Computational Geometry: Theory and Applications 105–106 (2022) 101882

Contents lists available at ScienceDirect

Computational Geometry: Theory and Applications

www.elsevier.com/locate/comgeo

# Universality of persistence diagrams and the bottleneck and Wasserstein distances

Peter Bubenik *, Alex Elchesen

ARTICLE INFO

*Article history:*
Received 28 April 2021
Received in revised form 22 December 2021
Accepted 12 April 2022
Available online 21 April 2022

*Keywords:*
Topological data analysis
Persistent homology
Persistence diagrams
Metric pairs
Formal sums

ABSTRACT

We prove that persistence diagrams with the $p$-Wasserstein distance is the universal $p$-subadditive commutative monoid on an underlying metric space with a distinguished subset. This result applies to persistence diagrams, to barcodes, and to multiparameter persistence modules. In addition, the 1-Wasserstein distance satisfies Kantorovich-Rubinstein duality.



## 1. Introduction

In computational settings persistent homology produces a persistence module that is isomorphic to a finite direct sum of interval modules. The barcode and persistence diagram summarize this collection of intervals [13,16]. Distances between these summaries include the barcode metric based on the dissimilarity distance between intervals [15,16] and the bottleneck and Wasserstein distances for persistence diagrams based on the supremum-norm distance in the plane [13,14]. More generally, given a poset $P$ and a collection $\mathscr{S}$ of indecomposable persistence modules on $P$, one may consider persistence modules on $P$ that are isomorphic to finite direct sums of elements of $\mathscr{S}$. Examples include two-parameter persistence modules isomorphic to a finite direct sum of block modules [7,12] and multi-parameter persistence modules isomorphic to a finite direct sum of rectangle modules [5].

We study distances in a setting that includes all of these examples. We start with a metric pair $(X, d, A)$ (Definition 3.1). For persistence diagrams, the relevant metric pair is $(\mathbb{R}^2_{\leqslant}, d, \Delta)$ or $(\overline{\mathbb{R}}^2_{\leqslant}, d, \Delta)$, where $d$ is the metric obtained from one of the $q$-norms on $\mathbb{R}^2$ (Example 3.2), and for barcodes it is $(\mathrm{Int}(\mathbb{R}), d, \emptyset)$, where $\mathrm{Int}(\mathbb{R})$ denotes the set of intervals in $\mathbb{R}$ and $d$ is either the length of the symmetric difference or the Hausdorff distance (Example 3.4). Given such a metric pair, we construct a free commutative monoid $(D(X, A), +, 0)$ of *persistence diagrams* on $(X, A)$ (Definition 2.2) together with a family of *Wasserstein distances* $W_p$ for all $p \in [1, \infty]$ (Definition 4.12). For persistence diagrams and barcodes we recover the metrics mentioned above (Example 4.15).

We introduce the notion of *p-subadditive commutative metric monoids*, an algebraic and metric object for discussing the above constructions. These are metric spaces which are also monoids and for which the metric is *p-subadditive* (Definition 4.1). We prove the following. For details see Definition 4.23.

* Corresponding author.
*E-mail address:* peter.bubenik@ufl.edu (P. Bubenik).

https://doi.org/10.1016/j.comgeo.2022.101882
0925-7721/

**Proposition 1.1** *(Functorial construction of persistence diagrams with Wasserstein distance). Given a metric pair* $(X, d, A)$*,* $(D(X, A), W_p, +, 0)$ *is a* $p$*-subadditive commutative metric monoid and the canonical inclusion* $i : (X, d, A) \to (D(X, A), W_p, 0)$ *is* 1*-Lipschitz. Furthermore, this construction is functorial.*

Our main result is the following. For details see Theorem 4.25.

**Theorem 1.2** *(Persistence diagrams with the Wasserstein distance as an adjoint). The forgetful functor from* $p$*-subadditive commutative metric monoids to metric pairs has a left adjoint given by the functor in Proposition 1.1.*

An equivalent statement of this result is that persistence diagrams with the Wasserstein distance are *universal* as follows.

**Theorem 1.3** *(Universality of persistence diagrams with the Wasserstein distance). Given a metric pair* $(X, d, A)$*,* $(D(X, A), W_p, +, 0)$ *is the universal* $p$*-subadditive commutative metric monoid obtained from* $(X, d, A)$*. That is, given any* $p$*-subadditive commutative metric monoid* $(N, \rho, +, 0)$ *and* 1*-Lipschitz map* $\varphi : (X, d, A) \to (N, \rho, 0)$*, there is a unique* 1*-Lipschitz monoid homomorphism* $\tilde{\varphi} : (D(X, A), W_p, +, 0) \to (N, \rho, +, 0)$ *such that* $\tilde{\varphi} i = \varphi$*.*

$$
\begin{tikzcd}
(X,d,A) \arrow[r, "i"] \arrow[dr, "\varphi"'] & (D(X,A),W_p,0) \\
& (N,\rho,0)
\end{tikzcd}
\quad
\begin{tikzcd}
(D(X,A),W_p,+,0) \arrow[d, dashed, "\exists!\tilde{\varphi}"] \\
(N,\rho,+,0)
\end{tikzcd}
\quad
\begin{tikzcd}
(X,d,A) \arrow[r, "i"] \arrow[dr, "\varphi"'] & (D(X,A),W_p,0) \arrow[d, dashed, "\tilde{\varphi}"] \\
& (N,\rho,0)
\end{tikzcd}
$$

From this it follows that the $p$-Wasserstein distance is the largest $p$-subadditive distance for persistence diagrams. For details see Theorem 5.1 and Definition 3.12.

**Corollary 1.4** *(Wasserstein distance as largest subadditive distance). Given a metric pair* $(X, d, A)$*, the* $p$*-Wasserstein distance* $W_p$ *is the largest* $p$*-subadditive metric on* $D(X, A)$ *compatible with* $d$*.*

The following related result is of independent interest. For details see Theorem 4.11.

**Theorem 1.5** *(Symmetric monoidal structures for pointed metric spaces). For each* $p \in [1, \infty]$ *there is a symmetric monoidal category* $\mathbf{Met}_*^{\mathbf{p}}$ *(Definitions 3.10 and 4.5 and Corollary 4.10) for which the category of commutative monoids internal to* $\mathbf{Met}_*^{\mathbf{p}}$ *is the category of* $p$*-subadditive commutative metric monoids.*

As a corollary to these results we obtain Converse Stability Theorems (Section 5.2). When $p = 1$, the Wasserstein distance satisfies Kantorovich-Rubinstein duality (Section 5.3).

### *1.1. Related work*

Various metrics in applied topology have been shown to be universal in the sense that they are maximal stable distances. These include the matching distance [17], the interleaving and bottleneck distance [19], the homology interleaving distance [6], and the Reeb graph edit distance [8]. In contrast, we show that the Wasserstein distances are universal in the sense of category-theory (Theorem 1.3).

A version of our Corollary 1.4 appears in [10]. Their version does not assume that the sum is finite but does restrict to the special case that the set $X$ is a set of objects in a Grothendieck category with local endomorphism rings and that the set $A$ consists of the zero object. Also, they do not show that their result follows from a functorial construction.

The Wasserstein distances between persistence diagrams has been studied extensively by Divol and Lacombe [18]. There they relate the Wasserstein distance between persistence diagrams to the classical Wasserstein distance on probability measures. Among other things, this allows for a version of Kantorovich-Rubinstein duality to be recovered for persistence diagrams.

Note that all of the persistence diagrams defined in the present paper are finite by definition. Extensions of the ideas presented here to countable persistence diagrams, such as those in [14] and [4], and signed persistence diagrams, as well as to the setting of Radon measures, can be found in the sequel [3].

Skraba and Turner [22] have shown that the Wasserstein distance for persistence diagrams of weighted cell complexes is stable.

## 2. Background

### 2.1. Metric spaces

In order to include various distances arising in persistent homology, we will use a less restrictive notion of metric space than is standard. Such relaxed metrics are referred to as extended pseudometrics in the literature, but we will refer to them as metrics for brevity.

**Definition 2.1.** A *metric space* is a tuple $(X,d)$ where $X$ is a set and $d : X \times X \to [0,\infty]$ is a function satisfying $d(x,x) = 0$ for all $x \in X$ (point triviality), $d(x, y) = d(y, x)$ for all $x, y \in X$ (symmetry), and $d(x, y) \leqslant d(x,z) + d(z, y)$ for all $x, y,z \in X$ (triangle inequality). Given metric spaces $(X,d_X)$ and $(Y,d_Y)$, a *metric map* is a function $f : X \to Y$ such that $d_Y(f(x), f(y)) \leqslant d_X(x, y)$ for all $x, y \in X$. Given a set $X$, a metric space $(Y,d)$, and a function $g : X \to Y$, the *pullback of d along g*, denoted $g^*d$, is the metric on $X$ defined by $g^*d(x,x') := d(g(x), g(x'))$.

### 2.2. Monoids

A *commutative monoid* $M = (M, +, 0)$ is a set $M$ together with an associative commutative binary operation $+ : M \times M \to M$ for which there exists an element $0 \in M$ satisfying $m + 0 = m$ for all $m \in M$, called the *identity* element. A *monoid homomorphism* between commutative monoids $M = (M, +_M, 0_M)$ and $N = (N, +_N, 0_N)$ is a map $f : M \to N$ such that $f(a +_M b) = f(a) +_N f(b)$ for all $a, b \in M$ and $f(0_M) = 0_N$. A subset $P \subset M$ is a *submonoid* if it contains 0 and $+$ restricts to a binary operation on $P$.

Given a set $X$, the *free commutative monoid on* $X$, denoted $D(X)$, is the set of all (finite) *formal sums* of elements of $X$, with the monoid operations being given by addition of formal sums. That is, $D(X)$ is the set of all functions $f : X \to \mathbb{N} \cup \{0\}$ with finite support and with the monoid operation given by the pointwise addition of functions. Formal sums are also called finite multisets. For $x \in X$, let $1_x : X \to \mathbb{N} \cup \{0\}$ be given by $1_x(x) = 1$ and $1_x(y) = 0$ for all other $y \in X$. As is customary, we denote $1_x$ by $x$. With this convention, we may write any formal sum $\alpha \in D(X)$ as $\alpha = x_1 + \cdots + x_n$, where $n \geqslant 0$ and $x_1,\ldots,x_n \in X$. We define the *canonical inclusion* $i : X \to D(X)$ by $i(x) = x$.

An equivalence relation $\sim$ on a commutative monoid $M$ is called a *congruence* if $a \sim b$ and $c \sim d$ implies $a + c \sim b + d$. If $\sim$ is a congruence then there is a well-defined commutative monoid structure on the set of equivalence classes $M/\sim$ given by $[a] + [b] := [a + b]$. Let $M$ be a commutative monoid and $P \subseteq M$ any submonoid. Define a relation $\sim$ on $M$ by

$$a \sim b \iff \exists x, y \in P \text{ such that } a + x = b + y.$$

Then $\sim$ is a congruence and we denote the commutative monoid $M/\sim$ by $M/P$ and refer to it as the *quotient of M by P*.

A *set pair* is a pair $(X, A)$ where $X$ is a set and $A$ is a nonempty subset of $X$. A map of pairs $f : (X, A) \to (Y, B)$ is a function $f : X \to Y$ such that $f(A) \subset B$. A *pointed set* is a pair $(X, \{x_0\})$, which is denoted $(X, x_0)$. Given pointed sets $(X, x_0)$ and $(Y, y_0)$, a *pointed function* $f : (X, x_0) \to (Y, y_0)$ is a function $f : X \to Y$ such that $f(x_0) = y_0$.

**Definition 2.2.** Given a pair $(X, A)$, let $D(X, A)$ denote the quotient monoid $D(X)/D(A)$. We call $D(X, A)$ the commutative monoid of *persistence diagrams* on $(X, A)$. Note that $D(X)/D(A) \cong D(X \setminus A)$. Given a map of pairs $f : (X, A) \to (Y, B)$, there is an induced monoid homomorphism $f_* : D(X, A) \to D(Y, B)$ given by $f_*(x_1 + \cdots + x_n) = f(x_1) + \cdots + f(x_n)$. Note that this also defines a pointed function $f_* : (D(X, A), 0) \to (D(Y, B), 0)$.

### 2.3. p-norms

Let $p \in [1,\infty]$ and $\mathbf{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n$. For $p < \infty$, the *p-norm of* $\mathbf{x}$ is defined by $\|\mathbf{x}\|_p = \left(\sum_{k=1}^n |x_k|^p\right)^{1/p}$ and for $p = \infty$, it is defined by $\|\mathbf{x}\|_\infty = \max_{1\leqslant k\leqslant n} |x_k|$. For $x = (x_1,\ldots,x_m)$, $y = (x_{m+1},\ldots,x_{m+n})$, and $z = (x_1,\ldots,x_{m+n})$, $\|(\|x\|_p, \|y\|_p)\|_p = \|z\|_p$. By the $\ell^p$*-distance on* $\mathbb{R}^n$ we mean the metric induced by the $p$-norm, i.e., $\|\mathbf{x} - \mathbf{y}\|_p$. The fact that each $\|\cdot\|_p$ is a norm relies on the Minkowski inequality: for all $p \in [1,\infty]$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{x} + \mathbf{y}\|_p \leqslant \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$. The $p$-norms are related as follows, which shows in particular that the $p$-norms are decreasing in $p$. For $\mathbf{x} \in \mathbb{R}^n$ and $1 \leqslant p \leqslant q \leqslant \infty$ we have $\|\mathbf{x}\|_q \leqslant \|\mathbf{x}\|_p \leqslant n^{\frac{1}{p}-\frac{1}{q}} \|\mathbf{x}\|_q$, and these inequalities are attained. Here we adopt the convention $\frac{1}{\infty} = 0$.

Let $\overline{\mathbb{R}}$ denote the set of extended real numbers $[-\infty, +\infty]$. The $\ell^p$-distance on $\mathbb{R}^n$ extends to $\overline{\mathbb{R}}^n$, with the understanding that it may take a value of $\infty$.

### 2.4. Basic category theory

A *category* $\mathbf{C}$ consists of a class $\mathrm{obj}(\mathbf{C})$ of *objects*, and for each pair of objects $X, Y \in \mathrm{obj}(\mathbf{C})$, a set $\mathbf{C}(X, Y)$ of *morphisms* (or *arrows*). The class of all morphisms of $\mathbf{C}$ is denoted $\mathrm{Hom}(\mathbf{C})$. A morphism $f \in \mathbf{C}(X, Y)$ is often denoted $f : X \to Y$. We will often simply write $X \in \mathbf{C}$ to indicate that $X$ is an object of $\mathbf{C}$. A category is *small* if $\mathrm{obj}(\mathbf{C})$ is a set as opposed to a proper class.

The objects and morphisms of a category **C** are required to satisfy the following axioms. For any objects $X, Y, Z \in \mathbf{C}$ and morphisms $f \in \mathbf{C}(X,Y)$, $g \in \mathbf{C}(Y,Z)$, there exists a morphism $g \circ f \in \mathbf{C}(X,Z)$, called the *composition of $f$ and $g$*. That is, there is a function $\mathbf{C}(X,Y) \times \mathbf{C}(Y,Z) \to \mathbf{C}(X,Z)$ given by $(f,g) \mapsto g \circ f$. We will often omit the $\circ$, writing $gf$ instead of $g \circ f$. Composition must be *associative*, meaning that $(hg)f = h(gf)$ whenever this composition is defined. Finally, for all $X \in \mathbf{C}$, there exists a morphism $\mathrm{id}_X : X \to X$ such that, for all $W, Y \in \mathbf{C}$ and $f : W \to X$, $g : X \to Y$, we have $\mathrm{id}_X f = f$ and $g\,\mathrm{id}_X = g$.

A *subcategory* **D** of **C** consists of a subclass obj(**D**) of obj(**C**) and a subclass Hom(**D**) of Hom(**C**) such that if $f : X \to Y \in \mathrm{Hom}(\mathbf{D})$ then $X, Y \in \mathrm{obj}(\mathbf{D})$, $\mathrm{id}_X \in \mathrm{Hom}(\mathbf{D})$ for all $X \in \mathrm{obj}(\mathbf{D})$, and $fg \in \mathrm{Hom}(\mathbf{D})$ whenever $f, g \in \mathrm{Hom}(\mathbf{D})$ and this composition is defined in **C**. This definition guarantees that **D** is a category in its own right. **D** is a *full subcategory* of **C** if $\mathbf{D}(X,Y) = \mathbf{C}(X,Y)$ for all $X, Y \in \mathbf{D}$.

Objects $X, Y \in \mathbf{C}$ are said to be *isomorphic* if there exist morphisms $f : X \to Y$ and $g : Y \to X$ such that $gf = \mathrm{id}_X$ and $fg = \mathrm{id}_Y$.

**Example 2.3.** Let **Set** denote the category whose objects are sets and whose morphisms are functions between sets. Composition is given by the composition of functions and the identity morphism on a set $S$ is the identity function on $S$. Isomorphisms in **Set** are bijective functions.

**Example 2.4.** Let **Met** denote the category whose objects are metric spaces $X = (X, d)$ and whose morphisms are metric maps (see Definition 2.1). Composition of metric maps is given by the composition of functions and the identity morphism on $X$ is the identity function on $X$. It is easily checked that the composition of metric maps is again a metric map, as is the identity function. Isomorphisms in **Met** are isometries.

**Example 2.5.** Let **CMon** denote the category whose objects are commutative monoids $M = (M, +, 0)$ and whose morphisms are monoid homomorphisms. Composition of monoid homomorphisms is given by the composition of functions and the identity morphism on $M$ is the identity function on $M$. Isomorphisms in **CMon** are monoid isomorphisms.

A *covariant functor* $F : \mathbf{C} \to \mathbf{D}$ consists of a map $F : \mathrm{obj}(\mathbf{C}) \to \mathrm{obj}(\mathbf{D})$ and, for each $X, Y \in \mathbf{C}$, a map $F : \mathbf{C}(X,Y) \to \mathbf{D}(F(X), F(Y))$. For all $X, Y, Z \in \mathbf{C}$ and $f \in \mathbf{C}(X,Y)$, $g \in \mathbf{C}(Y,Z)$, these maps must satisfy $F(gf) = F(g)F(f)$ and $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$.

A *contravariant functor* $F : \mathbf{C} \to \mathbf{D}$ consists of a map $F : \mathrm{obj}(\mathbf{C}) \to \mathrm{obj}(\mathbf{D})$ and, for each $X, Y \in \mathbf{C}$, a map $F : \mathbf{C}(X,Y) \to \mathbf{D}(F(Y), F(X))$. For all $X, Y, Z \in \mathbf{C}$ and $f \in \mathbf{C}(X,Y)$, $g \in \mathbf{C}(Y,Z)$, these maps must satisfy $F(gf) = F(f)F(g)$ and $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$. Note that a contravariant functor reverses the direction of arrows in the sense that if $f : X \to Y$ then $F(f) : F(Y) \to F(X)$.

**Example 2.6.** Any category admits an identity functor $1_{\mathbf{C}} : \mathbf{C} \to \mathbf{C}$ which maps objects and morphisms to themselves.

**Example 2.7.** If **D** is a subcategory of **C**, then the inclusion $\mathbf{D} \hookrightarrow \mathbf{C}$ is a covariant functor.

**Example 2.8.** Let $G : \mathbf{D} \to \mathbf{C}$ be a covariant functor. For each $C \in \mathbf{C}$, there is a covariant functor $\mathbf{C}(C, G-) : \mathbf{D} \to \mathbf{Set}$ which sends $D \in \mathbf{D}$ to $\mathbf{C}(C, GD) \in \mathbf{Set}$, and which sends a morphism $f : D \to D'$ in **D** to the set map $\mathbf{C}(C, Gf) : \mathbf{C}(C, GD) \to \mathbf{C}(C, GD')$ given by $g : C \to GD \mapsto (Gf)g : C \to GD'$.

Similarly, given a covariant functor $F : \mathbf{C} \to \mathbf{D}$ and $D \in \mathbf{C}$, there is a contravariant functor $\mathbf{D}(F-, D) : \mathbf{C} \to \mathbf{Set}$ which sends $C \in \mathbf{C}$ to $\mathbf{D}(FC, D) \in \mathbf{Set}$, and which sends a morphism $f : C \to C'$ in **C** to the set map $\mathbf{D}(Ff, D) : \mathbf{D}(FC', D) \to \mathbf{D}(FC, D)$ given by $h : FC' \to D \mapsto h(Ff) : FC \to D$.

As a special case, if $G = 1_{\mathbf{C}}$, for any $C \in \mathbf{C}$ we obtain the *covariant hom-functor* $\mathbf{C}(C, -)$. Similarly, if $F = 1_{\mathbf{C}}$ then for any $C \in \mathbf{C}$ we obtain the *contravariant hom-functor* $\mathbf{C}(-, C)$.

**Example 2.9.** The *forgetful functor* $U : \mathbf{CMon} \to \mathbf{Set}$ sends a commutative monoid to its underlying set, and sends a monoid homomorphism to the function defining it. That is, $U$ sends a commutative monoid to the set obtained by "forgetting" the monoid structure.

Similarly, there is a forgetful functor $U : \mathbf{Met} \to \mathbf{Set}$.

**Example 2.10.** The *free commutative monoid functor* $D : \mathbf{Set} \to \mathbf{CMon}$ sends a set $X$ to the free commutative monoid $D(X)$ (see Section 2.2) and sends a function $f : X \to Y$ to the monoid homomorphisms $D(f) : D(X) \to D(Y)$ given by $x_1 + \cdots + x_n \mapsto f(x_1) + \cdots + f(x_n)$.

A *natural transformation* $\alpha : F \Rightarrow G$ between functors $F, G : \mathbf{C} \to \mathbf{D}$ (either both covariant or both contravariant), denoted $\alpha : F \Rightarrow G$, consists of, for each $C \in \mathbf{C}$, a morphism $\alpha_C : F(C) \to G(C)$. If $F$ and $G$ are both covariant then we require that for any $D, E \in \mathbf{C}$ and any morphism $f \in \mathbf{C}(D, E)$, we have $\alpha_E F(f) = G(f)\alpha_D$. If $F$ and $G$ are both contravariant then we require that $\alpha_D F(f) = G(f)\alpha_E$. The morphism $\alpha_C$ is called *the component of* $\alpha$ *at* $C$. If all of the components of $\alpha$ are isomorphisms, then $\alpha$ is called a *natural isomorphism*. If $\alpha$ is a natural isomorphism from $F$ to $G$ then we say that $F$ and $G$ are *naturally isomorphic* and write $F \cong G$.

An *adjunction* between categories $\mathbf{C}$ and $\mathbf{D}$ consists of functors $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ such that, for each $D \in \mathbf{D}$ there is a natural isomorphism $\mathbf{D}(F-, D) \cong \mathbf{C}(-, GD)$ and for each $C \in \mathbf{C}$ there is a natural isomorphism $\mathbf{D}(FC, -) \cong \mathbf{C}(C, G-)$. In this case, we say that $F$ is the *left adjoint of* $G$ and $G$ *is the right adjoint of* $F$. We also write $F \dashv G$.

### 2.5. Universal properties and adjunctions

Consider a functor $U : \mathbf{D} \to \mathbf{C}$ between categories $\mathbf{D}$ and $\mathbf{C}$ and let $X \in \mathbf{C}$. An object $Y \in \mathbf{D}$ satisfies a *universal property* with *universal element* $i : X \to UY \in \mathbf{C}$ if for every object $Z \in \mathbf{D}$ and morphism $f : X \to UZ$ in $\mathbf{C}$ there is a unique morphism $g : Y \to Z$ in $\mathbf{D}$ such that $Ug \circ i = f$.

$$\begin{tikzcd} X \arrow[r, "i"] \arrow[dr, "f"'] & UY \\ & UZ \end{tikzcd} \qquad \begin{tikzcd} Y \arrow[d, dashed, "\exists! g"] \\ Z \end{tikzcd} \qquad \begin{tikzcd} X \arrow[r, "i"] \arrow[dr, "f"'] & UY \arrow[d, dashed, "Ug"] \\ & UZ \end{tikzcd} \tag{2.1}$$

For example, consider the forgetful functor $U : \mathbf{CMon} \to \mathbf{Set}$ and let $X \in \mathbf{Set}$. Then the free commutative monoid $D(X) \in \mathbf{CMon}$ satisfies a universal property with universal element the canonical inclusion $i : X \to UD(X) \in \mathbf{Set}$.

Given $C \in \mathbf{C}$ and a functor $U : \mathbf{D} \to \mathbf{C}$, the *comma category* $C \downarrow U$ is the category whose objects are pairs $(D, f : C \to UD)$, where $D \in \mathbf{D}$ and $f : C \to UD$ is a morphism in $\mathbf{C}$, and whose morphisms $(D, f : C \to UD) \to (D', f' : C \to UD')$ are morphisms $g : D \to D'$ in $\mathbf{D}$ for which $Ug \circ f = f'$. Then a more succinct way of stating (2.1) is that $(Y, i)$ is the initial object of the category $X \downarrow U$.

**Remark 2.11.** The *category of elements* of a functor $F : \mathbf{C} \to \mathbf{Set}$ is the category whose objects are pairs $(C, x)$, where $C \in \mathbf{C}$ and $x \in FC$, and whose morphisms $(C, x) \to (C', x')$ are morphisms $f : C \to C'$ in $\mathbf{C}$ for which $Ff(x) = x'$. Then $X \downarrow U$ is precisely the category of elements of the functor $\mathbf{C}(X, U-) : \mathbf{D} \to \mathbf{Set}$. Then yet another way of stating (2.1) is that there is a natural isomorphism $\mathbf{C}(X, U-) \cong \mathbf{D}(Y, -)$ defined by $i \in \mathbf{C}(X, UY)$. That is, $\mathbf{C}(X, U-)$ is represented by $Y$ via $i$.

The following lemma shows how a family of universal properties can be used to obtain an adjunction. We will use it frequently throughout.

**Lemma 2.12** *([21, Lemma 4.6.1]). A functor $U : \mathbf{D} \to \mathbf{C}$ admits a left adjoint if and only if for each object $X$ in $\mathbf{C}$ the comma category $X \downarrow U$ has an initial object.*

For example, the forgetful functor $U : \mathbf{CMon} \to \mathbf{Set}$ has a left adjoint $D : \mathbf{Set} \to \mathbf{CMon}$, the free commutative monoid functor. This follows from Lemma 2.12 since for any set $X$, $X \downarrow U$ has the initial object $(D(X), i_X)$, where $D(X)$ is the free commutative monoid and $i_X : X \to D(X)$ is the canonical inclusion (see Section 2.2).

A special case of (2.1) has $U : \mathbf{D} \to \mathbf{C}$ being the inclusion of a full subcategory.

**Definition 2.13.** A *reflective* subcategory of a category $\mathbf{C}$ is a full subcategory $\mathbf{D}$ of $\mathbf{C}$ such that the inclusion $\mathbf{D} \hookrightarrow \mathbf{C}$ has a left adjoint. The left adjoint of the inclusion is called the *reflector*.

### 2.6. Symmetric monoidal categories and internal objects

We will be interested in metric spaces which are also commutative monoids for which the metric and monoid structures are in some sense compatible. This idea is formalized categorically by the notion of a commutative monoid internal to a category. In order to make this notion precise, the category in question must have additional structure - that of a symmetric monoidal category. For example, a commutative topological monoid, analogous to a topological group, is a commutative monoid internal to **Top**, the (symmetric monoidal) category of topological spaces and continuous maps. In this section we sketch the formal definition of symmetric monoidal categories and commutative monoid objects internal to them. For a complete treatment, see [20, VII.1].

A *symmetric monoidal category* is a category $\mathbf{C}$ equipped with a functor $\otimes : \mathbf{C} \times \mathbf{C} \to \mathbf{C}$ called the *tensor product*, an object $1 \in \mathbf{C}$ called the *unit object*, a natural isomorphism $\alpha_{X,Y,Z} : (X \otimes Y) \otimes Z \to X \otimes (Y \otimes Z)$ called the *associator*, a natural isomorphism $\lambda_X : 1 \otimes X \to X$ called the *left unitor*, a natural isomorphism $\rho_X : X \otimes 1 \to X$ called the *right unitor*, and a natural isomorphism $B_{X,Y} : X \otimes Y \to Y \otimes X$ called the *braiding*. These natural isomorphisms must satisfy certain *coherence conditions* expressed by commutative diagrams. For example, the unitors and the associator must obey the *triangle equality*, which specifies that the diagram

$$\begin{tikzcd}
(X \otimes 1) \otimes Y \arrow[rr, "\alpha_{X,1,Y}"] \arrow[dr, "\rho_X \otimes \mathrm{id}_Y"'] & & X \otimes (1 \otimes Y) \arrow[dl, "\mathrm{id}_X \otimes \lambda_Y"] \\
& X \times Y &
\end{tikzcd}$$

commutes. For the other coherence conditions and their diagrams, see [20, VII.1].

**Example 2.14.** **Set** is a symmetric monoidal category with tensor product given by the cartesian product and with unit object being the one-point set $*$. The associator, left unitor, right unitor, and braiding are defined by the obvious bijections $(X \times Y) \times Z \cong X \times (Y \times Z)$, $* \times X \cong X$, $X \times * \cong X$, and $X \times Y \cong Y \times X$, respectively.

**Example 2.15.** **Met** is a symmetric monoidal category with tensor product $(X, d_X) \otimes (Y, d_Y) := (X \times Y, D_\infty)$, where $D_\infty((x, y), (x', y')) := \max(d_X(x, x'), d_Y(y, y'))$, and with unit object the one-point metric space $*$. The associator, unitors, and braiding are the same as those in **Set** (it only needs to be checked that these maps are isometries, i.e., isomorphisms in **Met**).

For any symmetric monoidal category $(\mathbf{C}, \otimes, 1)$, a notion of a commutative monoid defined within **C** can be made precise. A *commutative monoid object* in $(\mathbf{C}, \otimes, 1)$ (or a *commutative monoid internal to* $(\mathbf{C}, \otimes, 1)$) is a tuple $(M, \mu, e)$ consisting of an object $M \in \mathbf{C}$, a morphism $e : 1 \to M$ called the *unit*, and a morphism $\mu : M \otimes M \to M$ called the *product*. The unit and product morphisms are required to satisfy certain coherence conditions with the associator, unitors, and braiding of **C**, expressing associativity and commutativity of the product and the fact that the unit morphism serves as an identity for the product. For example, the unit $e : 1 \to M$ is required to make the diagram

$$\begin{tikzcd}
1 \otimes M \arrow[r, "e \otimes \mathrm{id}_M"] \arrow[dr, "\lambda_M"'] & M \otimes M \arrow[d, "\mu"] & M \otimes 1 \arrow[l, "\mathrm{id}_M \otimes e"'] \arrow[dl, "\rho_M"] \\
& M &
\end{tikzcd}$$

commute. See again [20, VII.1] for the other coherence conditions.

A morphism between commutative monoid objects $(M, \mu, e)$, $(M', \mu', e')$ in a symmetric monoidal category $(\mathbf{C}, \otimes, 1)$ is a morphism $f : M \to M'$ in **C** such that $f\mu = \mu'(f \otimes f)$ and $fe = e'$. Commutative monoid objects in $(\mathbf{C}, \otimes, 1)$, together with these morphisms, form a category which is denoted by $\mathbf{CMon}(\mathbf{C}, \otimes, 1)$ or more simply by $\mathbf{CMon}(\mathbf{C})$ when the tensor product and unit are fixed.

**Example 2.16.** Recall that **Set** is a symmetric monoidal category with tensor product the cartesian product and with unit object the one-point set (see Example 2.14). The commutative monoid objects internal to **Set** are just commutative monoids, and the corresponding morphisms are monoid homomorphisms. That is, we have an isomorphism of categories $\mathbf{CMon}(\mathbf{Set}) \cong \mathbf{CMon}$.

**Example 2.17.** Let **Met** be given the symmetric monoidal structure of Example 2.15. A commutative monoid internal to **Met** is a metric space $(M, d)$ together with a unit $e : * \to \mathbf{Met}$ and a multiplication operation $+ : M \times M \to M$. Note that $e$ simply picks out an element of $M$, which we will denote by 0. The fact that $+$ is metric map means that

$$d(a + b, a' + b') \leqslant \max(d(a, a'), d(b, b')) \quad \text{for all } a, a', b, b' \in M. \tag{2.2}$$

Thus a commutative monoid internal to **Met** is a tuple $(M, d, +, 0)$, where $(M, d)$ is a metric space, $(M, +, 0)$ is a commutative monoid, and such that (2.2) holds.

We will consider variations of the preceding example for pointed metric spaces and for different choices of tensor product.

## 3. Metric pairs and pointed metric spaces

In this section, we introduce the categories of metric spaces of interest to us. In Section 3.1, we introduce the categories of metric pairs and pointed metric spaces. We show that every metric pair gives rise to a pointed metric space by taking a quotient and that this construction is functorial. Moreover, we show that the category of pointed metric spaces is a reflective subcategory of the category of metric pairs, with reflector being the quotient functor. In Section 3.2, we introduce the $p$-strengthened triangle inequality ($p \in [1, \infty]$) and the corresponding subcategories of pointed metric spaces which satisfy it. These subcategories are required for our statement of universality.

### 3.1. Metric pairs, pointed metric spaces, and quotients

We now introduce the main categories of interest to us.

**Definition 3.1.** Let $\mathbf{Met_{pairs}}$ denote the category whose objects are of the form $(X, d, A)$, where $(X, d)$ is a metric space with $A$ a nonempty subset of $X$ and whose morphisms $f : (X, d, A) \to (Y, d', B)$ are metric maps $f : (X, d) \to (Y, d')$ such that $f(A) \subset B$. $(X, d, A)$ is called a *metric pair*.

**Example 3.2.** Consider the metric space $(\mathbb{R}^2, d)$ where $d$ is the metric induced by the $q$-norm, where $1 \leqslant q \leqslant \infty$. Let $\mathbb{R}^2_{\leqslant} = \{(x, y) \in \mathbb{R}^2 \mid x \leqslant y\}$ and similarly define subsets $\mathbb{R}^2_{\geqslant}$ and $\mathbb{R}^2_{=}$ where the latter is also denoted $\Delta$. Then we have metric pairs $(\mathbb{R}^2, d, \mathbb{R}^2_{\geqslant})$ and $(\mathbb{R}^2_{\leqslant}, d, \Delta)$. Similarly, we have the metric pair $(\overline{\mathbb{R}}^2_{\leqslant}, d, \Delta)$.

**Definition 3.3.** Let $\mathbf{Met}_*$ denote the full subcategory of $\mathbf{Met_{pairs}}$ whose objects are of the form $(X, d, \{x_0\})$, which we denote $(X, d, x_0)$. We call $x_0$ the *basepoint* and call $(X, d, x_0)$ a *pointed metric space*. A morphism $f : (X, d, x_0) \to (Y, d', y_0)$ is called a *pointed metric map*.

**Example 3.4.** Let $\mathrm{Int}(\mathbb{R})$ denote the set of intervals in $\mathbb{R}$ with $d(I, J)$ equal to the length (i.e. Lebesgue measure) of the symmetric difference $(I \cup J) \setminus (I \cap J)$. Then $(\mathrm{Int}(\mathbb{R}), d, \emptyset)$ is a pointed metric space. We may also equip $\mathrm{Int}(\mathbb{R})$ with the Hausdorff distance $d_H$ to obtain the pointed metric space $(\mathrm{Int}(\mathbb{R}), d_H, \emptyset)$.

We now show how to obtain a pointed metric space from a metric pair.

**Definition 3.5.** Given a metric pair $(X, d, A)$ consider the quotient set $X/A = (X \setminus A) \amalg \{A\}$. Let $\overline{d} : X/A \times X/A \to [0, \infty]$ be the induced metric. That is, $\overline{d}(A, A) = 0$, for $x \in X \setminus A$, $\overline{d}(x, A) = \overline{d}(A, x) = d(x, A)$, where $d(x, A) = \inf_{y \in A} d(x, y)$, and for $x, y \in X \setminus A$, $\overline{d}(x, y) = \min(d(x, y), d(x, A) + d(y, A))$.

Given a morphism $f : (X, d_X, A) \to (Y, d_Y, B)$, let $\overline{f} : X/A \to Y/B$ be the induced map. That is, $\overline{f}(A) = B$, and for $x \in X \setminus A$, $\overline{f}(x) = B$ if $f(x) \in B$ and otherwise $\overline{f}(x) = f(x)$.

There is a natural quotient map of pairs $q : (X, A) \to (X/A, \{A\})$ given by $q(x) = x$ if $x \in X \setminus A$ and $q(x) = A$ if $x \in A$. We will sometimes denote the image of $x$ by $[x]$, but we will drop the brackets when there is no ambiguity.

We will show that this quotient map may be used to define a functor from $\mathbf{Met_{pairs}}$ to $\mathbf{Met}_*$. First, we show that the quotient of a metric pair is indeed a pointed metric space.

**Lemma 3.6.** *If $(X, d, A)$ is a metric pair then $(X/A, \overline{d}, A)$ is a pointed metric space. Moreover, the quotient map $q : (X, d, A) \to (X/A, \overline{d}, A)$ is a metric map.*

**Proof.** We need to show that $\overline{d}$ is a metric. Point triviality and symmetry follow from the definition. It remains to prove the triangle inequality. There are three nontrivial cases.

For the first case, let $x, z \in X \setminus A$. We want to show that $\overline{d}(x, A) \leqslant \overline{d}(x, z) + \overline{d}(z, A)$. Since $\overline{d}(x, z) + \overline{d}(z, A) = \min(d(x, z), d(x, A) + d(z, A)) + d(z, A)$, it suffices to show that $\overline{d}(x, A) \leqslant d(x, z) + d(z, A)$ and $\overline{d}(x, A) \leqslant d(x, A) + 2d(z, A)$. The first inequality holds since $\overline{d}(x, A) = d(x, A) = \inf_{y \in A} d(x, y) \leqslant \inf_{y \in A}(d(x, z) + d(z, y)) = d(x, z) + d(z, A)$, and the second inequality holds trivially since $\overline{d}(x, A) = d(x, A) \leqslant d(x, A) + 2d(z, A)$. Thus $\overline{d}(x, A) \leqslant \overline{d}(x, z) + \overline{d}(z, A)$.

For the second case, let $x, y \in X \setminus A$. Then $\overline{d}(x, y) = \min(d(x, y), d(x, A) + d(A, y)) \leqslant d(x, A) + d(A, y) = \overline{d}(x, A) + \overline{d}(A, y)$.

For the third case, let $x, y, z \in X \setminus A$. We want to show that $\overline{d}(x, y) \leqslant \overline{d}(x, z) + \overline{d}(z, y) = \min(d(x, z), d(x, A) + d(A, z)) + \min(d(z, y), d(z, A) + d(A, y))$. The right hand side has four possible values. First, $\overline{d}(x, y) \leqslant d(x, y) \leqslant d(x, z) + d(z, y)$. Second, $\overline{d}(x, y) \leqslant d(x, A) + d(A, y) \leqslant d(x, z) + d(z, A) + d(A, y)$. Third, $\overline{d}(x, y) \leqslant d(x, A) + d(A, y) \leqslant d(x, A) + d(A, z) + d(z, y)$. Fourth, $\overline{d}(x, y) \leqslant d(x, A) + d(A, y) \leqslant d(x, A) + d(A, z) + d(z, A) + d(A, y)$. Therefore $\overline{d}(x, y) \leqslant \overline{d}(x, z) + \overline{d}(z, y)$.

To prove the second statement, let $x, y \in X$. There are three cases. First, if $x, y \in X \setminus A$ then $\overline{d}(q(x), q(y)) = \min(d(x, y), d(x, A) + d(y, A)) \leqslant d(x, y)$. Second, if $x \in X \setminus A$ and $y \in A$, then $\overline{d}(q(x), q(y)) = d(x, A) \leqslant d(x, y)$. Third, if $x, y \in A$ then $\overline{d}(q(x), q(y)) = \overline{d}(A, A) = 0 \leqslant d(x, y)$. This completes the proof. □

Next, we show that this map sends morphisms in $\mathbf{Met_{pairs}}$ to morphisms in $\mathbf{Met}_*$.

**Lemma 3.7.** *Given a morphism $f : (X, d_X, A) \to (Y, d_Y, B)$ of metric pairs, the induced map $\overline{f} : (X/A, \overline{d}_X, A) \to (Y/B, \overline{d}_Y, B)$ is a pointed metric map.*

**Proof.** We will prove that $\overline{f}$ is a metric map. Let $x \in X \setminus A$. First we show that $d_Y(f(x), B) \leqslant d_X(x, A)$. Indeed, $d_Y(f(x), B) = \inf_{y \in B} d_Y(f(x), y) \leqslant \inf_{x' \in A} d_Y(f(x), f(x')) \leqslant \inf_{x' \in A} d_X(x, x') = d(x, A)$. Then $\overline{d}_Y(\overline{f}(x), \overline{f}(A)) = \overline{d}_Y(f(x), B) = d_Y(f(x), B) \leqslant$

$d_X(x, A) = \overline{d}_X(x, A)$. Next let $x, x' \in X \setminus A$. Then $\overline{d}_Y(\overline{f}(x), \overline{f}(x')) = \overline{d}_Y(f(x), f(x')) = \min(d_Y(f(x), f(x')), d_Y(f(x), B) + d_Y(B, f(x'))) \leqslant \min(d_X(x, x'), d_X(x, A) + d_X(A, x')) = \overline{d}_X(x, x')$. □

With the above lemmas in hand, it is now easy to check that we have a functor.

**Definition 3.8.** Let $Q : \mathbf{Met_{pairs}} \to \mathbf{Met}_*$ be the functor that sends a metric pair $(X, d, A)$ to the pointed metric space $(X/A, \overline{d}, A)$ and that sends $f : (X, d_X, A) \to (Y, d_Y, B)$ to $\overline{f} : (X/A, \overline{d}_X, A) \to (Y/B, \overline{d}_Y, B)$.

**Theorem 3.9.** $\mathbf{Met}_*$ *is a reflective subcategory of* $\mathbf{Met_{pairs}}$ *with left adjoint* $Q$.

**Proof.** By Lemma 2.12, it suffices to show that, for each $X = (X, d_X, A) \in \mathbf{Met_{pairs}}$, there is a morphism $r : (X, d_X, A) \to (X/A, \overline{d}, A) = IQ(X, d_X, A)$ in $\mathbf{Met_{pairs}}$ for which $((X/A, \overline{d}, A), r)$ is an initial object in the comma category $X \downarrow I$. Here, $I : \mathbf{Met}_* \to \mathbf{Met_{pairs}}$ denotes the inclusion functor.

$$\begin{tikzcd} (X, d_X, A) \arrow[r, "r"] \arrow[dr, "f"'] & (X/A, \overline{d}_X, A) \arrow[d, dashed, "\exists! \overline{f}"] \\ & (Y, d_Y, y_0) \end{tikzcd} \tag{3.1}$$

To this end, define $r : (X, d_X, A) \to (X/A, \overline{d}_X, A)$ by $r(x) = A$ if $x \in A$ and $r(x) = x$ otherwise. Let us show that $r$ is a metric map. Let $x, x' \in X$. If $x, x' \notin A$ then $\overline{d}_X(rx, rx') = \overline{d}_X(x, x') = \min(d_X(x, x'), d_X(x, A) + d_X(A, x)) \leqslant d_X(x, x')$. If $x \notin A$ and $x' \in A$ then $\overline{d}_X(rx, rx') = \overline{d}_X(x, A) = d_X(x, A) \leqslant d_X(x, x')$.

To see that $r$ is a universal element, let $f : (X, d_X, A) \to (Y, d_Y, y_0) \in \mathbf{Met_{pairs}}$ be given. We want to show that there is a unique metric map $\overline{f} : (X/A, \overline{d}_X, A) \to (Y, d_y, y_0)$ such that $\overline{f} \circ r = f$. By the commutativity of (3.1), we are forced to define $\overline{f}(A) = y_0$ and $\overline{f}(x) = f(x)$ if $x \notin A$. This demonstrates uniqueness. To establish existence, it remains to show that $\overline{f}$ is a metric map. If $x \in X \setminus A$ then $d_Y(\overline{f}(x), \overline{f}(A)) = d_Y(f(x), y_0) = d_Y(f(x), f(x')) \leqslant d_X(x, x')$ for all $x' \in A$. Thus $d_Y(\overline{f}(x), \overline{f}(A)) \leqslant d_X(x, A) = \overline{d}_X(x, A)$. If $x, x' \in X \setminus A$ then $d_Y(\overline{f}(x), \overline{f}(x')) = d_Y(f(x), f(x')) \leqslant d_X(x, x')$. Furthermore $d_Y(\overline{f}(x), \overline{f}(x')) = d_Y(f(x), f(x')) \leqslant d_Y(f(x), y_0) + d_Y(y_0, f(x')) \leqslant d_X(x, A) + d_X(A, x')$. Therefore $d_Y(\overline{f}(x), \overline{f}(x')) \leqslant \overline{d}_X(x, x')$. □

### *3.2. The p-strengthened triangle inequality*

In this section, we introduce a convenient class of pointed metric spaces for each $p \in [1, \infty]$. These are pointed metric spaces which satisfy a slightly stronger version of the triangle inequality with respect the basepoint.

**Definition 3.10.** Let $(X, d)$ be a metric space, $x_0 \in X$, and $p \in [1, \infty]$. We say that the metric $d$ satisfies the *$p$-strengthened triangle inequality with respect to* $x_0$ if $d(x, y) \leqslant \left\|(d(x, x_0), d(x_0, y))\right\|_p$ for all $x, y \in X$. Let $\mathbf{Met}_*^{\mathbf{p}}$ denote the full subcategory of $\mathbf{Met}_*$ consisting of those objects $(X, d, x_0)$ for which $(X, d)$ satisfies the $p$-strengthened triangle inequality with respect to $x_0$.

Note that the 1-strengthened triangle inequality is just the triangle inequality. So $\mathbf{Met}_*^{\mathbf{1}} = \mathbf{Met}_*$. Also, for $1 \leqslant p \leqslant q \leqslant \infty$, $\mathbf{Met}_*^{\mathbf{q}}$ is a full subcategory of $\mathbf{Met}_*^{\mathbf{p}}$.

**Example 3.11.** Let $*$ denote the singleton set. Also let $*$ denote the pointed metric space $(*, 0, *)$. Then for all $p \in [1, \infty]$, $* \in \mathbf{Met}_*^{\mathbf{p}}$. In fact, it is the initial and terminal object in $\mathbf{Met}_*^{\mathbf{p}}$.

From any metric space $(X, d)$ and given basepoint $x_0 \in X$, we can obtain a metric that satisfies the $p$-strengthened inequality. The following definition is a more general construction.

**Definition 3.12.** Let $(X, d, A)$ be a metric pair and let $p \in [1, \infty]$. Define $d_p : X \times X \to [0, \infty]$ by

$$d_p(x, y) := \min(d(x, y), \left\|(d(x, A), d(A, y))\right\|_p).$$

In the special case that $A = \{x_0\}$ is a singleton, it is clear from the definition that $d_p$ satisfies the $p$-strengthened triangle inequality with respect to $x_0$. We still need to verify that $d_p$ is actually a metric.

**Lemma 3.13.** *Let* $p \in [1, \infty]$. *If* $(X, d, A)$ *is a metric pair then so is* $(X, d_p, A)$.

**Proof.** Point triviality and symmetry follow from the definition. To show the triangle inequality, let $x, y, z \in X$. We want to show that $\min(d(x, y), \|(d(x, A), d(A, y))\|_p) \leqslant \min(d(x, z), \|(d(x, A), d(A, z))\|_p) + \min(d(z, y), \|(d(z, A), d(A, y))\|_p)$. The right hand side has four possible values. First, $d(x, y) \leqslant d(x, z) + d(z, y)$. Second,

$$\begin{aligned}\|(d(x, A), d(A, y))\|_p &\leqslant \|(d(x, z) + d(z, A), d(A, y))\|_p \\ &= \|(d(x, z), 0) + (d(z, A), d(A, y))\|_p \leqslant d(x, z) + \|(d(z, A), d(A, y))\|_p .\end{aligned}$$

Third,

$$\begin{aligned}\|(d(x, A), d(A, y))\|_p &\leqslant \|(d(x, A), d(A, z) + d(z, y))\|_p \\ &= \|(d(x, A), d(A, z)) + (0, d(z, y))\|_p \leqslant \|(d(x, A), d(A, z))\|_p + d(z, y).\end{aligned}$$

Fourth,

$$\begin{aligned}\|(d(x, A), d(A, y))\|_p &\leqslant \|(d(x, A), d(A, z), d(z, A), d(A, y))\|_p \\ &= \left\|(\|(d(x, A), d(A, z))\|_p , \|(d(z, A), d(A, y))\|_p)\right\|_p \\ &\leqslant \|(d(x, A), d(A, z))\|_p + \|(d(z, A), d(A, y))\|_p .\end{aligned}$$

The result now follows from these inequalities. □

**Lemma 3.14.** *Let $p \in [1, \infty]$. If $f : (X, d, x_0) \to (Y, d', y_0)$ is a pointed metric map, then so is $f : (X, d_p, x_0) \to (Y, d'_p, y_0)$.*

**Proof.** Let $x, x' \in X$. Then

$$\begin{aligned}d'_p(f(x), f(x')) &= \min(d'(f(x), f(x')), \left\|(d'(f(x), y_0), d'(y_0, f(x')))\right\|_p) \\ &= \min(d'(f(x), f(x')), \left\|(d'(f(x), f(x_0)), d'(f(x_0), f(x')))\right\|_p) \\ &\leqslant \min(d(x, x'), \left\|(d(x, x_0), d(x_0, x'))\right\|_p) = d_p(x, x'). \quad \square\end{aligned}$$

The operation which sends a pointed metric space $(X, d, x_0)$ to $(X, d_p, x_0)$ is easily seen to be functorial.

**Definition 3.15.** Let $1 \leqslant p \leqslant q \leqslant \infty$. Let $S_{p,q} : \mathbf{Met}_*^{\mathbf{p}} \to \mathbf{Met}_*^{\mathbf{q}}$ be the functor that sends $(X, d, x_0)$ to $(X, d_q, x_0)$ and $f : (X, d, x_0) \to (Y, d', y_0)$ to $f : (X, d_q, x_0) \to (Y, d'_q, y_0)$. We will also denote $S_{1,p}$ by $S_p$.

**Theorem 3.16.** *Let $1 \leqslant p \leqslant q \leqslant \infty$. $\mathbf{Met}_*^{\mathbf{q}}$ is a reflective subcategory of $\mathbf{Met}_*^{\mathbf{p}}$ with left adjoint $S_{p,q}$. As a special case, $\mathbf{Met}_*^{\mathbf{p}}$ is a reflective subcategory of $\mathbf{Met}_*$ with left adjoint $S_p$.*

**Proof.** Let $(X, d, x_0) \in \mathbf{Met}_*^{\mathbf{p}}$, $(Y, d', y_0) \in \mathbf{Met}_*^{\mathbf{q}}$ and $f : (X, d, x_0) \to (Y, d', y_0) \in \mathbf{Met}_*^{\mathbf{p}}$.

$$\begin{tikzcd} (X, d, x_0) \arrow[r, "r"] \arrow[dr, "f"'] & (X, d_q, x_0) \arrow[d, dashed, "\exists! \overline{f}"] \\ & (Y, d', y_0) \end{tikzcd} \tag{3.2}$$

Let $r : (X, d, x_0) \to (X, d_q, x_0) \in \mathbf{Met}_*^{\mathbf{p}}$ be the identity function on $X$. We will show that $((X, d_q, x_0), r)$ is a universal element in the comma category $X \downarrow I$, where $I$ denotes the inclusion $\mathbf{Met}_*^{\mathbf{q}} \hookrightarrow \mathbf{Met}_*^{\mathbf{p}}$. Note that $r$ is a metric map since for all $x, x' \in X$, $d_q(x, x') \leqslant d(x, x')$. Let $f : (X, d, A) \to (Y, d', y_0)$ be a pointed metric map. By the commutativity of (3.2), we are forced to define $\overline{f} = f$. This establishes uniqueness. To establish existence, it remains to show that $\overline{f}$ is a metric map. Let $x, x' \in X$. We have $d'(f(x), f(x')) \leqslant d(x, x')$ and, since $(Y, d', y_0) \in \mathbf{Met}_*^{\mathbf{q}}$,

$$\begin{aligned}d'(f(x), f(x')) &\leqslant \left\|(d'(f(x), y_0), d'(y_0, f(x')))\right\|_q \\ &= \left\|(d'(f(x), f(x_0)), d'(f(x_0), f(x')))\right\|_q \leqslant \left\|(d(x, x_0), d(x_0, x'))\right\|_q .\end{aligned}$$

Therefore $d'(\overline{f}(x), \overline{f}(x')) \leqslant d_q(x, x')$. It follows by Lemma 2.12 that $S_{p,q}$ is left adjoint to the inclusion. □

By Lemmas 3.6 and 3.13, we have the following.

**Lemma 3.17.** *Given a metric $(X, d, A)$ and $p \in [1, \infty]$, we have the pointed metric space $(X/A, \overline{d}_p, A)$ where $\overline{d}_p$ is given by $\overline{d}_p(A, A) = 0$, for $x \in X \setminus A$, $\overline{d}_p(x, A) = d(x, A) = \overline{d}_p(A, x)$, and for $x, y \in X \setminus A$, $\overline{d}_p(x, y) = \min(d(x, y), \|d(x, A), d(A, y)\|_p)$.*

## 4. Commutative metric monoids and Wasserstein distance

In this section, we introduce $p$-subadditive commutative metric monoids. These are metric spaces which are also monoids and for which the monoid operation is, in a precise sense, compatible with the metric. The category of $p$-subadditive commutative metric monoids is the setting in which we state our universality results for the Wasserstein distances.

### 4.1. Commutative metric monoids

In this section, we introduce $p$-subadditive commutative metric monoids which are metric spaces that are simultaneously commutative monoids and which satisfy a certain compatibility condition between the metric and the monoid operation.

**Definition 4.1.** Let $p \in [1, \infty]$. A *$p$-subadditive commutative metric monoid* is a tuple $(M, d, +, 0)$ where $(M, d, 0)$ is a pointed metric space and $(M, +, 0)$ is a commutative monoid such that for all $a, b, a', b' \in M$,

$$d(a+b, a'+b') \leqslant \left\| (d(a,a'), d(b,b')) \right\|_p .$$

In this case, we say that the metric $d$ is *$p$-subadditive*. A *morphism of $p$-subadditive commutative metric monoids* $f : (M, d, +, 0) \to (N, \rho, +, 0)$ is a pointed metric map $f : (M, d, 0) \to (N, \rho, 0)$ such that $f : (M, +, 0) \to (N, +, 0)$ is a monoid homomorphism. Call such a map a *metric monoid homomorphism*. Let $\mathbf{CMetMon_p}$ denote the category of $p$-subadditive commutative metric monoids and metric monoid homomorphisms.

**Lemma 4.2.** *Let $p \in [1, \infty]$. Let $(M, d, +, 0)$ be a $p$-subadditive commutative metric monoid. Then for $n \geqslant 0$ and $a_1, \ldots, a_n, b_1, \ldots, b_n \in M$,*

$$d(a_1 + \cdots + a_n, b_1 + \cdots + b_n) \leqslant \left\| (d(a_i, b_i))_{i=1}^n \right\|_p .$$

**Proof.** The proof is by induction on $n$.

$$\begin{aligned} d(a_1 + \cdots + a_{n+1}, b_1 + \cdots + b_{n+1}) &\leqslant \| (d(a_1 + \cdots + a_n, b_1 + \cdots + b_n), d(a_{n+1}, b_{n+1})) \|_p \\ &\leqslant \left\| \left( \left\| (d(a_i, b_i))_{i=1}^n \right\|_p, d(a_{n+1}, b_{n+1}) \right) \right\|_p = \left\| (d(a_i, b_i))_{i=1}^{n+1} \right\|_p . \quad \square \end{aligned}$$

**Corollary 4.3.** *Let $p \in [1, \infty]$. Let $(M, d, +, 0)$ be a $p$-subadditive commutative metric monoid. Then for $n \geqslant 0$ and $a_1, \ldots, a_n, b_1, \ldots, b_n \in M$,*

$$d(a_1 + \cdots + a_n, b_1 + \cdots + b_n) \leqslant \min_{\sigma \in \Sigma_n} \left\| (d(a_i, b_{\sigma(i)}))_{i=1}^n \right\|_p ,$$

*where $\Sigma_n$ denotes the symmetric group on $n$ symbols.*

The following lemma shows that there is a forgetful functor $U_p : \mathbf{CMetMon_p} \to \mathbf{Met}_*^{\mathbf{p}}$.

**Lemma 4.4.** *Let $p \in [1, \infty]$. Let $(M, d, +, 0)$ be a $p$-subadditive commutative metric monoid. Then $(M, d, 0) \in \mathbf{Met}_*^{\mathbf{p}}$. Furthermore if $f : (M, d, +, 0) \to (N, \rho, +, 0)$ is a morphism of commutative metric monoids then $f : (M, d, 0) \to (N, \rho, 0) \in \mathbf{Met}_*^{\mathbf{p}}$.*

**Proof.** For the first statement, we show that $(M, d, 0)$ satisfies the $p$-strengthened triangle inequality at 0. Let $a, b \in M$. Then $d(a, b) = d(a+0, 0+b) \leqslant \| (d(a,0), d(0,b)) \|_p$. The second statement follows directly from the definitions. $\square$

### 4.2. Monoid objects in $\mathbf{Met}_*^{\mathbf{p}}$

In this section, we show that $\mathbf{Met}_*^{\mathbf{p}}$ can be equipped with a tensor product making it into a symmetric monoidal category. We then show that $p$-subadditive commutative metric monoids are precisely the commutative monoids internal to this symmetric monoidal category.

**Definition 4.5.** Let $p \in [1, \infty]$. Given pointed metric spaces $(X, d_X, x_0)$ and $(Y, d_Y, y_0)$, define $d_X \times_p d_Y : (X \times Y) \times (X \times Y) \to [0, \infty]$ by

$$(d_X \times_p d_Y)\big((x, y), (x', y')\big) = \left\| (d_X(x, x'), d_Y(y, y')) \right\|_p .$$

We call $d_X \times_p d_Y$ the *$p$-product metric*. Let $X \times_p Y$ denote the tuple $(X \times Y, d_X \times_p d_Y, (x_0, y_0))$.

The following lemma shows that $d_X \times_p d_Y$ defines a metric on the product $X \times Y$.

**Lemma 4.6.** *Let $p \in [1,\infty]$. If $X = (X, d_X, x_0)$ and $Y = (Y, d_Y, y_0)$ are pointed metric spaces then so is $X \times_p Y$.*

**Proof.** We show that $d_X \times_p d_Y$ is a metric for $X \times Y$. Point triviality and symmetry follow from the corresponding properties for $d_X$ and $d_Y$. It remains to prove the triangle inequality. For all $(x, y), (x', y'), (x'', y'') \in X \times Y$,

$$\begin{aligned}\left\|(d_X(x,x'), d_Y(y,y'))\right\|_p &\leqslant \left\|(d_X(x,x'') + d_X(x'',x'), d_Y(y,y'') + d_Y(y'',y'))\right\|_p \\ &= \left\|(d_X(x,x''), d_Y(y,y'')) + (d_X(x'',x'), d_Y(y'',y'))\right\|_p \\ &\leqslant \left\|(d_X(x,x''), d_Y(y,y''))\right\|_p + \left\|(d_X(x'',x'), d_Y(y'',y'))\right\|_p . \quad \square\end{aligned}$$

The product metric $d_X \times_p d_Y$ can be used to give a succinct and convenient description of $p$-subadditive commutative metric monoids.

**Lemma 4.7.** *A* $p$-subadditive commutative metric monoid *is a tuple $(M, d, +, 0)$ where $(M, d, 0)$ is a pointed metric space and $(M, +, 0)$ is a commutative monoid such that $+ : M \times_p M \to M$ is a metric map.*

**Proof.** Consider $(M, d, +, 0)$ where $(M, d, 0)$ is a pointed metric space and $(M, +, 0)$ is a commutative monoid. $+ : M \times_p M \to M$ is a metric map if and only if for all $(a, b), (a', b') \in M \times M$, $d(a + b, a' + b') \leqslant \left\|d(a, a'), d(b, b')\right\|_p$. $\square$

The following lemma shows that $\mathbf{Met}_*^{\mathbf{p}}$ is closed with respect to forming $p$-product metrics.

**Lemma 4.8.** *Let $p \in [1,\infty]$. If $X = (X, d_X, x_0)$, $Y = (Y, d_Y, y_0) \in \mathbf{Met}_*^{\mathbf{p}}$ then $X \times_p Y \in \mathbf{Met}_*^{\mathbf{p}}$.*

**Proof.** We show that $d_X \times_p d_Y$ satisfies the $p$-strengthened triangle inequality with respect to $(x_0, y_0)$. For all $(x, y), (x', y') \in X \times Y$,

$$\begin{aligned}\left\|(d_X(x,x'), d_Y(y,y'))\right\|_p &\leqslant \left\|(\left\|(d_X(x,x_0), d_X(x_0,x'))\right\|_p, \left\|(d_Y(y,y_0), d_Y(y_0,y'))\right\|_p)\right\|_p \\ &= \left\|(d_X(x,x_0), d_X(x_0,x'), d_Y(y,y_0), d_Y(y_0,y'))\right\|_p \\ &= \left\|(\|(d_X(x,x_0), d_Y(y,y_0))\|_p, \left\|(d_X(x_0,x'), d_Y(y_0,y'))\right\|_p)\right\|_p . \quad \square\end{aligned}$$

We want to show that $(\mathbf{Met}_*^{\mathbf{p}}, \times_p, *)$ is a symmetric monoidal category. We will first show that $(\mathbf{Met}_*, \times_p, *)$ is a symmetric monoidal category.

**Proposition 4.9.** *For each $p \in [1,\infty]$, $(\mathbf{Met}_*, \times_p, *)$ is a symmetric monoidal category.*

**Proof.** For the associator, consider $X, Y, Z \in \mathbf{Met}_*$ and $x, x' \in X$, $y, y' \in Y$ and $z, z' \in Z$. Then

$$\begin{aligned}((d_X \times_p d_Y) \times_p d_Z)(((x,y),z),((x',y'),z)) &= \left\|((d_X \times_p d_Y)((x,y),(x',y')), d_Z(z,z'))\right\|_p \\ &= \left\|(\left\|(d_X(x,x'), d_Y(y,y'))\right\|_p, d_Z(z,z'))\right\|_p = \left\|(d_X(x,x'), d_Y(y,y'), d_Z(z,z'))\right\|_p .\end{aligned}$$

The left unitor is an isometry since for $X \in \mathbf{Met}_*$ and $x, x' \in X$, $\left\|(d_X(x,x'), 0)\right\|_p = d_X(x,x')$. Similarly, the right unitor is an isometry. The braiding is given by the obvious isometry $X \times_p Y \cong Y \times_p X$. With these computations in hand, the rest of the axioms are easy to check. $\square$

**Corollary 4.10.** *For each $p \in [1,\infty]$, $(\mathbf{Met}_*^{\mathbf{p}}, \times_p, *)$ is a symmetric monoidal category, which we denote by $\mathbf{Met}_*^{\mathbf{p}}$.*

**Proof.** This follows immediately since $\mathbf{Met}_*^{\mathbf{p}}$ is a subcategory of $\mathbf{Met}_*$ which, by Lemma 4.8, is closed under the tensor product $\times_p$. $\square$

Let $\mathbf{CMon}(\mathbf{Met}_*, \times_{\mathbf{p}}, *)$ and $\mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$ denote the categories of commutative monoids internal to the symmetric monoidal categories $(\mathbf{Met}_*, \times_p, *)$ and $\mathbf{Met}_*^{\mathbf{p}}$, respectively. Recall that $\mathbf{CMetMon_p}$ denotes the category of $p$-subadditive commutative metric monoids. The following proposition shows that $\mathbf{CMetMon_p}$ is precisely the category of commutative monoids internal to the $\mathbf{Met}_*^{\mathbf{p}}$. Moreover, we show that $\mathbf{CMon}(\mathbf{Met}_*, \times_{\mathbf{p}}, *)$ and $\mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$ are in fact the same.

**Theorem 4.11.** *A commutative monoid in the symmetric monoidal category* $\mathbf{Met}_*^{\mathbf{p}}$ *is a* $p$*-subadditive commutative metric monoid and a morphism of commutative monoids in* $\mathbf{Met}_*^{\mathbf{p}}$ *is morphism of* $p$*-subadditive commutative metric monoids. That is,* $\mathbf{CMetMon_p} = \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$. *Moreover,* $\mathbf{CMon}(\mathbf{Met}_*, \times_{\mathbf{p}}, *) = \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$.

**Proof.** A commutative monoid in $\mathbf{Met}_*^{\mathbf{p}}$ is a pointed metric space $(M, d, m_0) \in \mathbf{Met}_*^{\mathbf{p}}$ together with a binary operation $+ : M \times M \to M$ that is associative, commutative, and for which $m_0$ is a unit, such that $+ : M \times_p M \to M \in \mathbf{Met}_*^{\mathbf{p}}$. That is, $+$ is $p$-subadditive. Thus a commutative monoid in $\mathbf{Met}_*^{\mathbf{p}}$ is a $p$-subadditive commutative metric monoid.

A morphism $f : (M, d, +_M, m_0) \to (N, \rho, +_N, n_0)$ of commutative monoids in $\mathbf{Met}_*^{\mathbf{p}}$ is a morphism $f : (M, d, m_0) \to (N, \rho, n_0) \in \mathbf{Met}_*^{\mathbf{p}}$ such that for all $a, b \in M$, $f(a +_M b) = f(a) +_N f(b)$ and $f(m_0) = n_0$. That is, $f : (M, d, m_0) \to (N, \rho, n_0) \in \mathbf{Met}_*$ such that $f : (M, +_M, m_0) \to (N, +_N, n_0)$ is a monoid homomorphism.

To see that $\mathbf{CMon}(\mathbf{Met}_*, \times_{\mathbf{p}}, *) = \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$, note that it suffices to require that $(M, d, m_0) \in \mathbf{Met}_*$, since the unit condition and $p$-subadditivity implies that for $a, b \in M$, $d(a, b) = d(a + m_0, m_0 + b) \leqslant \|(d(a, m_0), d(m_0, b))\|_p$. $\square$

*4.3. Wasserstein distance*

In this section, we introduce the $p$-Wasserstein distance $W_p$ on the space of diagrams $D(X, A)$ on a metric pair, and show that $(D(X, A), W_p)$, taken together with the monoid structure on $D(X, A)$, forms a $p$-subadditive commutative metric monoid.

Given a set pair $(X, A)$, recall that $D(X, A) = D(X)/D(A)$. As a special case, for a pointed set $(X, x_0)$, $D(X, x_0) = D(X)/D(x_0)$.

**Definition 4.12.** Let $p \in [1, \infty]$. Given a metric pair $(X, d, A)$ define $W_p[d, A] : D(X, A) \times D(X, A) \to [0, \infty]$ by

$$W_p[d, A](x_1 + \cdots + x_m, x'_1 + \cdots + x'_n) = \inf \left\| (d(x_k, x'_{\sigma(k)}))_{k=1}^{m+n} \right\|_p,$$

where the infimum is taken over $x_{m+1}, \ldots, x_{m+n}, x'_{n+1}, \ldots x'_{n+m} \in A$ and $\sigma \in \Sigma_{m+n}$, where $\Sigma_{m+n}$ denotes the symmetric group on $m + n$ symbols.

One may check that Definition 4.12 may be restated as follows.

**Lemma 4.13.**

$$W_p[d, A](x_1 + \cdots + x_m, x'_1 + \cdots + x'_n) = \min_{\sigma \in \Sigma_{m+n}} \left\| (d(x_k, x'_{\sigma(k)}))_{k=1}^{m+n} \right\|_p,$$

*where* $x_{m+1} = \cdots = x_{m+n} = A = x'_{n+1} = \cdots x'_{n+m}$ *and* $d(x, A) = \inf_{a \in A} d(x, a)$.

For brevity, we will sometimes denote $W_p[d, A]$ by $W_p$ when this can lead to no confusion.

**Remark 4.14.** Recall that for a metric pair $(X, d, A)$, $(X/A, \overline{d}, A)$ denotes the pointed metric space obtained by collapsing $A$ to a point (see Definition 3.5). Then $D(X, A) \cong D(X \setminus A) \cong D(X/A, A)$. Explicitly, we have monoid isomorphisms $\varphi : D(X \setminus A) \to D(X, A)$ and $\psi : D(X \setminus A) \to D(X/A, A)$ given by $x_1 + \cdots + x_n \mapsto x_1 + \cdots + x_n + D(A)$ and $x_1 + \cdots + x_n \mapsto [x_1] + \cdots + [x_n] + D(A)$, respectively. By Lemma 4.24 below, we have that $(D(X, A), W_p[d, A], +, 0)$ and $(D(X/A, A), W_p[\overline{d}_p, A], +, 0)$ are isometrically isomorphic, and so we may pass between the settings of metric pairs and pointed metric spaces whenever convenient.

**Example 4.15.** For the metric pair $(\mathbb{R}^2_\leqslant, d, \Delta)$ or $(\overline{\mathbb{R}}^2_\leqslant, d, \Delta)$ (Example 3.2) and $p \in [1, \infty]$, $W_p[d, \Delta]$ is the $p$-Wasserstein distance on (finite) persistence diagrams. For the metric pair $(\mathrm{Int}(\mathbb{R}), d, \emptyset)$ and $d$ the length of the symmetric difference (Example 3.4), $W_1[d, \emptyset]$ is the barcode metric.

The following lemma verifies that $W_p$ is indeed a metric on $D(X, x_0)$.

**Lemma 4.16.** *Let* $p \in [1, \infty]$. *If* $(X, d, x_0)$ *is a pointed metric space then* $(D(X, x_0), W_p, 0)$ *is a pointed metric space.*

**Proof.** Point triviality and symmetry follow from the definition. To prove the triangle inequality, let $\alpha = x_1 + \cdots + x_n$, $\beta = x'_1 + \cdots + x'_m$, $\gamma = x''_1 + \cdots + x''_p$, be elements of $D(X, x_0)$. Let $r = n + m + p$ and let $x_{n+1} = \cdots = x_r = x'_{m+1} = \cdots = x'_r = x''_{p+1} = \cdots = x''_r = x_0$. Let $\sigma$, $\tau \in S_r$ be permutations realizing $W_p(\alpha, \gamma)$, $W_p(\gamma, \beta)$, respectively. Let $\pi = \tau \circ \sigma \in S_r$. Then

$$
\begin{aligned}
W_p(\alpha,\beta) &\leqslant \left\| \left(d(x_k, x'_{\pi(k)})\right)_{k=1}^r \right\|_p \\
&\leqslant \left\| \left(d(x_k, x''_{\sigma(k)})\right)_{k=1}^r + \left(d(x''_{\sigma(k)}, x'_{\pi(k)})\right)_{k=1}^r \right\|_p \\
&\leqslant \left\| \left(d(x_k, x''_{\sigma(k)})\right)_{k=1}^r \right\|_p + \left\| \left(d(x''_{\sigma(k)}, x'_{\pi(k)})\right)_{k=1}^r \right\|_p \\
&= \left\| \left(d(x_k, x''_{\sigma(k)})\right)_{k=1}^r \right\|_p + \left\| \left(d(x''_{\sigma(\sigma^{-1}(\ell))}, x'_{\pi(\sigma^{-1}(\ell))})\right)_{\ell=1}^r \right\|_p \\
&= \left\| \left(d(x_k, x''_{\sigma(k)})\right)_{k=1}^r \right\|_p + \left\| \left(d(x''_{\ell}, x'_{\tau(\ell)})\right)_{\ell=1}^r \right\|_p = W_p(\alpha,\gamma) + W_p(\gamma,\beta). \quad \square
\end{aligned}
$$

By Remark 4.14, the preceding lemma shows that $W_p[d, A]$ is a metric on $D(X, A)$ for any metric pair $(X, d, A)$ and any $p \in [1, \infty]$.

Next, we show that $W_p$ is $p$-subadditive.

**Lemma 4.17.** *Let $p \in [1, \infty]$. If $(X, d, x_0)$ is a pointed metric space then $(D(X, x_0), W_p, +, 0)$ is a $p$-subadditive commutative metric monoid and hence is an object in* $\mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$.

**Proof.** Let $\alpha, \beta, \gamma, \delta \in D(X, x_0)$, where $\alpha = x_1 + \cdots + x_m$, $\beta = x_{m+n+1} + \cdots + x_{m+n+p}$, $\gamma = x'_1 + \cdots + x'_n$, $\delta = x'_{m+n+1} + \cdots + x'_{m+n+q}$. We want to show that $W_p(\alpha+\beta, \gamma+\delta) \leqslant \left\| (W_p(\alpha,\gamma), W_p(\beta,\delta)) \right\|_p$. Let $x_{m+1} = \cdots = x_{m+n} = x_{m+n+p+1} = x_{m+n+p+q} = x'_{n+1} = \cdots = x'_{m+n} = x'_{m+n+q+1} = x'_{m+n+p+q} = x_0$. Given $\sigma \in \Sigma_{m+n}$ and $\tau \in \Sigma_{p+q}$, let $\sigma * \tau \in \Sigma_{m+n+p+q}$ be defined by $\sigma * \tau(i) := \sigma(i)$ if $i \leqslant m+n$ and $\sigma * \tau(i) := \tau(i)$ otherwise. Then

$$
\begin{aligned}
\left\| (W_p(\alpha,\gamma), W_p(\beta,\delta)) \right\|_p &= \left\| \left( \min_{\sigma \in \Sigma_{m+n}} \left\| (d(x_i, x'_{\sigma(i)}))_{i=1}^{m+n} \right\|_p, \min_{\tau \in \Sigma_{p+q}} \left\| (d(x_{m+n+i}, x'_{m+n+\tau(i)})) \right\|_p \right) \right\|_p \\
&= \min_{\sigma \in \Sigma_{m+n}} \min_{\tau \in \Sigma_{p+q}} \left\| (d(x_i, x'_{\sigma*\tau(i)}))_{i=1}^{m+n+p+q} \right\|_p \\
&\geqslant \min_{\pi \in \Sigma_{m+n+p+q}} \left\| (d(x_i, x'_{\pi(i)}))_{i=1}^{m+n+p+q} \right\|_p = W_p(\alpha+\beta, \gamma+\delta). \quad \square
\end{aligned}
$$

**Lemma 4.18.** *Let $p \in [1, \infty]$. Let $(X, d, x_0) \in \mathbf{Met}_*^{\mathbf{p}}$. Then the inclusion map $i : (X, d, x_0) \hookrightarrow (D(X, x_0), W_p[d, x_0], 0)$ is an isometry (and hence a metric map).*

**Proof.** Let $x, y \in X$. $W_p[d, x_0](x, y) = \min\left(d(x, y), \|(d(x, x_0), d(x_0, y))\|_p\right) = d(x, y)$. $\square$

The preceding lemma shows that if $(X, d, x_0) \in \mathbf{Met}_*^{\mathbf{p}}$, then $i^* W_p = d$. On the other hand, if $(X, d, x_0) \in \mathbf{Met}_*$ but $d$ does not satisfy the $p$-strengthened inequality with respect to $x_0$, then the inclusion $i : (X, d, x_0) \hookrightarrow (D(X, x_0), W_p, x_0)$ is only guaranteed to be 1-Lipschitz, but is not in general an isometry. This is one reason for working with $\mathbf{Met}_*^{\mathbf{p}}$ as opposed to just $\mathbf{Met}_*$.

The following lemma shows that for a pointed metric map $f : (X, d_X, x_0) \to (Y, d_Y, y_0)$, the induced map $f_* : D(X, x_0) \to D(Y, y_0)$ is a metric map with respect to the Wasserstein distances.

**Lemma 4.19.** *Let $p \in [1, \infty]$. Given a pointed metric map $f : (X, d, x_0) \to (Y, d', y_0)$, the induced map $f_* : D(X, x_0) \to D(Y, y_0)$ is a morphism of $p$-subadditive commutative metric monoids $f_* : (D(X, x_0), W_p[d, x_0], +, 0) \to (D(Y, y_0), W_p[d', y_0], +, 0)$.*

**Proof.** By Definition 4.1, we need to show that $f_* : (D(X, x_0), W_p[d, x_0]) \to (D(Y, y_0), W_p[d', y_0])$ is a metric map and that $f_* : (D(X, x_0), +, 0) \to (D(Y, y_0), +, 0)$ is a monoid homomorphism. The latter is true by the definition of $f_*$ (Definition 2.2). Let $x_1 + \cdots + x_m, x'_1 + \cdots + x'_n \in D(X, x_0)$. Let $x_{m+1} = \cdots = x_{m+n} = x_0 = x'_{n+1} = \cdots = x'_{n+m}$ and thus $f(x_{m+1}) = \cdots = f(x_{m+n}) = y_0 = f(x'_{n+1}) = \cdots = f(x'_{n+m})$. Then

$$
\begin{aligned}
W_p[d', y_0](f(x_1) + \cdots + f(x_m), f(x'_1) + \cdots + f(x'_n)) &= \min_{\sigma \in \Sigma_{m+n}} \left\| (d'(f(x_i), f(x'_{\sigma(i)})))_{i=1}^{m+n} \right\|_p \\
&\leqslant \min_{\sigma \in \Sigma_{m+n}} \left\| (d(x_i, x'_{\sigma(i)}))_{i=1}^{m+n} \right\|_p = W_p[d, x_0](x_1 + \cdots + x_m, x'_1 + \cdots + x_n). \quad \square
\end{aligned}
$$

From the preceding lemmas, it is easy to see that the assignment that sends $(X, d, x_0) \in \mathbf{Met}_*^{\mathbf{p}}$ to $(D(X, x_0), W_p, +, 0)$ and that sends a pointed metric map $f : (X, d_X, x_0) \to (Y, d_Y, y_0)$ to the induced map $f_* : D(X, x_0) \to D(Y, y_0)$ is functorial.

**Definition 4.20.** Let $p \in [1, \infty]$. Let $D_p : \mathbf{Met}_*^{\mathbf{p}} \to \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$ be the functor given by sending $(X, d, x_0)$ to $(D(X, x_0), W_p, +, 0)$ and $f : (X, d, x_0) \to (Y, d', y_0)$ to $f_* : (D(X, x_0), W_p[d, x_0], +, 0) \to (D(Y, y_0), W_p[d', y_0], +, 0)$.

Recall that there is a forgetful functor $U_p : \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}}) \to \mathbf{Met}_*^{\mathbf{p}}$ given by sending $(M, d, +, 0) \in \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$ to $(M, d, 0) \in \mathbf{Met}_*^{\mathbf{p}}$ (see Lemma 4.4 and Theorem 4.11).

**Theorem 4.21.** *Let $p \in [1, \infty]$. The forgetful functor $U_p : \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}}) \to \mathbf{Met}_*^{\mathbf{p}}$ has left adjoint $D_p$.*

**Proof.** Let $(X, d, x_0) \in \mathbf{Met}_*^{\mathbf{p}}$, $(N, \rho, +, 0) \in \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$ and $\varphi : (X, d, x_0) \to (N, \rho, 0) \in \mathbf{Met}_*^{\mathbf{p}}$.

$$\begin{tikzcd} (X,d,x_0) \arrow[r, "i"] \arrow[dr, "\varphi"'] & (D(X,x_0), W_p[d,x_0], 0) \arrow[d, dashed, "\tilde{\varphi}"] & (D(X,x_0), W_p[d,x_0], +, 0) \arrow[d, dashed, "\exists! \tilde{\varphi}"] \\ & (N,\rho,0) & (N,\rho,+,0) \end{tikzcd} \tag{4.1}$$

By the commutativity of the left hand side of (4.1), we have that for all $x \in X$, $\tilde{\varphi}(x) = \tilde{\varphi}(i(x)) = \varphi(x)$. For $\tilde{\varphi}$ to be a monoid homomorphism, we have $\tilde{\varphi}(x_1 + \cdots + x_n) = \varphi(x_1) + \cdots + \varphi(x_n)$. Thus, if $\tilde{\varphi}$ exists it is unique. It remains to show that $\tilde{\varphi}$ is a metric map.

Let $\alpha = x_1 + \cdots + x_m \in D(X, x_0)$, $\beta = x'_1 + \cdots + x'_n \in D(X, x_0)$ and let $x_{m+1} = \cdots = x_{m+n} = x_0 = x'_{n+1} = \cdots = x'_{n+m}$. Then

$$\begin{aligned} \rho(\tilde{\varphi}(\alpha), \tilde{\varphi}(\beta)) &= \rho(\varphi(x_1) + \cdots + \varphi(x_{m+n}), \varphi(x'_1) + \cdots + \varphi(x'_{m+n})) \\ &\leqslant \min_{\sigma \in \Sigma_{m+n}} \left\| (\rho(\varphi(x_i), \varphi(x'_i)))_{i=1}^{m+n} \right\|_p \leqslant \min_{\sigma \in \Sigma_{m+n}} \left\| (d(x_i, x'_i))_{i=1}^{m+n} \right\|_p = W_p[d, x_0](\alpha, \beta). \end{aligned}$$

Thus $((D(X, x_0), W_p[d, x_0], 0), i)$ is a universal element, and the fact that $D_p$ is left adjoint to $U_p$ now follows from Lemma 2.12. $\square$

The above constructions can also be formalized using metric pairs instead of pointed metric spaces, as we will now demonstrate.

**Definition 4.22.** Let $\overline{U}_p : \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}}) \to \mathbf{Met}_{\mathbf{pairs}}$ be the functor given by sending $(M, d, +, 0)$ to $(M, d, \{0\})$ and $f : (M, d, +, 0) \to (N, \rho, +, 0)$ to $f : (M, d, \{0\}) \to (N, \rho, \{0\})$.

Note that $\overline{U}_p$ is just the composition $\mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}}) \xrightarrow{U_p} \mathbf{Met}_*^{\mathbf{p}} \hookrightarrow \mathbf{Met}_* \hookrightarrow \mathbf{Met}_{\mathbf{pairs}}$.

**Definition 4.23.** Let $\overline{D}_p : \mathbf{Met}_{\mathbf{pairs}} \to \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$ be the functor given by sending $(X, d, A)$ to $(D(X, A), W_p[d, A], +, 0)$ and $f : (X, d, A) \to (Y, d', B)$ to $f_* : (D(X, A), W_p[d, A], +, 0) \to (D(Y, B), W_p[d', B], +, 0)$.

Recall the functors $Q : \mathbf{Met}_{\mathbf{pairs}} \to \mathbf{Met}_*$ of Definition 3.8 and $S_p : \mathbf{Met}_* \to \mathbf{Met}_*^{\mathbf{p}}$ of Definition 3.15. We will show that $\overline{D}_p$ is the left adjoint of $\overline{U}_p$ (Theorem 4.25). This will follow from the following lemma, which shows that $\overline{D}_p$ is naturally isomorphic to the composition $D_p S_p Q$, together with the fact that each of $D_p, S_p, Q$ has a right adjoint, and the composition of these right adjoints is precisely $\overline{U}_p$.

**Lemma 4.24.** *The functors $\overline{D}_p : \mathbf{Met}_{\mathbf{pairs}} \to \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$ and $D_p S_p Q : \mathbf{Met}_{\mathbf{pairs}} \to \mathbf{CMon}(\mathbf{Met}_*^{\mathbf{p}})$ are naturally isomorphic.*

**Proof.** Let $(X, d, A) \in \mathbf{Met}_{\mathbf{pairs}}$. Then $\overline{D}_p(X, d, A) = (D(X, A), W_p[d, A], +, 0)$ and $D_p S_p Q(X, d, A) = (D(X/A, A), W_p[\overline{d}_p, A], +, 0)$. Recall (see Remark 4.14) that we have monoid isomorphisms $\varphi = \varphi_X : D(X \setminus A) \to D(X, A)$ and $\psi = \psi_X : D(X \setminus A) \to D(X/A, A)$. Let $\eta = \eta_X : D(X, A) \to D(X/A, A)$ be the composite monoid isomorphism $\psi_X \varphi_X^{-1}$. Explicitly, $\eta_X$ is given by $x_1 + \cdots + x_n + D(A) \mapsto [x_1] + \cdots + [x_n] + D(A)$. We will show that $\eta_X$ is a isometry. Let $x_1 + \cdots + x_m, x'_1 + \cdots + x'_n \in D(X \setminus A)$. Denote these elements by $\alpha$ and $\alpha'$, respectively. Let $x_{m+1} = \cdots = x_{m+n} = x'_{n+1} = x'_{n+m} = A$. By Lemma 4.13,

$$W_p[d, A](\varphi\alpha, \varphi\alpha') = \min_{\sigma \in \Sigma_{m+n}} \left\| (d(x_i, x'_{\sigma(i)}))_{i=1}^{m+n} \right\|_p .$$

On the other hand, by Lemma 3.17,

$$\begin{aligned} W_p[\overline{d}_p, A](\psi\alpha, \psi\alpha') &= \min_{\sigma \in \Sigma_{m+n}} \left\| (\overline{d}_p(x_i, x'_{\sigma(i)}))_{i=1}^{m+n} \right\|_p \\ &= \min_{\sigma \in \Sigma_{m+n}} \left\| (\min(d(x_i, x'_{\sigma(i)}), \left\| d(x_i, A), d(A, x_{\sigma(i)}) \right\|_p))_{i=1}^{m+n} \right\|_p . \end{aligned}$$

Therefore $W_p[\overline{d}_p, A](\psi\alpha, \psi\alpha') \leqslant W_p[d, A](\varphi\alpha, \varphi\alpha')$. On the other hand, let $\sigma_0 \in \Sigma_{n+m}$ be a permutation such that $W_p[\overline{d}_p, A](\psi\alpha, \psi\alpha') = \left\| (\min(d(x_i, x'_{\sigma_0(i)}), \left\| d(x_i, A), d(A, x_{\sigma_0(i)}) \right\|_p))_{i=1}^{m+n} \right\|_p$. If there is an index $i$ with $\left\| (d(x_i, A), d(A, x'_{\sigma_0(i)})) \right\|_p$

$$
\begin{tikzcd}
\mathbf{CMon}(\mathbf{Met}^{\mathbf{p}}_*) \arrow[dd, shift left, "\mathrm{U_p}"] \arrow[drr, shift left, "\overline{\mathrm{U}}_\mathrm{p}"] \arrow[dd, phantom, "\dashv"'] & & & & \mathbf{CMon}(\mathbf{Met}^{\mathbf{q}}_*) \arrow[dd, shift right, "\mathrm{U_q}"'] \arrow[dll, shift right, "\overline{\mathrm{U}}_\mathrm{q}"'] \arrow[dd, phantom, "\dashv"] \\
& & \mathbf{Met_{pairs}} \arrow[ull, dashed, shift left, "\overline{\mathrm{D}}_\mathrm{p}"] \arrow[urr, dashed, shift right, "\overline{\mathrm{D}}_\mathrm{q}"'] \arrow[dll, shift left, "\mathrm{S_p Q}"] \arrow[drr, shift right, "\mathrm{S_q Q}"'] \arrow[ull, phantom, "\top"] \arrow[urr, phantom, "\top"'] \arrow[dll, phantom, "\top"'] \arrow[drr, phantom, "\top"] & & \\
\mathbf{Met}^{\mathbf{p}}_* \arrow[uu, dashed, shift left, "\mathrm{D_p}"] \arrow[urr, hook, dashed, shift left] \arrow[rrrr, shift left, "\mathrm{S_{p,q}}"] \arrow[rrrr, phantom, "\perp"'] & & & & \mathbf{Met}^{\mathbf{q}}_* \arrow[uu, dashed, shift right, "\mathrm{D_q}"'] \arrow[ull, hook, dashed, shift right] \arrow[llll, hook, dashed, shift left]
\end{tikzcd}
$$

**Fig. 1.** For $1 \leqslant p \leqslant q \leqslant \infty$, the relationships between the categories $\mathbf{Met}^{\mathbf{p}}_*$, $\mathbf{Met}^{\mathbf{q}}_*$, $\mathbf{CMon}(\mathbf{Met}^{\mathbf{p}}_*)$, $\mathbf{CMon}(\mathbf{Met}^{\mathbf{q}}_*)$, and $\mathbf{Met_{pairs}}$ via the functors $U_r, D_r, \overline{U}_r, \overline{D}_r$, $S_r$ $(r \in \{p, q\})$, and $Q$, $S_{p,q}$. The solid triangles commute and the dashed triangles commute.

$< d(x_i, x'_{\sigma_0(i)})$ then we may choose a new permutation $\sigma_1$ with $x'_{\sigma_1(i)} = A$ and $x_{\sigma_1^{-1}\sigma_0(i)} = A$ and which is otherwise the same as $\sigma_0$. By induction we remove all such indices to obtain a new permutation $\sigma'_0 \in \Sigma_{n+m}$ with

$$\|(d(x_i, x'_{\sigma'_0(i)}))_{i=1}^{m+n}\|_p = \left\|(\overline{d}_p(x_i, x'_{\sigma_0(i)}))_{i=1}^{m+n}\right\|_p .$$

Therefore $W_p[\overline{d}_p, A](\psi\alpha, \psi\alpha') = W_p[d, A](\varphi\alpha, \varphi\alpha')$. Then

$$\begin{aligned} W_p[\overline{d}_p, A](\eta\alpha, \eta\alpha') &= W_p[\overline{d}_p, A](\psi\varphi^{-1}\alpha, \psi\varphi^{-1}\alpha') \\ &= W_p[d, A](\varphi\varphi^{-1}\alpha, \varphi\varphi^{-1}\alpha') = W_p[d, A](\alpha, \alpha'). \end{aligned}$$

Thus we have an isomorphism $\eta_X : \overline{D}_p(X, d, A) \to D_p S_p Q(X, d, A)$.

To see that these isomorphisms are natural, let $f : (X, d, A) \to (Y, d', B)$ be a morphism in $\mathbf{Met_{pairs}}$. The map $\overline{D}_p f = f_* : D(X, A) \to D(Y, B)$ is given by $x_1 + \cdots + x_n + D(A) \mapsto f(x_1) + \cdots + f(x_n) + D(B)$, while the map $D_p S_p Q f : D(X/A, A) \to D(Y/B, B)$ is given by $[x_1] + \cdots + [x_n] + D(A) \mapsto [f(x_1)] + \cdots + [f(x_n)] + D(B)$. Thus $(D_p S_p Q f)\eta_X = \eta_Y \overline{D}_p f$ and hence the maps $\eta_X$ assemble into a natural isomorphism $\eta : \overline{D}_p \Rightarrow D_p S_p Q$. $\square$

**Theorem 4.25.** *The forgetful functor $\overline{U}_p : \mathbf{CMon}(\mathbf{Met}^{\mathbf{p}}_*) \to \mathbf{Met_{pairs}}$ has left adjoint $\overline{D}_p$.*

**Proof.** Since $\overline{U}_p$ is given by the composition $\mathbf{CMon}(\mathbf{Met}^{\mathbf{p}}_*) \xrightarrow{U_p} \mathbf{Met}^{\mathbf{p}}_* \hookrightarrow \mathbf{Met}_* \hookrightarrow \mathbf{Met_{pairs}}$, by Theorems 3.9, 3.16, and 4.21, it has left adjoint the composite $D_p S_p Q$. By Lemma 4.24, $\overline{D}_p \cong D_p S_p Q$ and hence $\overline{D}_p$ is, up to natural isomorphism, the left adjoint of $\overline{U}_p$. $\square$

The relationship between the forgetful functors $U_p, \overline{U}_p$, the free functors $D_p, \overline{D}_p$, the quotient functors $Q$, and the functors $S_p, S_{p,q}$ is summarized in Fig. 1.

## 5. Applications

In this section, we give several applications of universality. The first application shows that for a pointed metric space $(X, d, x_0)$, $W_p$ is the largest $p$-subadditive metric on $D(X, x_0)$ which in some sense extends the metric $d$. This result implies an abstract form of converse stability from which we derive converse stability-type results in various settings. As a second application, we show how universality can be used to derive the correct form of Kantorovich-Rubinstein duality for persistence diagrams.

### 5.1. Maximality of the Wasserstein distances

The following theorem shows that $W_p$ is the largest $p$-subadditive metric extending the underlying metric.

**Theorem 5.1.** *Let $p \in [1, \infty]$ and let $(X, d, A)$ be a metric pair. Then $W_p[d, A]$ is the largest $p$-subadditive metric $\rho$ on $D(X, A)$ satisfying $i^*\rho = d_p$.*

**Proof.** Suppose that $\rho$ is a $p$-subadditive metric on $D(X, A)$ with $i^*\rho = d_p$. Then $(D(X, A), \rho, +, 0) \in \mathbf{CMon}(\mathbf{Met}^{\mathbf{p}}_*)$ and $i : (X, d, A) \hookrightarrow (D(X, A), \rho, +, 0)$ is 1-Lipschitz. By Theorem 4.25, there is a unique 1-Lipschitz map $\tilde{i} : (D(X, A), W_p[d, A], 0) \to (D(X, A), \rho, 0)$, and hence $\rho \leqslant W_p$. $\square$

If $d$ satisfies the $p$-strengthened triangle inequality then $d = d_p$, and so we immediately obtain the following.

**Corollary 5.2.** *Let $p \in [1, \infty]$ and let $(X, d, x_0) \in \mathbf{Met}_*^{\mathbf{p}}$. Then $W_p[d, x_0]$ is the largest $p$-subadditive metric $\rho$ on $D(X, x_0)$ satisfying $i^*\rho = d$.*

As another application of universality, we show that for a pointed metric space $(X, d, x_0)$, $W_p[d, x_0] = W_p[d_p, x_0]$.

**Corollary 5.3.** *Let $(X, d, x_0) \in \mathbf{Met}_*$. Then $W_p[d, x_0] = W_p[d_p, x_0]$.*

**Proof.** Note that $i^*W_p[d, x_0] = i^*W_p[d_p, x_0] = d_p$. Since $W_p[d, x_0]$ and $W_p[d_p, x_0]$ are both $p$-subadditive, we have $W_p[d, x_0] = W_p[d_p, x_0]$ by Theorem 5.1. $\square$

The preceding corollary justifies our use of the categories $\mathbf{Met}_*^{\mathbf{p}}$. By Lemma 4.18, if $(X, d, x_0) \in \mathbf{Met}_*^{\mathbf{p}}$, then $(X, d)$ embeds into $(D(X, x_0), W_p[d, x_0], 0)$. This corollary shows that we can always pass to $d_p$ without changing the Wasserstein distance, and so it suffices to work in $\mathbf{Met}_*^{\mathbf{p}}$.

### 5.2. Converse stability

We show that certain converse stability theorems follow from our results. The following is a completely formal (or "soft" [1]) converse stability result, from which specific converse stability theorems follow.

**Theorem 5.4** *(Abstract Converse Stability). Fix $p \in [1, \infty]$. Let $(X, x_0)$ be a pointed set and let $\rho$ be a $p$-subadditive metric on $D(X, x_0)$. Then $\rho \leqslant W_p[i^*\rho, x_0]$.*

**Proof.** Let $d = i^*\rho$ and consider the pointed metric space $(X, d, x_0)$. Since $\rho$ is $p$-subadditive by assumption and $i^*\rho = d$ by definition, the result immediately follows from Corollary 5.2. $\square$

**Example 5.5** *(Converse Algebraic Stability).* (See [19] for a version of this result that applies to all pointwise finite dimensional persistence modules.) Let $\mathbf{Vect}(K)^{\mathbb{R}}_{\mathrm{Fin}}$ denote the monoid of isomorphism classes of persistence modules which decompose as a finite direct sum of interval modules. We can identify $\mathbf{Vect}(K)^{\mathbb{R}}_{\mathrm{Fin}}$ with $D(\mathrm{Int}(\mathbb{R}), \emptyset)$ via the map that sends a direct sum of interval modules to the corresponding formal sum of intervals. Equip $\mathbf{Vect}(K)^{\mathbb{R}}_{\mathrm{Fin}}$ with the interleaving distance $d_I$ [9,11,19]. Note that $d_I$ is $\infty$-subadditive. Indeed, if $(\varphi, \psi)$ is an $\epsilon$-interleaving between $M$ and $N$ and $(\varphi', \psi')$ is an $\eta$-interleaving between $M'$ and $N'$, then $(\varphi \oplus \varphi', \psi \oplus \psi')$ is a $\max(\epsilon, \eta)$-interleaving between $M \oplus M'$ and $N \oplus N'$. Note that the interleaving distance for interval modules is $(d_H)_\infty$, the $\infty$-strengthening of the Hausdorff distance with respect to $\emptyset$. That is, $i^*d_I = (d_H)_\infty$, and hence by Theorem 5.4 and Corollary 5.3, $d_I \leqslant W_\infty[(d_H)_\infty, \emptyset] = W_\infty[d_H, \emptyset]$.

For a second version of this result, introduce an equivalence relation on $\mathbf{Vect}(K)^{\mathbb{R}}_{\mathrm{Fin}}$ given by $M \sim N$ if $d_I(M, N) = 0$. Then we can identify $\mathbf{Vect}(K)^{\mathbb{R}}_{\mathrm{Fin}}/\sim$ with $D(\overline{\mathbb{R}}^2_{\leqslant}, \Delta)$ via the map that sends a direct sum of interval modules to the corresponding persistence diagram. By Theorem 5.4, $d_I \leqslant W_\infty[d, \Delta]$, where $d$ is the $\ell^\infty$-distance. The distance $W_\infty[d, \Delta]$ is the *bottleneck distance*. Note that this distance restricted to the images of interval modules is $d_\infty$, the $\infty$-strengthening of the $\ell^\infty$-distance.

**Example 5.6** *(Converse Algebraic Stability Theorem for generalized persistence modules).* Consider generalized persistence modules $M : \mathbf{P} \to \mathbf{A}$. If $\mathbf{P}$ is equipped with certain additional structure, such as a *subadditive projection on translations* or a *superadditive family of translations*, then $\mathbf{A}^{\mathbf{P}}$ can be equipped with an interleaving distance $d_I$ [1]. As in Example 5.5, the interleaving distance $d_I$ is $\infty$-subadditive. Let Ind be a set of indecomposable generalized persistence modules in $\mathbf{A}^{\mathbf{P}}$ with basepoint the zero module 0. Then we have the set of generalized barcodes $D(\mathrm{Ind}, 0)$. There is a bijection from the set of isomorphism classes of generalized persistence modules in $\mathbf{A}^{\mathbf{P}}$ that are isomorphic to a finite direct sum of elements of Ind to $D(\mathrm{Ind}, 0)$, which sends direct sums to formal sums. By Theorem 5.4, $d_I \leqslant W_\infty[d_I, 0]$, where the latter is also called the bottleneck distance [5].

### 5.3. Kantorovich-Rubinstein duality

The classical Kantorovich-Rubinstein duality theorem says that the classical 1-Wasserstein distance $w_1(\mu, \nu)$ between probability measures $\mu$ and $\nu$ on a complete and separable metric space $(X, d)$ is equal to $\sup \int_X f d(\mu - \nu)$, where the supremum is taken over all 1-Lipschitz functions. A version of Kantorovich-Rubinstein duality holds for persistence diagrams as well. We will show that

$$W_1(\textstyle\sum_{i=1}^n a_i, \sum_{j=1}^m b_j) = \sup\{\sum_{i=1}^n k(a_i) - \sum_{j=1}^m k(b_j)\},$$

where now the supremum is taken over all 1-Lipschitz functions $k : X \to \mathbb{R}$ with $k(x_0) = 0$.

To motivate the form that Kantorovich-Rubinstein duality takes for persistence diagrams, we first show how Theorem 1.3 can be used to derive the inequality

$$W_1(\sum_{i=1}^n a_i, \sum_{j=1}^m b_j) \geqslant \sup\{\sum_{i=1}^n k(a_i) - \sum_{j=1}^m k(b_j)\}.$$

We will then use the classical Kantorovich-Rubinstein duality theorem to show that this is in fact an equality.

Let $(X, d, x_0) \in \mathbf{Met}_*$ and consider the commutative metric monoid $(\mathbb{R}, |\cdot|, +, 0)$, where $|\cdot|$ denotes the metric induced by absolute value. The inequality $|(a+b)-(c+d)| \leqslant |a-c|+|b-d|$ implies that $(\mathbb{R}, |\cdot|, +, 0) \in \mathbf{CMon}(\mathbf{Met}_*)$. Let $h : X \to \mathbb{R}$ be a 1-Lipschitz map. Define $k : X \to \mathbb{R}$ by $k(x) = h(x) - h(x_0)$ for all $x \in X$. Then $k(x_0) = 0$ and $|k(x) - k(y)| = |h(x) - h(x_0) - (h(y) - h(x_0))| = |h(x) - h(y)| \leqslant d(x, y)$ so that $k$ is a pointed metric map. By Theorem 1.3, there is a unique morphism of 1-subadditive commutative metric monoids $\tilde{k} : (D(X, x_0), W_1, +, 0) \to (\mathbb{R}, |\cdot|, +, 0)$ such that $\tilde{k} \circ i = k$. Explicitly, $\tilde{k}$ is given by $\sum_i c_i \mapsto \sum_i k(c_i)$ for $c_i \in X$. Then for $\alpha = \sum_{i=1}^n a_i$, $\beta = \sum_{j=1}^m b_j \in D(X, x_0)$ with $m \geqslant n$, we have $|\tilde{k}(\alpha) - \tilde{k}(\beta)| = \left|\tilde{k}\left(\sum_{i=1}^n a_i\right) - \tilde{k}\left(\sum_{j=1}^m b_j\right)\right| = \left|\sum_{i=1}^n k(a_i) - \sum_{j=1}^m k(b_j)\right| = \left|\sum_{i=1}^n h(a_i) - \sum_{j=1}^m h(b_j) + (m-n)h(x_0)\right|$. Since $\tilde{k}$ is 1-Lipschitz, we obtain the inequality $\left|\sum_{i=1}^n h(a_i) - \sum_{j=1}^m h(b_j) + (m-n)h(x_0)\right| \leqslant W_1\left(\sum_{i=1}^n a_i, \sum_{j=1}^m b_j\right)$. Therefore

$$\sup\left\{\sum_{i=1}^n h(a_i) - \sum_{j=1}^m h(b_j) + (m-n)h(x_0) \;\middle|\; h : X \to \mathbb{R},\ \text{1-Lipschitz}\right\} \leqslant W_1\left(\sum_{i=1}^n a_i, \sum_{j=1}^m b_j\right), \tag{5.1}$$

or equivalently,

$$\sup\left\{\sum_{i=1}^n k(a_i) - \sum_{j=1}^m k(b_j) \;\middle|\; k : X \to \mathbb{R},\ k(x_0) = 0,\ \text{1-Lipschitz}\right\} \leqslant W_1\left(\sum_{i=1}^n a_i, \sum_{j=1}^m b_j\right).$$

To see that this inequality is in fact an equality, consider $\alpha = a_1 + \cdots + a_n$, $\beta = b_1 + \cdots + b_m \in D(X, x_0)$ and consider the classical 1-Wasserstein distance $w_1(\sum_{i=1}^n \delta_{a_i} + (r-n)\delta_{x_0}, \sum_{i=1}^m \delta_{b_j} + (r-m)\delta_{x_0})$, where $\delta_x$ is the Dirac measure at $x$ and $r = m + n$. It is known that for sums of Dirac measures the computation of $w_1$ is equivalent to the *linear assignment problem*. In other words, letting $\tilde{\alpha} = \sum_{i=1}^n \delta_{a_i} + (r-n)\delta_{x_0}$ and $\tilde{\beta} = \sum_{i=1}^m \delta_{b_j} + (r-m)\delta_{x_0}$, we have

$$w_1\left(\tilde{\alpha}, \tilde{\beta}\right) = \min_{\sigma \in S_{n+m}} \left\|(d(a_i, b_{\sigma(i)})_{i=1}^{n+m})\right\|_1 = W_1[d, x_0](\alpha, \beta),$$

where $a_{n+1} = \cdots = a_{n+m} = b_{m+1} = \cdots = b_{n+m} = x_0$. It follows from Kantorovich-Rubinstein duality for measures [23] that

$$\sup\left\{\int_X h d(\tilde{\alpha} - \tilde{\beta}) \mid h : X \to \mathbb{R},\ \text{1-Lipschitz}\right\} = w_1(\tilde{\alpha}, \tilde{\beta}) = W_1[d, x_0](\alpha, \beta). \tag{5.2}$$

Since the left-hand side of (5.2) is precisely the left-hand side of (5.1), we obtain the desired equality.

For a direct proof of Kantorovich-Rubinstein duality in this setting using linear programming see [2, Appendix C].

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research was partially supported by the Southeast Center for Mathematics and Biology, an NSF-Simons Research Center for Mathematics of Complex Biological Systems, under National Science Foundation Grant No. DMS-1764406 and Simons Foundation Grant No. 594594. This material is based upon work supported by, or in part by, the Army Research Laboratory and the Army Research Office under contract/grant number W911NF-18-1-0307.

## References

[1] Peter Bubenik, Vin de Silva, Jonathan Scott, Metrics for generalized persistence modules, Found. Comput. Math. 15 (6) (2015) 1501–1531.
[2] Peter Bubenik, Alex Elchesen, Universality of persistence diagrams and the bottleneck and Wasserstein distances, arXiv preprint, arXiv:1912.02563v2, 2020.
[3] Peter Bubenik, Alex Elchesen, Virtual persistence diagrams, signed measures, Wasserstein distances, and Banach spaces, J. Appl. Comput. Topol. (2022), https://doi.org/10.1007/s41468-022-00091-9.
[4] Andrew J. Blumberg, Itamar Gal, Michael A. Mandell, Matthew Pancia, Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces, Found. Comput. Math. 14 (4) (2014) 745–789.
[5] Håvard Bakke Bjerkevik, On the stability of interval decomposable persistence modules, Discrete Comput. Geom. 66 (1) (2021) 92–121.

[6] Andrew J. Blumberg, Michael Lesnick, Universality of the homotopy interleaving distance, arXiv preprint, arXiv:1705.01690, 2017.
[7] Magnus Bakke Botnan, Michael Lesnick, Algebraic stability of zigzag persistence modules, Algebraic Geom. Topol. 18 (6) (2018) 3133–3204.
[8] Ulrich Bauer, Claudia Landi, Facundo Mémoli, The Reeb graph edit distance is universal, Found. Comput. Math. (2020) 1–24.
[9] Peter Bubenik, Jonathan A. Scott, Categorification of persistent homology, Discrete Comput. Geom. 51 (3) (2014) 600–627.
[10] Peter Bubenik, Jonathan Scott, Donald Stanley, An algebraic Wasserstein distance for generalized persistence modules, arXiv preprint, arXiv:1809.09654, 2018.
[11] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, Steve Y. Oudot, Proximity of persistence modules and their diagrams, in: Proceedings of the 25th Annual Symposium on Computational Geometry - SCG '09, 2009.
[12] Jérémy Cochoy, Steve Oudot, Decomposition of exact pfd persistence bimodules, Discrete Comput. Geom. 63 (2) (2020) 255–293.
[13] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, Stability of persistence diagrams, Discrete Comput. Geom. 37 (1) (2007) 103–120.
[14] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, Yuriy Mileyko, Lipschitz functions have $L_p$-stable persistence, Found. Comput. Math. 10 (2) (2010) 127–139.
[15] Gunnar Carlsson, Afra Zomorodian, Anne Collins, Leonidas Guibas, Persistence barcodes for shapes, in: SGP '04: Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, ACM Press, New York, NY, USA, 2004, pp. 124–135.
[16] Anne Collins, Afra Zomorodian, Gunnar Carlsson, Leonidas J. Guibas, A barcode shape descriptor for curve point cloud data, Comput. Graph. 28 (6) (2004) 881–894.
[17] Michele d'Amico, Patrizio Frosini, Claudia Landi, Natural pseudo-distance and optimal matching between reduced size functions, Acta Appl. Math. 109 (2) (2010) 527–554.
[18] Vincent Divol, Théo Lacombe, Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport, J. Appl. Comput. Topol. 5 (1) (2021) 1–53.
[19] Michael Lesnick, The theory of the interleaving distance on multidimensional persistence modules, Found. Comput. Math. 15 (3) (2015) 613–650.
[20] Saunders Mac Lane, Categories for the Working Mathematician, second edition, Graduate Texts in Mathematics, vol. 5, Springer-Verlag, New York, 1998.
[21] E. Riehl, Category Theory in Context. Aurora: Dover Modern Math Originals, Dover Publications, 2017.
[22] Primoz Skraba, Katharine Turner, Wasserstein stability for persistence diagrams, arXiv preprint, arXiv:2006.16824, 2020.
[23] Cédric Villani, Topics in Optimal Transportation, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.