{g_i\}_i$ as such that $\psi(g_i) = a$ is finite dimensional. By the same argument as above we can deduce that the deaths must occur at or after d and all the births at or after b, implying that in the corresponding bars the births must come before the first death (which is d) since ψ is a persistence homomorphism. □

Lemma 7.15. *Given an epimorphism between p.f.d. persistence modules $f : \mathcal{A} \twoheadrightarrow \mathcal{B}$ such that* $\ker f$ *is finitely generated,*

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) \leq ||\ker f||_p$$

Proof. Consider the following short exact sequence

$$0 \to \ker f \xhookrightarrow{i} \mathcal{A} \xrightarrow{f} \mathcal{B} \to 0$$

It simplifies the argument to add an arbitrarily small positive perturbation to the death times in $\mathcal{A}$ that appear in $\ker f$, so we can assume that the death times in $\ker f$ are unique. At each point, we only consider classes in $\ker f$ which are not on the diagonal. For brevity, we also refer to the indecomposables as bars (rather than intervals or indecomposables).

We consider a death-birth interpolation on $\ker f$. Let $f_t : \mathcal{A}_t \to \mathcal{B}$ as in Lemma 7.12 for $t = [0, 1]$. By [2, Theorem 4.2], there exists an injective set map between the summands of $\ker f_t$ and $\mathcal{A}_t$. Denote this map ϕ_t. The proof mirrors the proof of the cellular stability theorem. Consider an interval $t \in (a_i, a_{i+1})$, such that the ordering of the death times in $\mathcal{A}_t$ does not change between a_i and a_{i+1}. This implies that ϕ_t is constant in (a_i, a_{i+1}). As only the death times in the image of ϕ change from $\mathcal{A}_{a_i}$ to $\mathcal{A}_{a_{i+1}}$,

$$\begin{aligned} W_p(\mathrm{Dgm}(\mathcal{A}_{a_i}), \mathrm{Dgm}(\mathcal{A}_{a_i+1}))^p &\leq \sum_{x \in \mathrm{Dgm}(\ker f_t)} (\mathbf{d}(\phi_t(x)^{a_i}) - \mathbf{d}(\phi_t(x)^{a_{i+1}}))^p \\ &= W_p(\mathrm{Dgm}(\ker f_{a_i}), \mathrm{Dgm}(\ker f_{a_{i+1}}))^p \end{aligned}$$

The death time for each bar is linear as a function of t in the death-birth interpolation on $\ker f$. Therefore, we can divide $[0, 1]$ into intervals where the ordering of the death times is consistent. This is the same argument as in the cellular case. Now we note that the interpolation on $\ker f$ is linear so we can consider intervals where the ordering is consistent.

Assuming there are finitely many such intervals which cover $[0, 1]$ except at a finite number of times, where we have equality of death times. These can be covered taking one-sided limits. Hence summing up over the intervals, just as in Theorem 4.8, we obtain an upper bound of

$$W_p(\mathrm{Dgm}(\ker f_0), \mathrm{Dgm}(\ker f_1))^p = ||\ker f||_p^p.$$

Since $W_p(\mathrm{Dgm}(\mathcal{A}_1), \mathrm{Dgm}(\mathcal{B})) = 0$, it follows that Wasserstein distance can only be smaller, yielding the result.

If $\mathcal{A}$ is finitely generated, the space of set death times is finite and so the existence of the decomposition of $[0, 1]$ into pieces where ϕ_i is constant is immediate. If $\mathcal{A}$ is p.f.d., we are only guaranteed to have a countable number of death times, however we will show it is sufficient for $\ker f$ to be finitely generated. This follows from Lemma 7.14. Since each bar may only introduce a finite number of non-constant points and there are a finite number of bars in $\ker f$ (by the assumption it is finitely generated), the result follows.

□

There are two subtleties which we point out in the proof above:

(1) If $\mathcal{A}$ is freely generated, the epimorphism is either an isomorphism or the distance is infinite. From the perspective of persistence diagram this is correct, as the epimorphism indicates that the birth times are the same, but the fact that the morphism is not an isomorphism implies that the essential classes are not the same and so an infinite distance is appropriate.

(2) It is insufficient in the above proof for $\ker f$ to be p.f.d. If $\ker f$ contains a countable rather than finite number of bars (finitely generated), it is straightforward to construct examples where open intervals in $[0, 1]$ where ϕ_t is constant do not exist, even though each bar only has a finite number of "pairing switches," the proof relies on doing the interpolation the entire module at once – hence there may still be an infinite number of "pairing switches." We remove the requirement of $\ker f$ to be finitely generated in Lemma 7.21.

We also have the dual statement.

Lemma 7.16. *Given a monomorphism between p.f.d. perisistence modules $f : \mathcal{A} \hookrightarrow \mathcal{B}$, where* $\operatorname{coker} f$ *is finitely generated,*

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) \leq ||\operatorname{coker} f||_p$$

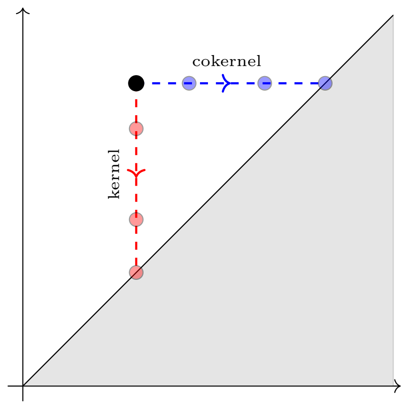


FIGURE 3. The transport plans corresponding to the kernel (red) and cokernel (blue) respectively. For the kernel interpolation to 0, we interpolate the death time to the birth time, and for the cokernel interpolation to 0, we interpolate the birth to the death time.

Proof. The proof is dual to Lemma 7.15 using a birth-death interpolation on $\operatorname{coker} f$. The proof of finiteness is also dual, with only a finite number of bars in $\mathcal{B}$ mapping to a single bar in $\operatorname{coker} f$, by the p.f.d. assumption. The remainder of the proof is the same with the only difference on equality of birth rather than death times. □

Notice that if $\mathcal{A}$ is free, the monomorphism implies that $\mathcal{B}$ must also be free. Furthermore, the cokernel precisely captures the difference in birth times.

Porism 7.17. *Given an monomorphism between freely generated p.f.d. persistence modules $f : \mathcal{A} \hookrightarrow \mathcal{B}$,*

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) \leq ||\operatorname{coker} f||_p$$

Proof. If the cokernel contains a infinite interval, then the norm is infinite and the statement holds trivially. If $||\operatorname{coker} f||_p$ is finite, then the linear interpolation in the proof of Lemma 7.16 may be applied even in the case of countably many intervals. This is because as $\mathcal{B}$ is freely generated, the induced matching remains consistent throughout the interpolation and so there need only be a countable, rather than finite number of intervals. □

Remark 7.18. *We do not explicitly deal with a module which contain both finite and infinite intervals. However, since we restrict to p.f.d. modules and a decomposition exists, we can study the matchings between only finite intervals or only infinite intervals. This is because while a morphism can induce a matching between an infinite interval and finite interval, the corresponding distance in this case is infinite.*

Remark 7.19. *Lemmas 7.15 and 7.16 depend critically on the induced matchings from epic and monic morphisms respectively. While we have chosen to be as explicit as possible in the proof, we note that the results on matchings follow from* [2, Theorem 5.7], *which shows that while the matchings are functorial in the subcategories where morphisms are either all monomorphisms or all epimorphisms. This enables us to relate the distance of the interpolation to the distance it induces.*

We now extend Lemmas 7.16 and 7.15 to remove the assumption of finitely generated kernel and cokernel respectively. We will show that any bounded p-energy p.f.d. module can be approximated by a finitely generated module. First, we define the matching p-distance induced by a monomorphism $\mathcal{A} \overset{f}{\hookrightarrow} \mathcal{B}$

$$W_p^f(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) = \left(\sum_{x \in \mathrm{Dgm}(\mathcal{A})} (\mathbf{b}(x) - \mathbf{b}(f_\star(x)))^p + \sum_{x' \in \mathrm{Dgm}(\mathcal{B}) - \operatorname{im} f_\star} \mathbf{d}(x')^p + \mathbf{b}(x')^p \right)^{1/p},$$

where $f_\star$ is the injective set map between intervals induced by f. Note that as f is a monomorphism the matched death times coincide. Dually given an epimorphism $\mathcal{B} \overset{g}{\twoheadrightarrow} \mathcal{A}$, the induced matching p-distance is

$$W_p^g(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) = \left(\sum_{x \in \mathrm{Dgm}(\mathcal{A})} (\mathbf{d}(g^\star(x)) - \mathbf{d}(x))^p + \sum_{x' \in \mathrm{Dgm}(\mathcal{B}) - \operatorname{im} g^\star} \mathbf{d}(x')^p + \mathbf{b}(x')^p \right)^{1/p},$$

where $g^\star$ is the injective set map induced by g (which we remind the reader goes in opposite direction to g).

Lemma 7.20. *Consider three p.f.d. persistence modules such that* $\mathcal{A} \overset{f}{\hookrightarrow} \mathcal{B} \overset{g}{\hookrightarrow} \mathcal{C}$ *or dually* $\mathcal{C} \overset{g}{\twoheadrightarrow} \mathcal{B} \overset{f}{\twoheadrightarrow} \mathcal{A}$.

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) \leq W_p^h(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{C})),$$

where $h = g_\star \circ f_\star$ *if* f *anf* g *are monic, and* $h = g^\star \circ f^\star$ *if* f *and* g *are epic.*

Proof. The proof is similar to Lemma 7.8. We prove the case of monic f and g as the epic is identical. As the Wasserstein distance is the infiumum over all matchings, the matching distance is an upper bound, i.e.

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B}) \leq W_p^f(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})$$

To complete the proof we show $W_p^f(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B}) \leq W_p^{g\circ f}(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})$.

$$\begin{aligned}\left(W_p^f(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B}))\right)^p &= \sum_{x \in \mathrm{Dgm}(\mathcal{A})} (\mathbf{b}(x) - \mathbf{b}(f_\star(x)))^p + \sum_{x' \in \mathrm{Dgm}(\mathcal{B}) - \mathrm{im}\, f_\star} \mathbf{d}(x')^p + \mathbf{b}(x')^p \\ &\leq \sum_{x \in \mathrm{Dgm}(\mathcal{A})} (\mathbf{b}(x) - \mathbf{b}(g_\star \circ f_\star(x)))^p + \sum_{x' \in \mathrm{Dgm}(\mathcal{C}) - \mathrm{im}\, g_\star \circ f_\star} \mathbf{d}(x')^p + \mathbf{b}(x')^p\end{aligned}$$

The inequality follows since for any point $x \in \mathcal{A}$, $\mathbf{b}(g_\star \circ f_\star(x)) \leq \mathbf{b}(f_\star(x))$ and the set of unmatched intervals is nondecreasing, i.e. $|\mathrm{Dgm}(\mathcal{C}) - \mathrm{im}\, g_\star \circ f_\star| \geq |\mathrm{Dgm}(\mathcal{B}) - \mathrm{im}\, f_\star|$ since f and g are monic. ☐

Lemma 7.21. *Given an epimorphism between p.f.d. persistence modules* $f : \mathcal{A} \twoheadrightarrow \mathcal{B}$,

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) \leq ||\ker f||_p$$

Proof. If $\ker f$ does not have bounded p-energy, the statement holds trivially as the upper bound is infinite. Hence we can assume $||\ker f||_p < \infty$. By Lemma 2.10, there exists a finitely generated approximation to $\ker f$ which we denote $\overline{\ker f}$ such that $W_p(\overline{\ker f}, \ker f) < \varepsilon$. There exists a commutative diagram

$$\begin{tikzcd} \ker f \arrow[r, hook, "i"] \arrow[d, two heads, "\varphi"] & \mathcal{A} \arrow[d, two heads, "\psi"] \\ \overline{\ker f} \arrow[r, "i'"] & \mathcal{A}' \end{tikzcd} \tag{5}$$

where $\mathcal{A}' = \mathcal{A}/i(\ker\varphi)$. This module can be described explicitly,

$$\mathcal{A}' = \bigoplus_{x \in \mathrm{Dgm}(\ker f) - \mathrm{im}\, \varphi^\star} \{\mathbf{b}(i_\star(x), \mathbf{b}(x)\} \oplus \bigoplus_{x' \in \mathrm{Dgm}(\overline{\ker f})} \{\mathbf{b}(i_\star \circ \varphi^\star(x')), \mathbf{d}(i_\star \circ \varphi^\star(x'))\}.$$

It shortens the intervals which we quotient out in obtaining the finitely generated approximation. From the explicit description, as the birth times are in bijection, twe can directly check that the p-matching distance $W_p^\psi(\mathcal{A}, \mathcal{A}') < \varepsilon$. However, there is no reason that $\mathcal{A}'$ should be compatible with f. Rather, as in describing the interpolation, we must take the pushout of φ and i, which we denote $\mathcal{A}''$. We use $\mathcal{A}'$ to bound the distance $W_p(\mathrm{Dgm}(\mathcal{A}, \mathcal{A}''))$. Consider the following commutative diagram

$$\begin{tikzcd} \ker f \arrow[r, hook, "i"] \arrow[d, two heads, "\varphi"] & \mathcal{A} \arrow[d, two heads, "j"] \arrow[ddr, bend left, two heads, "\psi"] & \\ \overline{\ker f} \arrow[r, hook] \arrow[drr, hook, bend right, "i'"'] & \mathcal{A}'' \arrow[dr, dashed, two heads, "h"] & \\ & & \mathcal{A}' \end{tikzcd} \tag{6}$$

The diagram commutes by universality. We note j is epic since pushouts preserve epimorphisms and h must be epic since $h \circ j = \psi$. By Lemma 7.20, $\mathcal{A} \twoheadrightarrow \mathcal{A}' \twoheadrightarrow \mathcal{A}''$ implies that

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{A}'') \leq W_p^\psi(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{A}') \leq \varepsilon.$$

$$\begin{aligned} W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B}) &\leq W_p(\mathrm{Dgm}(\mathcal{A}''), \mathrm{Dgm}(\mathcal{B}) + \varepsilon \\ &\leq ||\overline{\ker f}||_p + \varepsilon \\ &\leq ||\ker f||_p + 2\varepsilon \end{aligned}$$

The second inequality follows from Lemma 7.15 and the last inequality is by construction. Taking the limit of ε to 0 yields the result. $\square$

The dual statement follows similarly.

Lemma 7.22. *Given an monomorphism between p.f.d. persistence modules* $f : \mathcal{A} \hookrightarrow \mathcal{B}$*,*

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) \leq ||\operatorname{coker} f||_p$$

Proof. If $\operatorname{coker} f$ does not have bounded p-energy, the statement holds trivially as the upper bound is infinite. Hence we can assume $||\operatorname{coker} f||_p < \infty$. By Lemma 2.10, there exists a finitely generated approximation to $\operatorname{coker} f$ which we denote $\overline{\operatorname{coker} f}$ such that $W_p(\overline{\operatorname{coker} f}, \operatorname{coker} f) < \varepsilon$. There exists a commutative diagram

(7)
$$\begin{tikzcd} \mathcal{B}' \arrow[r, "j'"] \arrow[d, hook, "\varphi"] & \overline{\operatorname{coker} f} \arrow[d, hook, "\psi"] \\ \mathcal{B} \arrow[r, two heads, "j"] & \operatorname{coker} f \end{tikzcd}$$

We define $\mathcal{B}'$ explicitly:

$$\mathcal{B}' = \bigoplus_{x \in \mathrm{Dgm}(\operatorname{coker} f) - \operatorname{im} \psi_\star} \{\mathbf{d}(x), \mathbf{d}(j^\star(x))\} \oplus \bigoplus_{x' \in \mathrm{Dgm}(\overline{\operatorname{coker} f})} \{\mathbf{b}(j^\star \circ \psi_\star(x')), \mathbf{d}(j^\star \circ \psi_\star(x'))\}.$$

We can directly check that $\mathcal{B}'$ is a submodule of $\mathcal{B}$ since we shorten intervals by increasing the birth time and that the matching p-distance $W_p^{varphi}(\mathrm{Dgm}(\mathcal{B}, \mathcal{B}')) \leq \varepsilon$. Commutativity can also be directly verified. As before, to apply interpolation with respect to f, we must construct the pullback $\mathcal{B}''$.

(8)
$$\begin{tikzcd} \mathcal{B}' \arrow[rrd, bend left, "j'"] \arrow[rd, hook, dashed, "h"] \arrow[rdd, bend right, "\varphi"'] & & \\ & \mathcal{B}'' \arrow[r, two heads] \arrow[d, hook, "i"] & \overline{\operatorname{coker} f} \arrow[d, hook, "\psi"] \\ & \mathcal{B} \arrow[r, two heads, "j"] & \operatorname{coker} f \end{tikzcd}$$

Commutativity follows from universality, and $i \circ h = \varphi$ implies that h is monic. Since $\mathcal{B}' \hookrightarrow \mathcal{B}'' \hookrightarrow \mathcal{B}$, by Lemma 7.20 it follows that $W_p(\mathrm{Dgm}(\mathcal{B}), \mathrm{Dgm}(\mathcal{B}'') \leq \varepsilon$. Since $\overline{\operatorname{coker} f}$ is finitely generated we may apply Lemma 7.16 to

$$0 \to \mathcal{A} \hookrightarrow \mathcal{B}'' \twoheadrightarrow \overline{\operatorname{coker} f} \to 0$$

$$\begin{aligned} W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B}) &\leq W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B}'') + \varepsilon \\ &\leq ||\overline{\operatorname{coker} f}||_p + \varepsilon \\ &\leq ||\operatorname{coker} f||_p + 2\varepsilon \end{aligned}$$

$\square$

We combine the above results to obtain statements about short exact sequences.

Theorem 7.23. *Given a short exact sequence of p.f.d. persistence modules*

$$0 \to \mathcal{A} \overset{\varphi}{\hookrightarrow} \mathcal{B} \overset{\psi}{\twoheadrightarrow} \mathcal{C} \to 0$$

then

(i) $W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) \leq ||\mathcal{C}||_p$ $\qquad$ (ii) $W_p(\mathrm{Dgm}(\mathcal{B}), \mathrm{Dgm}(\mathcal{C})) \leq ||\mathcal{A}||_p$

Proof. Given an monic map $\mathcal{A} \overset{\varphi}{\hookrightarrow} \mathcal{B}$, there is a short exact sequence

$$0 \to \mathcal{A} \overset{\varphi}{\hookrightarrow} \mathcal{B} \twoheadrightarrow \operatorname{coker} \varphi \to 0,$$

applying Lemma 7.22, we obtain (i). Given an epic map $\mathcal{B} \overset{\psi}{\twoheadrightarrow} \mathcal{C}$, there is a short exact sequence

$$0 \to \ker \psi \hookrightarrow \mathcal{B} \overset{\psi}{\twoheadrightarrow} \mathcal{C} \to 0$$

applying Lemma 7.21, we obtain (ii). $\square$

As an immediate corollary,

Corollary 7.24. *Given a morphism between persistence modules $f : \mathcal{A} \to \mathcal{B}$,*

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) \leq ||\ker f \oplus \mathrm{coker}\, f||_p$$

Proof. First note that for $p < \infty$,

$$||\ker f \oplus \mathrm{coker}\, f||_p^p = ||\ker f||_p^p + ||\mathrm{coker}\, f||_p^p$$

and for $p = \infty$

$$||\ker f \oplus \mathrm{coker}\, f||_\infty = \max\{||\ker f||_\infty, ||\mathrm{coker}\, f||_\infty\}$$

We construct the following two short exact sequences

$$0 \to \ker f \hookrightarrow \mathcal{A} \xrightarrow{f} \mathrm{im}\, f \to 0$$
$$0 \to \mathrm{im}\, f \hookrightarrow \mathcal{B} \twoheadrightarrow \mathrm{coker}\, f \to 0$$

Theorem 7.23, implies

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathrm{im}\, f))^p \leq ||\ker f||^p$$

$$W_p(\mathrm{Dgm}(\mathcal{B}), \mathrm{Dgm}(\mathrm{im}\, f))^p \leq ||\mathrm{coker}\, f||^p$$

We note that $||\ker f||^p$ bounds the change in birth times, i.e. the death times are unchanged and $||\mathrm{coker}\, f||^p$ bounds the change in death times, i.e. the birth times are unchanged. Comparing with Equation 4, we deduce the result. □

Remark 7.25. *Note that the weaker result*

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B})) \leq ||\ker f||_p + ||\mathrm{coker}\, f||_p$$

follows directly from the triangle inequality for the Wasserstein distance.

Example 7.26. *One is tempted to use the realization of a morphism between persistence modules as a matching between indecomposables in order realize the transport plan. Unfortunately, this is not the case. Consider*

$$\mathcal{A} \cong [x_2, x_4), \qquad \mathcal{B} \cong [x_1, x_4) \oplus [x_2, x_3), \qquad \mathcal{C} \cong [x_1, x_3),$$

with $x_1 < x_2 < x_3 < x_4$. Furthermore, let $[a]$ denote the lone summand in $\mathcal{A}$, $[c]$ the lone summand in $\mathcal{C}$, and $[b_1] = [x_1, x_4)$ and $[b_2] = [x_2, x_3)$. These three modules can be put into a short exact sequence

$$0 \to \mathcal{A} \xhookrightarrow{f} \mathcal{B} \xrightarrow{g} \mathcal{C} \to 0$$

with the maps given explicitly by

$$f([a]) \mapsto [b_1] + [b_2], \qquad g([b_1]) \mapsto [c], \qquad g([b_2]) \mapsto -[c]$$

The maps and the example described below are shown in Figure 4. Note that the image of the first morphism is $[x_2, x_4)$ and the image of the second is $[x_1, x_3)$ as required by injectivity and surjectivity of the short exact sequence. This is an example of a non-trivial extension, i.e. $\mathcal{B}$ is not isomorphic to $\mathcal{A} \oplus \mathcal{C}$.

In realizing the transport plan from $\mathcal{B}$ to $\mathcal{C}$, when interpolating from x_4 to x_3, the longer bar, i.e. $[x_1, x_4)$ becomes shorter. That is, let for any $t \in [x_3, x_4]$, the longer bar ($[b_1]_t$) in $\mathcal{B}_t$ is given by $[x_1, t)$. When $t = x_3$, there is a pairing switch and upon further interpolating $\mathcal{A}$ to 0, the shorter bar $[x_2, x_3)$ ($[b_2]$) becomes shorter. That is, for $t \in [x_2, x_3]$, $[b_2]_t$ is given by $[x_2, t)$ with $[b_1]_t$ unchanged.

One can view this as either a pairing switch in between death times, or that the pairing in $\mathcal{B}$ changes – by the elder rule the relation maps to the youngest generator and so maps to the summand born at x_2 rather than the one born at x_1. Note that this precisely mirrors the tracking which occurs in the cellular proof of stability.

We are now ready to prove our main result. Note that the case of $p = \infty$ is the stability result from [2]. We remark that the proof below can be modified to cover the case $p = \infty$ as well.

Theorem 7.27. *Given modules $\mathcal{A}$, $\mathcal{B}$ and an interpolating object $(\mathcal{C}, \varphi, \psi)$, for $p < \infty$*

$$W_p(\mathrm{Dgm}(\mathcal{A}), \mathrm{Dgm}(\mathcal{B}))^p \leq ||\ker \varphi||_p^p + ||\mathrm{coker}\, \varphi||_p^p + ||\ker \psi||_p^p + ||\mathrm{coker}\, \psi||_p^p$$

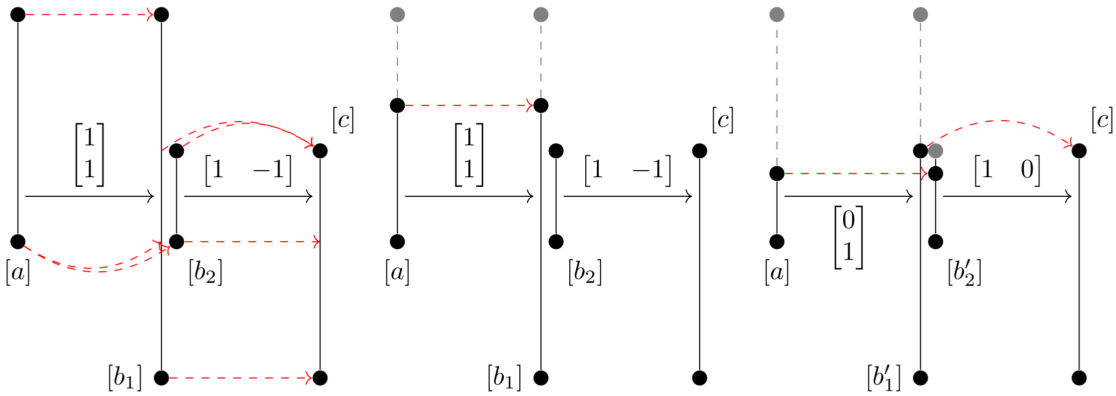


FIGURE 4. The left shows the example of where the morphisms do not directly correspond to the transport plan. In the middle is the first part of the interpolation which respects the morphism. In the second part shown on the right, the morphisms has changed. More accurately, a change of basis is required as the element in the image, $[b_1]+[b_2]$, is killed by the relation in the image, hence the pairing changes. The second morphism can then be constructed via the pushout construction described earlier in this section.

Proof. By Corollary 7.24,

$$W_p(\mathrm{Dgm}(\mathcal{C}), \mathrm{Dgm}(\mathcal{A}))^p \leq ||\ker \varphi||_p^p + ||\operatorname{coker} \varphi||_p^p$$

$$W_p(\mathrm{Dgm}(\mathcal{C}), \mathrm{Dgm}(\mathcal{B}))^p \leq ||\ker \psi||_p^p + ||\operatorname{coker} \psi||_p^p$$

Since we have the p-th power, it only remains to prove that we can add these to obtain the result. We observe that $\operatorname{coker}\varphi$ as a transport plan increases birth times from $\mathcal{A}$, while $\operatorname{coker}\psi$ decreases birth times of $\mathcal{C}$. By considering the short exact sequences with $\operatorname{im}\varphi$, we can infer a matching of birth and death times in $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$. That is, every birth time in $\mathcal{A}$ matches a birth time in $\operatorname{im}\varphi$ or a point on the diagonal. Every birth time in $\operatorname{im}\varphi$ matches a birth time in $\mathcal{C}$. Every birth time in $\mathcal{C}$ matches to a point on the diagonal or a birth time in $\operatorname{im}\psi$, which finally matches a birth time in $\mathcal{B}$. For any point which matches to the diagonal the inequality trivially holds. To bound the cost of the matching for the other points, consider any birth time in $\mathcal{A}$ and we note that $\mathbf{b}_{\mathcal{A}} \leq \mathbf{b}_{\mathcal{C}}$ and $\mathbf{b}_{\mathcal{B}} \leq \mathbf{b}_{\mathcal{C}}$. It follows that

$$|\mathbf{b}_{\mathcal{B}} - \mathbf{b}_{\mathcal{A}}|^p \leq |\mathbf{b}_{\mathcal{C}} - \mathbf{b}_{\mathcal{A}}|^p + |\mathbf{b}_{\mathcal{C}} - \mathbf{b}_{\mathcal{B}}|^p$$

The argument is similar for death times.

□

We now show that any matching produces an interpolating object with a matching algebraic distance. Again we consider $p < \infty$ to simplify the statement of the Theorem, but remark that the proof holds for this case as well.

Theorem 7.28. *Let $\mathcal{M}$ denote a matching between two diagrams corresponding to modules $\mathcal{A}$ and $\mathcal{B}$. For $p < \infty$, every matching $\mathcal{M}$ induces an interpolating object $(\mathcal{C}, \varphi, \psi)$ such that the p-Wasserstein transportation cost of the matching $\mathcal{M}$ is*

$$\sum_{(x,y)\in\mathcal{M}} |\mathbf{b}(x) - \mathbf{b}(y)|^p + |\mathbf{d}(x) - \mathbf{d}(y)|^p = ||\ker \varphi||_p^p + ||\operatorname{coker} \varphi||_p^p + ||\ker \psi||_p^p + ||\operatorname{coker} \psi||_p^p$$

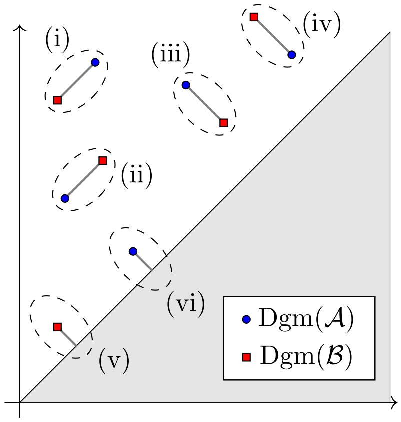
Proof. We show that for two persistence module $\mathcal{A}$ and $\mathcal{B}$, any matching between the corresponding persistence diagrams induces an interpolating object which induce the same distance. We do this by explicitly constructing $\mathcal{C}$.

We can assume that

$$\mathcal{A} = \bigoplus_{(\mathbf{b}_i,\mathbf{d}_i)\in \mathrm{Dgm}(\mathcal{A})} \mathbb{I}(\mathbf{b}_i, \mathbf{d}_i],$$

with half-open intervals. If not, the structure theorem of p.f.d. persistence modules and tells us that there is a morphism $\alpha : \oplus_{(\mathbf{b}_i,\mathbf{d}_i)\in \mathrm{Dgm}(\mathcal{A})}\mathbb{I}(\mathbf{b}_i, \mathbf{d}_i] \to \mathcal{A}$ such that $\ker\alpha$ and $\operatorname{coker}\alpha$ contains only intervals over single points which occur when the endpoints of the intervals in the decomposition differ. We can then cover these other interval shapes by composing ϕ with α. Similarly, we can assume $\mathcal{B} = \bigoplus_{(\mathbf{b}_i,\mathbf{d}_i)\in\mathrm{Dgm}(\mathcal{A})}\mathbb{I}(\mathbf{b}_i,\mathbf{d}_i]$.

There are six cases which cover all possible types of pairs within a matching as shown in the following persistence diagram. We observe that a morphism exists between points if the target point is below and to the left of the source point.



When we match a point $(\mathbf{b},\mathbf{d})$ to the diagonal we are effectively matching it to the point $\left(\frac{\mathbf{b}+\mathbf{d}}{2},\frac{\mathbf{b}+\mathbf{d}}{2}\right)$, which corresponds to an empty interval. We construct $\mathcal{C}$ by taking a direct sum over the pairs in the matching $\mathcal{M}$;

$$\mathcal{C} = \bigoplus_{((\mathbf{b}_{\mathcal{A}},\mathbf{d}_{\mathcal{A}}),(\mathbf{b}_{\mathcal{B}},\mathbf{d}_{\mathcal{B}}))\in\mathcal{M}} \mathbb{I}(\max\{\mathbf{b}_{\mathcal{A}},\mathbf{b}_{\mathcal{B}}\},\max\{\mathbf{d}_{\mathcal{A}},\mathbf{d}_{\mathcal{B}}\}].$$

For example, we contribute $\mathbb{I}(\mathbf{b}_{\mathcal{A}},\mathbf{d}_{\mathcal{A}}]$ in case (i) and contribute $\mathbb{I}(\frac{\mathbf{d}_{\mathcal{A}}+\mathbf{b}_{\mathcal{A}}}{2},\mathbf{d}_{\mathcal{A}}]$ in case (v). We consider the obvious morphisms $\phi:\mathcal{C}\to\mathcal{A}$ and $\psi:\mathcal{C}\to\mathcal{B}$ sending the generator of $\mathbb{I}(\max\{\mathbf{b}_{\mathcal{A}},\mathbf{b}_{\mathcal{B}}\},\max\{\mathbf{d}_{\mathcal{A}},\mathbf{d}_{\mathcal{B}}\}]$ to that of $\mathbb{I}(\mathbf{b}_{\mathcal{A}},\mathbf{d}_{\mathcal{A}}]$ and $\mathbb{I}(\mathbf{b}_{\mathcal{B}},\mathbf{d}_{\mathcal{B}}]$ respectively. The cokernels and kernels of ϕ and ψ are generated by the shifts in birth and deaths times respectively with

$$\ker\phi = \bigoplus_{\substack{((\mathbf{b}_{\mathcal{A}},\mathbf{d}_{\mathcal{A}}),(\mathbf{b}_{\mathcal{B}},\mathbf{d}_{\mathcal{B}}))\in\mathcal{M}\\ \mathbf{d}_{\mathcal{A}}<\mathbf{d}_{\mathcal{B}}}} \mathbb{I}(\mathbf{d}_{\mathcal{A}},\mathbf{d}_{\mathcal{B}}]$$

$$\ker\psi = \bigoplus_{\substack{((\mathbf{b}_{\mathcal{A}},\mathbf{d}_{\mathcal{A}}),(\mathbf{b}_{\mathcal{B}},\mathbf{d}_{\mathcal{B}}))\in\mathcal{M}\\ \mathbf{d}_{\mathcal{A}}>\mathbf{d}_{\mathcal{B}}}} \mathbb{I}(\mathbf{d}_{\mathcal{B}},\mathbf{d}_{\mathcal{A}}]$$

$$\operatorname{coker}\phi = \bigoplus_{\substack{((\mathbf{b}_{\mathcal{A}},\mathbf{d}_{\mathcal{A}}),(\mathbf{b}_{\mathcal{B}},\mathbf{d}_{\mathcal{B}}))\in\mathcal{M}\\ \mathbf{b}_{\mathcal{A}}<\mathbf{b}_{\mathcal{B}}}} \mathbb{I}(\mathbf{b}_{\mathcal{A}},\mathbf{b}_{\mathcal{B}}]$$

$$\operatorname{coker}\psi = \bigoplus_{\substack{((\mathbf{b}_{\mathcal{A}},\mathbf{d}_{\mathcal{A}}),(\mathbf{b}_{\mathcal{B}},\mathbf{d}_{\mathcal{B}}))\in\mathcal{M}\\ \mathbf{b}_{\mathcal{B}}<\mathbf{b}_{\mathcal{A}}}} \mathbb{I}(\mathbf{b}_{\mathcal{B}},\mathbf{b}_{\mathcal{A}}].$$

When we compute the sum of the p-th powers of the p-norms we see that

$$\|\ker\phi\|_p^p + \|\ker\psi\|_p^p + \|\operatorname{coker}\phi\|_p^p + \|\operatorname{coker}\psi\|_p^p = \sum_{((\mathbf{b}_{\mathcal{A}},\mathbf{d}_{\mathcal{A}}),(\mathbf{b}_{\mathcal{B}},\mathbf{d}_{\mathcal{B}}))\in\mathcal{M}} |\mathbf{d}_{\mathcal{A}}-\mathbf{d}_{\mathcal{B}}|^p + |\mathbf{b}_{\mathcal{A}}-\mathbf{b}_{\mathcal{B}}|^p$$

which is precisely the cost for the transportation plan $\mathcal{M}$.

□

Together Theorems 7.27 and 7.28 imply that the diagram distance is equivalent to the algebraic distance.

7.1. **A Lower Bound for Short Exact Sequences.** In the proof of equivalence between algebraic and diagram distance, the triangle inequality between terms in a short exact sequence follow naturally. In this section, we prove a more surprising result. We prove a *lower bound* for the norm of the middle term of a short exact sequence in terms of the other two terms. This is connected to the space of *extensions* of persistence modules. One consequence of our result is that the trivial extension has the smallest possible p-norm.

Lemma 7.29. *Given a short exact sequence of persistence modules*

$$0 \to \mathcal{A} \overset{\varphi}{\hookrightarrow} \mathcal{B} \overset{\psi}{\twoheadrightarrow} \mathcal{C} \to 0$$

then

$$||\mathcal{A} \oplus \mathcal{C}||_p \leq ||\mathcal{B}||_p$$

Before proving the case of general p, we prove two special cases $p = 1$ and $p = \infty$ whose proof is straightforward.

Lemma 7.30. *Given a short exact sequence of persistence modules*

$$0 \to \mathcal{A} \overset{\varphi}{\hookrightarrow} \mathcal{B} \overset{\psi}{\twoheadrightarrow} \mathcal{C} \to 0$$

then

$$||\mathcal{A} \oplus \mathcal{C}||_1 = ||\mathcal{B}||_1$$

Proof. Observe that for any module $\mathcal{F}$, $\|\mathcal{F}\|_1 = \int_{\mathbb{R}} \mathbf{rk}(\mathcal{F}_t)dt$. We then have

$$\|\mathcal{A} \oplus \mathcal{C}\|_1 = \int_{\mathbb{R}} \mathbf{rk}((\mathcal{A} \oplus \mathcal{C})_t)dt = \int_{\mathbb{R}} \mathbf{rk}(\mathcal{A}_t) + \mathbf{rk}(\mathcal{C}_t)dt = \int_{\mathbb{R}} \mathbf{rk}(\mathcal{B}_t)dt = \|\mathcal{B}\|_1$$

where the third equality holds by exactness restricted to each $t \in \mathbb{R}$. □

Lemma 7.31. *Given a short exact sequence of persistence modules*

$$0 \to \mathcal{A} \overset{\varphi}{\hookrightarrow} \mathcal{B} \overset{\psi}{\twoheadrightarrow} \mathcal{C} \to 0$$

then

$$||\mathcal{A} \oplus \mathcal{C}||_\infty \leq ||\mathcal{B}||_\infty$$

Proof. Assume that

$$||\mathcal{A} \oplus \mathcal{C}||_\infty > ||\mathcal{B}||_\infty$$

There must exist a summand in either $\mathcal{A}$ or $\mathcal{C}$ which is more persistent than any summand in $\mathcal{B}$. Consider the case where a summand in $\mathcal{A}$ achieves the norm, i.e. is the most persistent bar. By the injectivity of φ, the image must be at least as persistent contradicting the assumption. Alternatively, if the norm is achieved in $\mathcal{C}$, the surjectivity of the ψ again contradicts the assumption since there must exist a summand which is at least as persistent as any summand in $\mathcal{C}$. □

Figure 4 illustrates this contradiction. We now prove the general result.

Proof of Lemma 7.29. We first assume that $\mathcal{A}$ and $\mathcal{C}$ are finitely generated. Recall that in homological algebra the short exact sequence

$$0 \to \mathcal{A} \hookrightarrow \mathcal{B} \twoheadrightarrow \mathcal{C} \to 0$$

is known as an extension of $\mathcal{C}$ by $\mathcal{A}$. To prove the result, we show that all extensions have a larger norm than the trivial extension, i.e. $\mathcal{A} \oplus \mathcal{C}$.

We do this iteratively, beginning with the trivial extension and construct a sequence of extensions, each increasing the norm. First, we recount some basic facts from homological algebra. The equivalence classes of extensions are in 1-1 correspondence with $\mathrm{Ext}^1(\mathcal{C}, \mathcal{A})$, where the 0 element corresponds to the trivial extension, see [43](Theorem 3.4.3). Taking a projective resolution of $\mathcal{C}$,

$$0 \to R_{\mathcal{C}} \to G_{\mathcal{C}} \to \mathcal{C} \to 0,$$

where $R_{\mathcal{C}}$ and $G_{\mathcal{C}}$ are freely generated and for persistence modules correspond to deaths and births respectively. That is $R_{\mathcal{C}}$ is the space of boundaries and $G_{\mathcal{C}}$ is the space of cycles. Applying $\mathrm{Hom}(-, \mathcal{A})$, we obtain the following exact sequence

$$\mathrm{Hom}(G_{\mathcal{C}}, \mathcal{A}) \to \mathrm{Hom}(R_{\mathcal{C}}, \mathcal{A}) \to \mathrm{Ext}^1(\mathcal{C}, \mathcal{A}) \to 0$$

Thus for any $\mathcal{B}$, there exists a homomorphism $\gamma : R_{\mathcal{C}} \to \mathcal{A}$, such that

$$\begin{tikzcd}
0 \arrow[r] & R_{\mathcal{C}} \arrow[r, "j"] \arrow[d, "\gamma"] & G_{\mathcal{C}} \arrow[r] \arrow[d] & \mathcal{C} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathcal{A} \arrow[r] & \mathcal{B} \arrow[r] & \mathcal{C} \arrow[r] & 0
\end{tikzcd}$$

where $\mathcal{B}$ is the pushout

$$R_{\mathcal{C}} \xrightarrow{(\gamma,j)} \mathcal{A} \oplus G_{\mathcal{C}}$$

or equivalently

$$\mathcal{B} = \operatorname{coker}(\gamma, j)$$

As $\gamma = 0$ corresponds to $\mathcal{A} \oplus \mathcal{C}$. We denote

$$\gamma_0 = 0, \qquad \mathcal{B}_0 = \mathcal{A} \oplus \mathcal{C}$$

Let γ_n denote the homomorphism which corresponds to $\mathcal{B}$, where n is the rank of $R_{\mathcal{C}}$. Order the generators in order of increasing birth time. Define γ_i as the restriction of γ_n to the first i generators, sending the remaining ones to 0. By construction each corresponding $\mathcal{B}_i$ is an extension.

We now prove that each step increases the norm. Recall, we have

$$R_{\mathcal{C}} \xrightarrow{(\gamma_i,j)} \mathcal{A} \oplus G_{\mathcal{C}} \to \mathcal{B}_i \to 0$$

where $G_{\mathcal{C}}$ is free and $\mathcal{A}$ is not free but is finitely generated. Going from γ_{i-1} to γ_i potentially changes one of the relations, resulting in a different pairing. This is called a cascade [20]. Let $r(i)$ denote the added relation. If this does not change the pairing we continue to $i+1$. If it does, let $(g(i), r(i))$ denote the resulting indecomposable. Observe that all other affected generators satisfy

(9) $$g(i) > g_1(i) > g_2(i) > \ldots > g_k(i)$$

and the affected relations satisfy

(10) $$r(i) < r_1(i) < g_2(i) < \ldots < r_k(i)$$

Note that these may come from $R_{\mathcal{C}}$, $\mathcal{A}$, or both. Here we assume a familiarity with the standard persistence algorithm. Simulating the persistence algorithm between generators and relations – if we view the map between relations and generators as a matrix with the rows indexed by generators sorted in filtration order top-down and the columns indexed by relations sorted in filtration order left to right – a summand is a pivot of relations and generators such that the the pivot is the lowest non-zero entry in the corresponding column. See Figure 5 for an illustration. It is clear that a change in a column can only affect columns which are to the right of it and generators which are above the row of the pivot.

Furthermore, the resulting indecomposables are (g_i, r_i). This can be shown again inductively. Consider the $k \times k$ matrix of the affected generators and relations. Given the first pivot, we can zero out the column above the pivot using row operations and zero out the row using column operations. Hence, the we are left with a $(k-1) \times (k-1)$ matrix where the pivot must again be in the lower left hand corner.

This pairing maximizes the norm over all possible pairings. This follows from the Rearrangement inequality [40, p.78] for $p = 2$. We prove the case for general p in Corollary B.2(see Appendix B). This implies that

$$||\mathcal{B}_i||_p \leq ||\mathcal{B}_{i+1}||_p$$

completing the proof for finitely generated modules.

To conclude the proof for the p.f.d. case, we provide an approximation argument similar to Lemmas 7.21 and 7.22. If $\mathcal{B}$ does not have bounded p-energy, the result holds trivially[4]. If $\mathcal{B}$ has bounded p-energy we can consider a finitely generated ε-approximation of $\mathcal{B}$ denoted by $\mathcal{B}'$. We can construct the following commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{A} \arrow[r, hook, "i"] \arrow[d, two heads, "\varphi"] & \mathcal{B} \arrow[r, two heads, "j"] \arrow[d, two heads, "q_{\mathcal{B}}"] & \mathcal{C} \arrow[r] \arrow[d, two heads, "\psi"] & 0 \\
0 \arrow[r] & \operatorname{im} q_{\mathcal{B}} \circ i \arrow[r, hook] & \mathcal{B}' \arrow[r, two heads] & \operatorname{coker}(j|_{\mathcal{B}'}) \arrow[r] & 0
\end{tikzcd}$$

where $\operatorname{coker}(j|_{\mathcal{B}'})$ is cokernel of the restriction of j to $\mathcal{B}'$. This is well defined since by Lemma 2.10, there is also a monomorphism $\mathcal{B}' \hookrightarrow \mathcal{B}$. We observe that by Lemma 7.8,

$$|| \ker \phi||_p \leq || \ker q_{\mathcal{B}}||_p \leq \varepsilon$$
$$|| \ker \psi||_p \leq || \ker q_{\mathcal{B}}||_p \leq \varepsilon$$

[4]We remark that the case of infintie bars is not interesting in this case as it can be shown that they always split trivially.

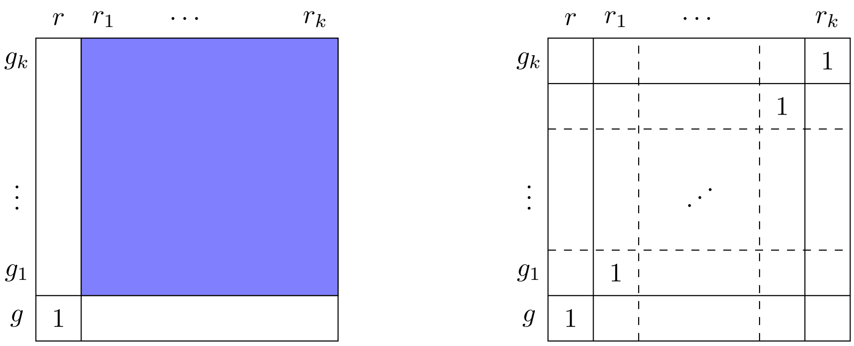


FIGURE 5. The generators are sorted in increasing filtration order top to bottom and the relations are sorted in filtration order left to right as a matrix. (Left) The added relation to γ_i can create a new pivot (shown by the entry 1) in $\mathcal{B}$. The shaded region shows the relations and generators which can be affected. Generators after g since the pivot is defined as the lowest non-zero entry. The relations before r cannot be affected as those columns remain reduces. (Right) Since we only show the affected relations and generators, the matrix after reduction must have the following shape, as each new pivot can only affect the upper right-hand submatrix.

We can use this to bound the p-norm of the direct sum. For ε sufficiently small, there exists a constant K independent of ε such that

$$\begin{aligned}||\mathcal{A} \oplus \mathcal{C}||_p^p = ||\mathcal{A}||_p^p + ||\mathcal{C}||_p^p &\leq (||\operatorname{im} q_{\mathcal{B}}||_p + \varepsilon)^p + (||\operatorname{coker}(j|_{\mathcal{B}'})||_p + \varepsilon)^p \\ &\leq ||\operatorname{im} q_{\mathcal{B}} \oplus \operatorname{coker}(j|_{\mathcal{B}'})||_p^p + K\varepsilon\end{aligned}$$

Taking the p-th root of the above expression yields

$$\left(||\operatorname{im} q_{\mathcal{B}} \oplus \operatorname{coker}(j|_{\mathcal{B}'})||_p^p + K\varepsilon\right)^{1/p} \leq ||\operatorname{im} q_{\mathcal{B}} \oplus \operatorname{coker}(j|_{\mathcal{B}'})||_p + K'\varepsilon$$

Since $p > 1$ the above is a concave function, we can upper bound the expression by a linear function where the slope is some constant K'. Applying the result for finitely generated modules from above

$$||\mathcal{A} \oplus \mathcal{C}||_p||\operatorname{im} q_{\mathcal{B}} \oplus \operatorname{coker}(j|_{\mathcal{B}'})||_p + K'\varepsilon \leq ||\mathcal{B}'||_p + K'\varepsilon \leq ||\mathcal{B}||_p + K'\varepsilon$$

Taking the limit $\varepsilon \to 0$ completes the proof. □

Remark 7.32. *This together with the triangle ineqaulity (from the equivalence with the Wasserstein distance on diagrams) gives a Minkowski-type bound related to short exact sequences of persistence modules,*

$$||\mathcal{A} \oplus \mathcal{C}||_p \leq ||\mathcal{B}||_p \leq ||\mathcal{A}||_p + ||\mathcal{C}||_p$$

where the second inequality follows from the triangle inequality. We note that if $p < \infty$, then

$$||\mathcal{A} \oplus \mathcal{C}||_p^p = ||\mathcal{A}||_p^p + ||\mathcal{C}||_p^p$$

which gives

$$\left(||\mathcal{A}||_p^p + ||\mathcal{C}||_p^p\right)^{1/p} \leq ||\mathcal{B}||_p \leq ||\mathcal{A}||_p + ||\mathcal{C}||_p$$

or equivalently

$$||\mathcal{A}||_p^p + ||\mathcal{C}||_p^p \leq ||\mathcal{B}||_p^p \leq (||\mathcal{A}||_p + ||\mathcal{C}||_p)^p$$

7.2. **Application of Algebraic Stability.** Here we show how the algebraic framework can be used to obtain the results in Section 4 directly. Here we reprove Theorem 4.8 directly on the chain complexes. There are not necessarily finite, but the induced persistence modules must be p.f.d.

Theorem 7.33. *Given $f, g : K \to \mathbb{R}$, let $\mathcal{F}$ and $\mathcal{G}$ denote the persistence modules corresponding to the respective sub-level set filtrations, then if $\mathcal{F}$ and $\mathcal{G}$ are p.f.d.*

$$W_p(\mathrm{Dgm}(\mathcal{F}), \mathrm{Dgm}(\mathcal{G})) \leq ||f - g||_p$$

Proof. Consider the resulting filtered chain complexes $\mathrm{C}_k(f)$ and $\mathrm{C}_k(g)$, where the filtrations are induced by the sub-level sets of the functions f and g. We observe that the filtered chain complexes can be considered as persistence modules, where each simplex generates a bar. We then directly construct the interpolating object at the chain level. For the interpolating object, we consider the sub-level set filtration induced by the function $\max(f, g)$. We obtain the following diagram graded by dimension,

$$\begin{tikzcd}
 & 0 \arrow[d] & & & \\
0 \arrow[r] & \mathrm{C}_k(\max(f,g)) \arrow[r, hook, "i_k"] \arrow[d, hook, "j_k"] & \mathrm{C}_k(f) \arrow[r, two heads] & \operatorname{coker} i_k \arrow[r] & 0 \\
 & \mathrm{C}_k(g) \arrow[d, two heads] & & & \\
 & \operatorname{coker} j_k \arrow[d] & & & \\
 & 0 & & &
\end{tikzcd}$$

where the maps i_k and j_k map a chain to itself. It can be directly verified that these chain maps respect the filtration and that

$$\begin{aligned}
||\operatorname{coker} i_k||_p^p &= \sum_{\substack{\sigma\in K\\ \dim(\sigma)=\mathrm{k}}} (\max(f,g)(\sigma) - f(\sigma))^p \\
||\operatorname{coker} j_k||_p^p &= \sum_{\substack{\sigma\in K\\ \dim(\sigma)=\mathrm{k}}} (\max(f,g)(\sigma) - g(\sigma))^p
\end{aligned} \tag{11}$$

Applying the homology functors to the two short exact sequences for all k, we obtain two long exact sequences.

$$\begin{aligned}
&\dots \xrightarrow{\delta_{k+1}} \mathrm{H}_k(\max(f,g)) \xrightarrow{j_k^*} \mathcal{G} \twoheadrightarrow \mathrm{H}_k(\operatorname{coker} j_k) \xrightarrow{\delta_k} \mathrm{H}_{k-1}(\max(f,g)) \xrightarrow{j_{k-1}^*} \dots \\
&\dots \xrightarrow{\delta'_{k+1}} \mathrm{H}_k(\max(f,g)) \xrightarrow{i_k^*} \mathcal{F} \twoheadrightarrow \mathrm{H}_k(\operatorname{coker} i_k) \xrightarrow{\delta'_k} \mathrm{H}_{k-1}(\max(f,g)) \xrightarrow{i_{k-1}^*} \dots
\end{aligned} \tag{12}$$

Recall that $\operatorname{coker} i_k$ and $\operatorname{coker} j_k$ are chain complexes and since homology is a subquotient, we may define surjective maps

$$\operatorname{coker} i_k \twoheadrightarrow \mathrm{H}_k(\operatorname{coker} i_k) \qquad\qquad \operatorname{coker} j_k \twoheadrightarrow \mathrm{H}_k(\operatorname{coker} j_k)$$

By Lemma 7.8, it follows that

$$||\mathrm{H}_k(\operatorname{coker} i_k)||_p \le ||\operatorname{coker} i_k||_p \qquad\qquad ||\mathrm{H}_k(\operatorname{coker} j_k)||_p \le ||\operatorname{coker} j_k||_p \tag{13}$$

Fixing the homological dimension k, we can extract the relevant terms from the long exact sequences into the following diagram

$$\begin{tikzcd}
 & & 0 \arrow[d] & & & \\
 & & \ker j_k^* \arrow[d] & & & \\
0 \arrow[r] & \ker i_k^* \arrow[r] & \mathrm{H}_k(\max(f,g)) \arrow[r, "i_k^*"] \arrow[d, "j_k^*"] & \mathcal{F} \arrow[r] & \operatorname{coker} i_k^* \arrow[r] & 0 \\
 & & \mathcal{G} \arrow[d] & & & \\
 & & \operatorname{coker} j_k^* \arrow[d] & & & \\
 & & 0 & & &
\end{tikzcd}$$

By exactness of the sequences in Equation 12,

$$\operatorname{coker}\delta_{k+1} \cong \ker i_k^* \qquad \operatorname{im}\delta_k \cong \operatorname{coker} i_k^* \qquad \operatorname{coker}\delta_{k+1} \cong \ker i_k^* \qquad \operatorname{im}\delta_k \cong \operatorname{coker} i_k^*$$

Inserting these isomorphisms into Theorem 7.27,

$$\begin{aligned} W_p(\mathcal{F},\mathcal{G})^p &\leq \left|\left|\bigoplus_k \ker i_k^* \oplus \operatorname{coker} i_k^* \oplus \ker j_k^* \oplus \operatorname{coker} j_k^*\right|\right|_p^p \\ &= \left|\left|\bigoplus_k \operatorname{coker}\delta_k \oplus \operatorname{im}\delta_k \oplus \operatorname{coker}\delta'_k \oplus \operatorname{im}\delta'_k\right|\right|_p^p \end{aligned} \tag{14}$$

Using the short exact sequences,

$$0 \to \ker\delta_k \to \mathrm{H}_k(\operatorname{coker} i_k) \to \operatorname{im}\delta_k \to 0$$
$$0 \to \ker\delta'_k \to \mathrm{H}_k(\operatorname{coker} j_k) \to \operatorname{im}\delta'_k \to 0$$

and Lemma 7.29 yields

$$\begin{aligned} ||\ker\delta_{k-1} \oplus \operatorname{im}\delta_k||_p^p &\leq ||\mathrm{H}_k(\operatorname{coker} i_k)||_p^p \\ ||\ker\delta'_{k-1} \oplus \operatorname{im}\delta'_k||_p^p &\leq ||\mathrm{H}_k(\operatorname{coker} j_k)||_p^p \end{aligned} \tag{15}$$

Rearranging the terms in Equation 14 and using Equation 15,

$$\begin{aligned} W_p(\mathcal{F},\mathcal{G})^p &\leq \left|\left|\bigoplus_k \operatorname{coker}\delta_k \oplus \operatorname{im}\delta_k \oplus \operatorname{coker}\delta'_k \oplus \operatorname{im}\delta'_k\right|\right|_p^p \\ &\leq \left|\left|\bigoplus_k \mathrm{H}_k(\operatorname{coker} i_k) \oplus \mathrm{H}_k(\operatorname{coker} j_k)\right|\right|_p^p \\ &\leq \left|\left|\bigoplus_k \operatorname{coker} i_k \oplus \operatorname{coker} j_k\right|\right|_p^p \\ &= \sum_{\sigma\in K} (\max(f,g)(\sigma) - f(\sigma))^p + (\max(f,g)(\sigma) - g(\sigma))^p \\ &= \sum_{\sigma\in K} (f(\sigma) - g(\sigma))^p = ||f-g||_p^p \end{aligned}$$

The third inequality follows from Equation 13 and the final equality follows from Equation 11 and observing that for any simplex, at least one of the two terms

$$\max(f,g)(\sigma) - f(\sigma) \qquad \max(f,g)(\sigma) - g(\sigma)$$

must be zero.

The case of $p = \infty$ follows similarly by considering the maximum rather than the sum. □

Remark 7.34. *The bound is equivalent to Theorem 4.8. To achieve this, we required Lemma 7.29. We remark that without this result a slightly weaker bound can be achieved using the following inequalities*

$$\begin{aligned} ||\ker i_k^*||_p &\leq ||\mathrm{H}_{k+1}(\operatorname{coker} i_{k+1})||_p \qquad & ||\operatorname{coker} i_k^*||_p &\leq ||\mathrm{H}_k(\operatorname{coker} i_k)||_p \\ |\ker j_k^*||_p &\leq ||\mathrm{H}_{k+1}(\operatorname{coker} j_{k+1})||_p \qquad & ||\operatorname{coker} j_k^*||_p &\leq ||\mathrm{H}_k(\operatorname{coker} j_k)||_p, \end{aligned}$$

which yields

$$W_p(\mathrm{Dgm}(f),\mathrm{Dgm}(g))^p \leq \sum_k \left(2||\operatorname{coker} i_k||_p^p + 2||\operatorname{coker} j_k||_p^p\right) = 2||f-g||_p^p$$

7.3. **Properties of Algebraic Distance.** Finally we explore the requirements for the algebraic distance to define an (extended) pseudometric in more general settings. While this follows for p.f.d. one-parameter persistence modules by the Equivalence Theorem 7.28. This has been extensively explored in the context of *noise systems* [38] and *amplitudes* [27]. We refer the reader to these papers for a more complete discussion in a variety of general settings. Rather, in this section we consider, we consider the special case of the algebraic distance in terms of p-norm of a module. Formally this can be seen as a function such that

$$||\mathcal{A}||_p : \mathcal{A} \mapsto \mathbb{R}^{\geq 0} \cup \{\infty\}$$

The requirement for non-negativity is obvious and even in simple cases the need to allow for the possibility of an infintie distance often arises. We also require the p-norm to have the following properties:

(1) *Zero Module*: The zero module $\not$ has norm 0, i.e. $||\not||_p = 0$

(2) *Monotonicity:* Given $\mathcal{A} \hookrightarrow \mathcal{B}$ or $\mathcal{B} \twoheadrightarrow \mathcal{A}$, implies

$$||\mathcal{A}||_p \leq ||\mathcal{B}||_p$$

.

(3) *Direct Sums:* The norm must be well-behaved under a direct sum: if $||\mathcal{A}||_p = ||\mathcal{B}||_p$, then for any $\mathcal{C}$,

$$||\mathcal{A} \oplus \mathcal{C}||_p = ||\mathcal{B} \oplus \mathcal{C}||_p$$

(4) Subadditivity: Given a short exact sequence,

$$0 \to \mathcal{A} \hookrightarrow \mathcal{B} \twoheadrightarrow \mathcal{C} \to 0,$$

implies

$$||\mathcal{B}||_p \leq ||\mathcal{A}||_p + ||\mathcal{C}||_p$$

We remark on the relation of the above conditions and amplitudes and noise systems. We observe that the above three conditions are more restrictive than amplitudes – which only require monotonicity and subadditivity [27, Definition 3.1]. We remark however that noise systems may be taken to closed under direct sums, which is sufficient to deduce the above property (see [38, Proposition 7.1]).

Recall that a psuedometric $d(\cdot, \cdot)$ must satisfy:

(i) $d(x, y) \leq 0$
(ii) $d(x, x) = 0$
(iii) $d(x, y) = d(y, x)$
(iv) $d(x, z) \leq d(x, y) + d(y, z)$

For the algebraic distance, (i) holds from our assumption on p-norms. Likewise, for (ii), for a module $\mathcal{A}$, we can construct an interpolating object $(\mathcal{A}, \mathrm{id}, \mathrm{id})$. As the kernels and cokernels are trivial, the result holds by the assumption the norm of the zero module and the behaviour of the norms under direct sums. Likewise, the definition of an interpolating object is symmetric, which implies (iii).

All that remains is to verify the triangle inequality. To prove this we will require one additional property of p-norms.

(5) *Duality*: For every short exact sequence

$$0 \to \mathcal{A} \hookrightarrow \mathcal{B} \twoheadrightarrow \mathcal{C} \to 0,$$

we can find a short exact sequence

$$0 \to \mathcal{D} \hookrightarrow \mathcal{E} \twoheadrightarrow \mathcal{F} \to 0,$$

such that

$$||\mathcal{A}||_p = ||\mathcal{F}||_p, \qquad ||\mathcal{B}||_p = ||\mathcal{E}||_p, \qquad ||\mathcal{C}||_p = ||\mathcal{D}||_p.$$

This condition seems extraneous but will often hold in most cases we consider. Perhaps the most obvious example includes taking duality between homology and cohomology. In the case of vector spaces, one can use the non-canonical isomorphism between a vector space and its dual. We leave the categorical foundations of the algebraic distance for future work (as well as the question of whether this is needed).

We conclude this section by showing that if the p-norm satisifies the above conditions, then then algebraic distance satisfies the triangle inequality.

Lemma 7.35. *Given an interpolating object $(\mathcal{E}, i, j)$ between $\mathcal{A}$ and $\mathcal{B}$ and an interpolating object $(\mathcal{F}, r, s)$ between $\mathcal{B}$ and $\mathcal{C}$. There exists an interpolating object $(\mathcal{G}, \varphi, \psi)$ such that*

$$c(\varphi, \psi) \leq c(i, j) + c(r, s)$$

which implies that

$$W_p^{\mathrm{alg}}(\mathcal{A}, \mathcal{C}) \leq W_p^{\mathrm{alg}}(\mathcal{A}, \mathcal{B}) + W_p^{\mathrm{alg}}(\mathcal{B}, \mathcal{C})$$

Proof. We do this constructively. By assumption,

$$c(i, j) = ||\ker i \oplus \ker j \oplus \operatorname{coker} i \oplus \operatorname{coker} j||_p$$
$$c(r, s) = ||\ker r \oplus \ker s \oplus \operatorname{coker} r \oplus \operatorname{coker} s||_p$$

We construct $(\mathcal{G}, \varphi, \psi)$ such that

$$c(\varphi, \psi) \leq c(i, j) + c(r, s)$$

Define $\mathcal{G} = \ker(j+k)$, which yields the following diagram

$$\begin{tikzcd}
\mathcal{G} \arrow[r, "\beta"] \arrow[d, "\alpha"] \arrow[rr, bend left, dashed, "\psi"] \arrow[dd, bend right, dashed, "\varphi"'] & \mathcal{F} \arrow[r, "s"] \arrow[d, "r"] & \mathcal{C} \\
\mathcal{E} \arrow[r, "j"] \arrow[d, "i"] & \mathcal{B} & \\
\mathcal{A} & &
\end{tikzcd}$$

By definition $\varphi = i \circ \alpha$ and $\psi = s \circ \beta$, so there are short exact consequences

$$0 \to \ker \alpha \hookrightarrow \ker \varphi \twoheadrightarrow \alpha(\ker \varphi) \to 0$$

$$0 \to \ker \beta \hookrightarrow \ker \psi \twoheadrightarrow \beta(\ker \psi) \to 0$$

Furthermore since $\operatorname{im} \varphi \subseteq \operatorname{im} i$ and $\operatorname{im} \psi \subseteq \operatorname{im} s$, there are short exact sequences

$$0 \to \operatorname{im} i / \operatorname{im} \varphi \hookrightarrow \operatorname{coker} \varphi \twoheadrightarrow \operatorname{coker} i \to 0$$

$$0 \to \operatorname{im} s / \operatorname{im} \psi \hookrightarrow \operatorname{coker} \psi \twoheadrightarrow \operatorname{coker} \ell \to 0$$

By the Duality Property (5), for the short exact sequence

$$0 \to \ker \beta \hookrightarrow \ker \psi \twoheadrightarrow \beta(\ker \psi) \to 0$$

there exists a short exact sequence such that

$$0 \to \mathcal{X} \hookrightarrow \mathcal{Y} \twoheadrightarrow \mathcal{Z} \to 0$$

such that

$$||\mathcal{X}||_p = ||\beta(\ker \psi)||_p, \qquad ||\mathcal{Y}||_p = ||\ker \psi||_p, \qquad ||\mathcal{Z}||_p = ||\ker \beta||_p$$

Applying the same to

$$0 \to \operatorname{im} s / \operatorname{im} \psi \hookrightarrow \operatorname{coker} \psi \twoheadrightarrow \operatorname{coker} s \to 0$$

and taking the direct sum over all four sequences and using Direct Sum Property (3)

$$\begin{aligned} ||\ker \varphi \oplus \ker \psi \oplus \operatorname{coker} \varphi \oplus \operatorname{coker} \psi||_p \leq{} & ||\ker \alpha \oplus \beta(\ker \psi) \oplus \operatorname{im} i / \operatorname{im} \varphi \oplus \operatorname{coker} s||_p + \\ & ||\alpha(\ker \varphi) \oplus \ker \beta \oplus \operatorname{coker} i \oplus \operatorname{im} s / \operatorname{im} \psi||_p \end{aligned}$$

We observe that there are injective maps $\alpha(\ker \varphi) \hookrightarrow \ker i$ and $\beta(\ker \psi) \hookrightarrow \ker s$, so by the Monotonicity Property (2),

$$||\alpha(\ker \varphi)||_p \leq ||\ker i||_p \qquad\qquad ||\beta(\ker \psi)||_p \leq ||\ker s||_p$$

By construction, β restricted to $\ker \alpha$ is injective. Commutativity implies that $\beta(\ker \alpha) \subseteq \ker r$. Hence,

$$||\ker \alpha||_p \leq ||\ker r||_p.$$

Likewise,

$$||\ker \beta||_p \leq ||\ker j||_p$$

We now claim that $||\operatorname{im} i / \operatorname{im} \varphi||_p \leq ||\operatorname{coker} r||_p$. Pick a non-trivial element $x \in \operatorname{im} i / \operatorname{im} \varphi$. There exists a lift of x to $\mathcal{E}$ denoted y. By construction, $\ker j \subseteq \operatorname{im} \alpha$, so any lift has the property that $j(x)$ is non-trivial. Finally, if $j(x) \in \operatorname{im} r$ it would again imply that $x \in \operatorname{im} \alpha$. We conclude that $j(x) \in \operatorname{coker} r$. Hence

$$||\operatorname{im} i / \operatorname{im} \varphi||_p \leq ||\operatorname{coker} r||_p$$

Likewise,

$$||\operatorname{im} s / \operatorname{im} \psi||_p \leq ||\operatorname{coker} j||_p$$

Substituting the inequalities and using the Subadditive Property (4), we conclude

$$\begin{aligned} ||\ker \varphi \oplus \ker \psi \oplus \operatorname{coker} \varphi \oplus \operatorname{coker} \psi||_p \leq{} & ||\ker i \oplus \ker j \oplus \operatorname{coker} i \oplus \operatorname{coker} j||_p + \\ & ||\ker r \oplus \ker s \oplus \operatorname{coker} r \oplus \operatorname{coker} s||_p \end{aligned}$$

which proves the result. $\square$

8. Discussion

We have investigated Wasserstein stability for persistence diagrams which has a scarcity of results despite becoming increasingly important for applications. While we have presented numerous results (far more than we originally intended), we believe this is a starting point for further investigation. Below we outline possible further questions and directions.

- **Cellular Stability Theorem:** This surprisingly straightforward proof can be extended to other settings of interval decomposable modules, e.g. zig-zag persistence [10], exact and weakly-exact multiparameter modules [5]. Another interesting direction is to consider implication for the study of random functions, e.g. discrete Gaussian random fields.
- **Algebraic Wasserstein Stability:** The algebraic formulation is a clear step toward understanding Wasserstein stability in more general settings where interval decompositions do not exist, including but not limited to multiparameter persistence, sheaf-based persistence, or even more general categorical or algebraic settings. However, there are multiple obstacles before the techniques developed here can be applied, including what is a suitable definition of norm. There have been some suggested following a previous version of this paper [3, 27].

  The notion of an interpolating object is reminiscent of erosion distance [34], it would be interesting to understand the precise relationship between the two notions. Furthermore, while the interpolating object is convenient for stating the results, most approximation results for persistence diagrams use interleaving. In the future work we will describe a notion of interleaving for Wasserstein distance based on the *theory of partial maps*.
- **Applications:** In addition to providing stability bounds for a number of topological summaries, can the results be used for better understanding the behaviour under a small number of outliers. There are also important questions on the stability of distance based filtrations. While explicit bounds independent of the point set size may not exist, the configurations are specific and do not occur if there sufficient randomness, just as on a Poisson point process the expected size of the Delauanay complex is linear rather $n^{\lceil \frac{d}{2} \rceil}$. How one could quantify this to obtain better bounds is an interesting and important question.

  The Wasserstein bounds are potentially useful for investigation of stochastic topology where the 2-Wasserstein distance is most commonly used. While the bottleneck distance is far too coarse, we believe the combination of lower and upper bounds on the persistence norm can produce new local-to-global techniques for understanding random phenomena.

  Finally, this opens up an avenue for bridging persistence and spectral analysis. With $p = 2$, the Cellular Stability Theorem implies a stability in terms of eigenfunction expansions of functions. The obvious result relating squared differences of coefficients with distances between diagrams is limited to fixed, sufficiently nice triangulations. However, we intend to investigate this in future work.

References

[1] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *The Journal of Machine Learning Research*, 18(1):218–252, 2017.

[2] Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. In *Proceedings 30th Annual Symposium on Computational Geometry*, page 355. ACM, 2014.

[3] Håvard Bakke Bjerkevik and Michael Lesnick. ℓ^p-distances on multiparameter persistence modules. *arXiv preprint arXiv:2106.13589*, 2021.

[4] Omer Bobrowski and Goncalo Oliveira. Random čech complexes on riemannian manifolds. *Random Structures & Algorithms*, 54(3):373–412, 2019.

[5] Magnus Bakke Botnan, Vadim Lebovici, and Steve Oudot. On rectangle-decomposable 2-parameter persistence modules. *arXiv preprint arXiv:2002.08894*, 2020.

[6] Peter Bubenik. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102, 2015.

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

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

[9] Peter Bubenik and Tane Vergili. Topological spaces of persistence modules and their properties. *Journal of Applied and Computational Topology*, 2(3-4):233–269, 2018.

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

[11] Mathieu Carriere, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 664–673. JMLR. org, 2017.

[12] F Chazal, WW Crawley-Boevey, and V de Silva. The observable structure of persistence modules. *Homology, Homotopy and Applications*, 18(2):247–265, 2016.

[13] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules. *arXiv preprint arXiv:1207.3674*, 2012.

[14] Frédéric Chazal, Vin De Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.

[15] Chao Chen and Herbert Edelsbrunner. Diffusion runs low on persistence fast. In *2011 International Conference on Computer Vision*, pages 423–430. IEEE, 2011.

[16] Chao Chen, Xiuyan Ni, Qinxun Bai, and Yusu Wang. A topological regularizer for classifiers via persistent homology. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2573–2582, 2019.

[17] Yu-Min Chung and Austin Lawson. Persistence curves: A canonical framework for summarizing persistence diagrams. *arXiv preprint arXiv:1904.07768*, 2019.

[18] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37:103–120, 2007.

[19] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l_p-stable persistence. *Foundations of computational mathematics*, 10(2):127–139, 2010.

[20] David Cohen-Steiner, Herbert Edelsbrunner, and Dmitriy Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the twenty-second annual symposium on Computational geometry*, pages 119–126, 2006.

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

[22] Justin Curry, Sayan Mukherjee, and Katharine Turner. How many directions determine a shape and other sufficiency results for two topological transforms. *arXiv preprint arXiv:1805.09782*, 2018.

[23] Vincent Divol and Théo Lacombe. Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *Journal of Applied and Computational Topology*, pages 1–53, 2020.

[24] Herbert Edelsbrunner and John Harer. *Computational Topology*. American Mathematical Society, 2010.

[25] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Proceedings 41st annual symposium on foundations of computer science*, pages 454–463. IEEE, 2000.

[26] Robert Ghrist, Rachel Levanger, and Huy Mai. Persistent homology and euler integral transforms. *Journal of Applied and Computational Topology*, 2(1-2):55–60, 2018.

[27] Barbara Giunti, John S Nolan, Nina Otter, and Lukas Waas. Amplitudes on abelian categories. *arXiv preprint arXiv:2107.09036*, 2021.

[28] Dejan Govc and Primoz Skraba. An approximate nerve theorem. *Foundations of Computational Mathematics*, 18(5):1245–1297, 2018.

[29] Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems*, pages 1634–1644, 2017.

[30] Christoph D Hofer, Roland Kwitt, and Marc Niethammer. Learning representations of persistence barcodes. *Journal of Machine Learning Research*, 20(126):1–45, 2019.

[31] Sara Kališnik. Tropical coordinates on the space of persistence barcodes. *Foundations of Computational Mathematics*, 19(1):101–129, 2019.

[32] Genki Kusano, Yasuaki Hiraoka, and Kenji Fukumizu. Persistence weighted gaussian kernel for topological data analysis. In *International Conference on Machine Learning*, pages 2004–2013, 2016.

[33] Jacob Leygonie, Steve Oudot, and Ulrike Tillmann. A framework for differential calculus on persistence barcodes. *arXiv preprint arXiv:1910.00960*, 2019.

[34] Amit Patel. Generalized persistence diagrams. *Journal of Applied and Computational Topology*, 1(3-4):397–419, 2018.

[35] Adrien Poulenard, Primoz Skraba, and Maks Ovsjanikov. Topological function optimization for continuous shape matching. In *Computer Graphics Forum*, volume 37, pages 13–25. Wiley Online Library, 2018.

[36] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4741–4748, 2015.

[37] Vanessa Robins and Katharine Turner. Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. *Physica D: Nonlinear Phenomena*, 334:99–117, 2016.

[38] Martina Scolamiero, Wojciech Chachólski, Anders Lundman, Ryan Ramanujam, and Sebastian Öberg. Multidimensional persistence and noise. *Foundations of Computational Mathematics*, 17(6):1367–1406, 2017.

[39] Primoz Skraba, Gugan Thoppe, and D Yogeshwaran. Randomly weighted $d-$ complexes: Minimal spanning acycles and persistence diagrams. *arXiv preprint arXiv:1701.00239*, 2017.

[40] J Michael Steele. *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities*. Cambridge University Press, 2004.

[41] Katharine Turner, Sayan Mukherjee, and Doug M Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA*, 3(4):310–344, 2014.

[42] Zhichao Wang, Qian Li, Gang Li, and Guandong Xu. Polynomial representation for persistence diagram. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6123–6132, 2019.

[43] Charles A Weibel. *An introduction to homological algebra*. Number 38. Cambridge university press, 1995.

Appendix A. Proofs for Section 2

Lemma A.1 (Lemma 2.6)**.** *For any $p' \leq p$, $W_p(X,Y) \leq W_{p'}(X,Y)$.*

Proof. For any matching between diagrams, $\mathbf{M} : X \to Y$, we have

$$\left(\sum_{x\in X} (\mathbf{d}(x) - \mathbf{d}(\mathbf{M}(x)))^p + (\mathbf{b}(x) - \mathbf{b}(\mathbf{M}(x)))^p\right)^{1/p}$$
$$= \left(\sum_{x\in X} (\mathbf{d}(x) - \mathbf{d}(\mathbf{M}(x)))^p + (\mathbf{b}(x) - \mathbf{b}(\mathbf{M}(x)))^p\right)^{\frac{p'}{p'p}}$$
$$\leq \left(\sum_{x\in X} (\mathbf{d}(x) - \mathbf{d}(\mathbf{M}(x)))^{pp'/p} + (\mathbf{b}(x) - \mathbf{b}(\mathbf{M}(x)))^{pp'/p}\right)^{1/p'}$$
$$= \left(\sum_{x\in X} (\mathbf{d}(x) - \mathbf{d}(\mathbf{M}(x)))^{p'} + (\mathbf{b}(x) - \mathbf{b}(\mathbf{M}(x)))^{p'}\right)^{1/p'}$$

To prove the inequality, we show that for $0 < a \leq 1$

$$(x+y)^a \leq x^a + y^a$$

This follows from the fact that the above inequality holds if and only if, $(1+t)^a \leq 1+t^a$ with $0 \leq t \leq 1$. The result follows since the function $1 + t^a - (1+t)^a$ is positive. Taking the derivative of this function, we obtain $a(t^{a-1} - (1+t))^{a-1})$. Since $0 < a < 1$, the derivative is positive and at $t = 0$, the function is 0, hence the function is positive. Since the inequality is true for any matching begin with the matching for p', the resulting matching induces a smaller norm for p, and since the distance is infimum over all matchings the Wasserstein distance is smaller. □

Appendix B. Proof of the Rearrangement Inequality for $p \geq 1$

Lemma B.1. *Given a sequence*

$$a_n \leq a_{n-1} \leq \ldots \leq a_1 \leq b_1 \leq b_2 \leq \ldots \leq b_n$$

and an increasing convex function f, the cost of a matching $\mathbf{M} : \{b_i\} \mapsto \{a_j\}$, is

$$f(\mathbf{M}) = \sum_{i=1}^{n} f(b_i - \mathbf{M}(b_i)).$$

The identity matching, i.e. $b_i \mapsto a_i$, maximizes this sum.

Proof. This follows the proof of the rearrangement inequality in [40, (p. 78)]. Begin with any matching $\mathbf{M} = \mathbf{M}_0$. Define $\mathbf{M}_1$ by performing an inversion:

$$(b_i, \mathbf{M}_{i-1}(b_i)) \mapsto (b_i, a_i)$$
$$(\mathbf{M}_{i-1}^{-1}(a_i), a_i)) \mapsto (\mathbf{M}_{i-1}^{-1}(a_i), \mathbf{M}_{i-1}(b_i))$$

That is we work from the middle of the sequence outward. Since all pairs with index less than i are paired, it follows that

$$\mathbf{M}(b_i) \leq a_i \leq b_i \leq \mathbf{M}^{-1}(a_i)$$

Now we show that $f(\mathbf{M}_i) \leq f(\mathbf{M}_{i+1})$. If the matching is an identity on i, the statement is trivial. Let j denote the index of $\mathbf{M}(b_i)$ and j' the index of $\mathbf{M}^{-1}(a_i)$. We must show

$$f(b_{j'} - a_i) + f(b_i - a_j) \leq f(b_{j'} - a_j) + f(b_i - a_i)$$

As we are only concerned with differences, without loss of generality we can set $a_j' = 0$ and rearranging terms gives

$$b_i^p - (b_i - a_i)^p \leq b_{j'}^p - (b_{j'} - a_i)^p$$

Using the substitutions

$$x_1 = b_i - a_i, \qquad x_2 = b_{j'} - a_i \qquad d = a_i$$

we obtain

$$f(x_1 + d) - f(x_1) \leq f(x_2 + d) - x_2$$

For convex functions, the quantity

$$\frac{f(y) - f(z)}{y - z}$$

is non-decreasing in both y and z. Since the function is increasing, the quantity is always positive for $y \leq z$. This implies

$$\frac{f(x_2+d)-f(x_2)}{d} \leq \frac{f(x_1+d)-f(x_2)}{d+x_2-x_1} \leq \frac{f(x_1+d)-f(x_1)}{d}$$

□

Corollary B.2. *For any sequence as above and $p \geq 1$, the identity matching maximizes*

$$\sum_{i=1}^{n}(b_i - \mathbf{M}(b_i))^p$$

Proof. As the exponentiation function x^p is an increasing convex function for non-negative x, Lemma B.1 implies the result. □

Appendix C. Notation

- K – a finite CW complex
- $\mathrm{Dgm}_k(K, f)$ or $\mathrm{Dgm}_k(f)$ – the k-th dimensional persistence diagram for the filtration induced on K by a monotone function f.
- $\mathrm{Dgm}(K, f)$ or $\mathrm{Dgm}(f)$ – $\bigoplus_k \mathrm{Dgm}_k(K, f)$
- $\mathcal{F}$ - a persistence module
- $\mathrm{Dgm}(\mathcal{F})$ – the persistence diagram of a persistence module
- $\check{\mathrm{C}}(\mathcal{P})$ – the Čech complex
- $\mathcal{R}(\mathcal{P})$ – the Vietoris-Rips complex, where $\mathcal{R}_\delta$ refers to the complex at parameter δ
- $W_p \cdot\cdot$ - the p-th Wasserstein distance
- $W_p^{\mathrm{alg}} \cdot\cdot$ - the p-th algebraic Wasserstein distance
- $\mathrm{C}_k(\cdot)$ – k-dimensional chain complex
- $\mathrm{H}_k(\cdot)$ – k-dimensional homology
- Ω – the inverse map taking points from a persistence diagram to an underlying CW-complex
- $\mathbf{M}$ – a bijective matching between sets
- $\mathbf{C}$ – a corrspondence
- $\hookrightarrow$ – monomorphism
- $\twoheadrightarrow$ – epimorphism
- $\mathbf{rk}$ – the rank
- v, w – geometric points
- x – points in persistence diagrams
- $\mathbf{b}(x)$ – birth time of point x
- $\mathbf{d}(x)$ – death time of point x
- ℓ – lifetime of point x, i.e. $\mathbf{d}(x) - \mathbf{b}(x)$

School of Mathematical Sciences, Queen Mary University of London, London, UK
Email address: p.skraba@qmul.ac.uk

Mathematical Sciences Institute, Australian National University, Canberra, Australia
Email address: katharine.turner@anu.edu.au