FREE COMPLETE WASSERSTEIN ALGEBRAS

RADU MARDARE, PRAKASH PANANGADEN, AND GORDON D. PLOTKIN

Abstract. We present an algebraic account of the Wasserstein distances W_p on complete metric spaces. This is part of a program of a quantitative algebraic theory of effects in programming languages. In particular, we give axioms, parametric in p, for algebras over metric spaces equipped with probabilistic choice operations. The axioms say that the operations form a barycentric algebra and that the metric satisfies a property typical of the Wasserstein distance W_p. We show that the free complete such algebra over a complete metric space is that of the Radon probability measures on the space with the Wasserstein distance as metric, equipped with the usual binary convex sum operations.

1. Introduction

We present an algebraic account of the Wasserstein distances W_p on complete metric spaces. This is part of a program begun in [7] to establish a quantitative algebraic theory of effects in programming languages. This program employs a quantitative version of equational logic where equations give an upper bound on the distance between two elements, and algebras are extended metric spaces equipped with nonexpansive operations.

In particular, in [7] we gave axioms, parametric in p, for algebras equipped with probabilistic choice operations and showed the free complete such algebras over a complete 1-bounded separable metric space X are the probability measures on X with the Wasserstein distance, equipped with the usual binary convex sum operations. However, we only did this for the case where the metric space was 1-bounded and separable.

In this note we prove such a result for all complete metric spaces, but only for algebras over metric spaces, rather than extended ones. We avoid the 1-bounded restriction by the standard idea of using probability measures with finite p-moment; we avoid the restriction to separable metric spaces by using Radon probability measures.

Section 2 discusses the p-Wasserstein distance between probability measures on a metric space X. In Theorem 1 we show that the distance is a metric space on the Radon probability measures with finite p-moment; to do so, we make use of a result in [2] that the triangle inequality holds for all probability measures if X is separable. Then, in Theorem 2 we show that if X is complete then the p-Wasserstein metric on the Radon probability measures with finite p-moment is also complete and is generated by the probability measures with finite support. To do this, we make use of the well-known result that if X is complete and separable then the p-Wasserstein metric of all probability measures over X with finite p-moment is also complete and separable, being generated by the rational measures with finite support in a countable basis of X (see [11, Theorem 6.18] and the bibliographic discussion there).

Section 3 discusses the algebraic aspect of these spaces. In particular, in Theorem 4 we show that the space of Radon probability measures on a complete metric space, equipped with binary convex sums, forms the free complete p-Wasserstein algebra over X. (The case $p = 1$ was essentially already proved in [4], as it is an immediate consequence of Theorem 5.2.1 there.) There is a certain disconnect between our work and that of [7] in that here we use standard metric spaces, whereas there the more general framework of extended metric spaces is employed. We sketch how to bridge this disconnect at the end of Section 3.

2. The Wasserstein distance

We begin with some technical preliminaries on Radon probability measures and couplings and their support. For general background on probability measures on topological and metric spaces see [5, 9]. By probability measure we mean a Borel probability measure. Given such a probability measure μ on a Hausdorff space X, we say that a Borel set B is *compact inner regular (for μ)* if:

$$\mu(B) = \sup\{\mu(C) \mid C \text{ compact}, C \subseteq B\}$$

Then μ is *Radon* if all Borel sets are compact inner regular for it, and *tight* if X is compact inner regular for it (equivalently, if for any Borel set B and $\varepsilon > 0$ there is a compact set C such that $\mu(B\backslash C) < \varepsilon$). Every tight probability measure on a metric space is Radon and every probability measure on a separable complete metric space is tight.

The *support* of a probability measure μ on a topological space X is

$$\mathrm{supp}(\mu) =_{\mathbf{def}} \{x \in X \mid \mu(U) > 0 \text{ for all open } U \text{ containing } x\}$$

Note that the support is always a closed set. If μ is Radon then $\mathrm{supp}(\mu)$ has measure 1. If the support of μ is finite then μ is Radon and can be written uniquely, up to order, as a finite convex sum of Dirac measures, viz:

$$\mu = \sum_{x \in \mathrm{supp}(\mu)} \mu(s)\delta(s)$$

(writing $\mu(s)$ instead of $\mu(\{s\})$). We say that μ is *rational* if all the $\mu(s)$ are.

The following very useful lemma is due to Basso [1]. It enables us, as he did, to establish results about probability measures on metric spaces by applying results about probability measures on separable metric spaces to their supports.

Lemma 1. *Every probability measure on a metric space has separable support.*

Proof. For every $\varepsilon > 0$ let C_ε be a maximal set of points in $\mathrm{supp}(\mu)$ with the property that all points are at distance $\geq \varepsilon$ apart. As any two open balls with centre in C_ε and radius $1/2\varepsilon$ are disjoint and each such ball has μ-measure > 0, C_ε is countable (were it uncountable, we would have an uncountable collection of reals with any denumerable subset summable, and no such collection exists). By the maximality of C_ε, any point in $\mathrm{supp}(\mu)$ is at distance $< \varepsilon$ from some point in C_ε. It follows that the countable set $\bigcup_{n\geq 0} C_{2^{-n}}$ is dense in $\mathrm{supp}(\mu)$. $\square$

A *coupling* γ between two probability measures μ and ν on a topological space X is a probability measure on X^2 whose left and right marginals (= pushforwards along the projections) are, respectively, μ and ν. We gather some facts about such couplings:

Lemma 2. *Let γ be a coupling between two probability measures μ and ν on a topological space X. Then:*

$$\mathrm{supp}(\gamma) \subseteq \mathrm{supp}(\mu) \times \mathrm{supp}(\nu)$$

Also, if μ and ν are tight, so is γ.

Proof. For the first part, suppose that $(x, y) \in \mathrm{supp}(\gamma)$ and let μ be an open neighbourhood of x. Then $U \times Y$ is an open neighbourhood of (x, y) and so $\mu(U) = \gamma(U \times Y) > 0$. So $x \in \mathrm{supp}(\mu)$. Similarly $y \in \mathrm{supp}(\nu)$.

For the second part, choose $\varepsilon > 0$. As μ and ν are tight, X and Y are compact inner regular for them. So there are compact sets $C \subseteq X$ and $D \subseteq Y$ such that $\mu(X \backslash C) < 1/2\varepsilon$ and $\nu(Y \backslash D) < 1/2\varepsilon$. Then we have:

$$\begin{array}{rcl}
\gamma((X \times Y)\backslash(C \times D)) & = & \gamma((X \times Y)\backslash((C \times Y) \cap (X \times D))) \\
& = & \gamma(((X \times Y)\backslash(C \times Y)) \cup ((X \times Y)\backslash(X \times D))) \\
& = & \gamma(((X\backslash C) \times Y) \cup (X \times (Y\backslash D))) \\
& \leq & \gamma((X\backslash C) \times Y) + \gamma(X \times (Y\backslash D)) \\
& = & \mu(X\backslash C) + \nu(Y\backslash D) \\
& < & \varepsilon
\end{array}$$

So, as ε was arbitrary, $X \times Y$ is compact inner regular for γ, as required. $\square$

It follows from the lemma that any coupling between Radon probability measures on metric spaces is itself Radon.

We now turn to the Wasserstein distance. A probability measure μ on a metric space X is said to have *finite p-th moment*, where $p > 1$, if, for some (equivalently all) $x_0 \in X$, the integral

$$\int d(x_0, -)^p \mathrm{d}\mu$$

is finite. The p-Wasserstein distance function W_p is defined on probability measures with finite p-th moment by:

$$\mathrm{W}_p(\mu, \nu) = \inf_{\gamma} \left(\int d_X^p \mathrm{d}\gamma \right)^{1/p}$$

where γ runs over the couplings between μ and ν (the integrals are finite as μ and ν have finite pth moment).

We need two lemmas relating probability measures on a metric space with probability measures on a closed subset of the space. Let C be a closed subset of a metric space X. We write $\mathrm{i}_*(\mu)$ for the pushforward of a finite measure μ on C along the inclusion $\mathrm{i} : C \to X$ of C in X, so $\mathrm{i}_*(\mu)(B) = \mu(B \cap C)$; and we write $\mathrm{r}(\nu)$ for the restriction of a finite measure ν on X to a finite measure on C, so $\mathrm{r}(\nu)(B) = \nu(B)$. Note that $\mathrm{r}(\mathrm{i}_*(\mu)) = \mu$.

Lemma 3. *Let C be a closed subset of a metric space X.*

(1) If μ is a Radon probability measure on C, then $\mathrm{i}_*(\mu)$ is a Radon probability measure on X with the same support as μ. Further, $\mathrm{i}_*(\mu)$ has finite p-th moment if μ does.
(2) If ν is a Radon probability measure on X with support included in C then $\mathrm{r}(\nu)$ is a Radon probability measure on C, and we have:

$$\nu = \mathrm{i}_*(\mathrm{r}(\nu))$$

If, further, ν has finite p-th moment, so does $\mathrm{r}(\nu)$.

Proof. (1) Let μ be a Radon probability measure on C. Then C is non-empty. Further, it is straightforward to check that $\mathrm{i}(\mu)$ is a Radon probability measure on X with the same support as C. Next, choosing $x_0 \in C$, as

$$\int d_X(x_0, -)^p \mathrm{d}(\mathrm{i}_*\mu) = \int d_C(x_0, -)^p \mathrm{d}\mu$$

we see that $\mathrm{i}_*(\mu)$ has finite p-th moment if μ does.

(2) Let ν be a Radon probability measure on X with support included in C. Then C is nonempty. Further, it is straightforward to check that $\mathrm{r}(\nu)$ is a Radon probability measure on C. Regarding the equality, for any Borel set B of X we have:

$$\mathrm{i}_*(\mathrm{r}(\nu))(B) = \mathrm{r}(\nu)(B \cap C) = \nu(B \cap C) = \nu(B)$$

with the last equality holding as ν is Radon and the support of ν is included in C. Finally, $\mathrm{r}(\nu)$ has finite p-th moment if ν does, as, choosing $x_0 \in C$, we have:

$$\int d_C(x_0, -)^p \mathrm{d}(\mathrm{r}(\nu)) = \int d_X(x_0, -)^p \mathrm{d}(\mathrm{i}_*(\mathrm{r}(\nu))) = \int d_X(x_0, -)^p \mathrm{d}\nu$$

□

Lemma 4. *Let C be a closed subset of a metric space X. Then*
(1) For any $\mu, \nu \in P_p(C)$, we have:

$$\mathrm{W}_p(\mu, \nu) = \mathrm{W}_p(\mathrm{i}_*(\mu), \mathrm{i}_*(\nu))$$

(2) For any $\mu, \nu \in P_p(X)$ whose support is included in C, we have:

$$\mathrm{W}_p(\mu, \nu) = \mathrm{W}_p(\mathrm{r}(\mu), \mathrm{r}(\nu))$$

Proof. (1) Let γ be a coupling between μ and ν. Then $\mathrm{i}_*(\gamma)$ is a coupling between $\mathrm{i}_*(\mu)$ and $\mathrm{i}_*(\nu)$, as:

$$\begin{array}{rcl} \mathrm{i}_*(\gamma)(B \times X) & = & \gamma((B \times X) \cap (C \times C)) \\ & = & \gamma((B \cap C) \times C) \\ & = & \mu(B \cap C) \\ & = & \mathrm{i}_*(\mu)(B) \end{array}$$

and similarly for ν. We also have:

$$\int d_X^p \mathrm{d}(\mathrm{i}_*\gamma) = \int d_C^p \mathrm{d}\gamma$$

As γ was chosen arbitrarily, we therefore have:

$$\mathrm{W}_p(\mu,\nu) \geq \mathrm{W}_p(\mathrm{i}_*(\mu), \mathrm{i}_*(\nu))$$

For the reverse inequality, let γ be a coupling between $\mathrm{i}_*(\mu)$ and $\mathrm{i}_*(\nu)$. By Lemma 3.1, $\mathrm{i}_*(\mu)$ and $\mathrm{i}_*(\nu)$ are Radon. So, by the remark after Lemma 2, γ is also Radon. Also, $\mathrm{supp}(\gamma) \subset C \times C$, since $\mathrm{supp}(\mathrm{i}_*(\mu)) \subseteq C$ and $\mathrm{supp}(\mathrm{i}_*(\nu)) \subseteq C$. So, by Lemma 3.2, $\mathrm{r}(\gamma)$ is a Radon probability measure on C^2 and $\gamma = \mathrm{i}_*(\mathrm{r}(\gamma))$.

Further, $\mathrm{r}(\gamma)$ is a coupling between μ and ν, for:

$$\begin{aligned} \mathrm{r}(\gamma)(B \times C) &= \gamma(B \times C) \\ &= \gamma((B \times X) \cap (C \times C)) \\ &= \gamma(B \times X) \quad (\text{as } \mathrm{supp}(\gamma) \subset C \times C) \\ &= \mu(B) \\ &= \mathrm{r}(\mu)(B) \end{aligned}$$

and similarly for ν. We then have:

$$\int d_X^p \mathrm{d}\gamma = \int d_X^p \mathrm{d}(\mathrm{i}_*\mathrm{r}\gamma) = \int d_C^p \mathrm{d}(\mathrm{r}\gamma)$$

As γ was chosen arbitrarily, we therefore have, as required:

$$\mathrm{W}_p(\mathrm{i}_*(\mu), \mathrm{i}_*(\nu)) \geq \mathrm{W}_p(\mu,\nu)$$

(2) Using part (1), we have:

$$\mathrm{W}_p(\mu,\nu) = \mathrm{W}_p(\mathrm{i}_*(\mathrm{r}(\mu)), \mathrm{i}_*(\mathrm{r}(\nu))) = \mathrm{W}_p(\mathrm{r}(\mu), \mathrm{r}(\nu))$$

□

With these technical lemmas established, we can now prove two theorems on spaces of Radon probability measures. For any metric space X and $p > 1$, define $P_p(X)$ to be the set of Radon probability measures on X with finite p-th moment, equipped with the W_p distance. For the first theorem we use the result in [2].

Theorem 1. *Let X be a metric space. Then $P_p(X)$ is a metric space.*

Proof. First $\mathrm{W}_p(\mu,\mu) = 0$ for any $\mu \in P_p(X)$ as $\Delta_*\mu$, the pushforward of μ along the diagonal $\Delta : X \to X \times X$ is a coupling between μ and itself. For the converse, suppose $\mathrm{W}_p(\mu,\nu) = 0$ for some $\mu,\nu \in P_p(X)$. Then $\mu,\nu \in P_1(X)$, and we have $\mathrm{W}_1(\mu,\nu) = 0$ as:

$$\mathrm{W}_1(\mu,\nu)^{1/p} = (\inf_\gamma \int d_X \mathrm{d}\gamma)^{1/p} \leq \inf_\gamma (\int d_X \mathrm{d}\gamma)^{1/p} \leq \inf_\gamma (\int d_X^p \mathrm{d}\gamma)^{1/p} = \mathrm{W}_p(\mu,\nu)$$

It follows that $\mu = \nu$ as it is shown in, e.g., [3, 1] that W_1 is a metric on $P_1(X)$.

Symmetry is evident. To show the triangle inequality, suppose that $\mu,\nu,\omega \in P_p(X)$. Let C be the closed set $\mathrm{supp}(\mu) \cup \mathrm{supp}(\nu) \cup \mathrm{supp}(\omega)$, separable by Lemma 1. Then $\mathrm{r}(\mu)$, $\mathrm{r}(\nu)$ and $\mathrm{r}(\omega)$ are probability measures on the separable space C, and so by [2], we have $\mathrm{W}_p(\mathrm{r}(\mu), \mathrm{r}(\omega)) \leq \mathrm{W}_p(\mathrm{r}(\mu), \mathrm{r}(\nu)) + \mathrm{W}_p(\mathrm{r}(\nu), \mathrm{r}(\omega))$. Then, by Lemmas 4.2 and 3.2, we see that $\mathrm{W}_p(\mu,\omega) \leq \mathrm{W}_p(\nu,\omega) + \mathrm{W}_p(\nu,\omega)$, as required. □

For the second theorem we use the result that the metric space of all probability measures with finite p-moment on a complete and separable space X is also complete and separable, being generated by the rational measures with finite support in a countable basis of X.

Theorem 2. *Let X be a complete metric space. Then $P_p(X)$ is a complete metric space generated by the finitely supported probability measures on X.*

Proof. To show that $P_p(X)$ is complete, let $\langle \mu_i \rangle_i$ be a Cauchy sequence in $P_p(X)$. Let C be the closure in $P_p(X)$ of $\bigcup_i \mathrm{supp}(\mu_i)$. As it is the closure of a countable union of separable sets, C is separable, and it is evidently complete. Applying Lemmas 3.2 and 4.2, we see that $\langle \mathrm{r}(\mu_i) \rangle_i$ is a Cauchy sequence in $P_p(C)$. Let μ be its limit there. As Lemma 4.1 shows that i_* is an isometric embedding, we see that $\langle \mathrm{i}_*(\mathrm{r}(\mu_i)) \rangle_i$ is a Cauchy sequence in $P_p(C)$ with limit $\mathrm{i}_*(\mu)$. But, by Lemma 3.2, $\langle \mathrm{i}_*(\mathrm{r}(\mu_i)) \rangle_i$ is exactly $\langle \mu_i \rangle_i$.

To see that the finitely supported probability measures are dense in $P_p(X)$, choose $\mu \in P_p(X)$ and $\varepsilon > 0$. Then, taking C to be the separable closed set $\mathrm{supp}(\mu)$, we have $\mathrm{r}(\mu) \in P_p(C)$. Then there is a finitely supported probability measure $\alpha \in P_p(C)$ at distance $\leq \varepsilon$ from $\mathrm{r}(\mu)$, and so we see that $\mathrm{i}_*(\alpha)$ is at distance $\leq \varepsilon$ from μ. Finally, $\mathrm{i}_*(\alpha)$ is finitely supported as, by Lemma 3.1, it has the same support as α. □

3. Wasserstein algebras

Wasserstein algebras are quantitative algebras, by which we mean algebras based on some category of metric spaces. Here we work with the category of metric spaces and nonexpansive maps, and its subcategory of complete metric spaces. These categories have all finite products with the one-point metric space as the final object and with the max metric on binary products $X \times Y$, where:

$$d_{X \times Y}(\langle x, y \rangle, \langle x', y' \rangle) = \max\{d_X(x, x'), d_Y(y, y')\}$$

We remark that nonexpansive maps are continuous.

A *finitary signature* Σ is a collection of operation symbols f and an assignment of an arity $n \in \mathbb{N}$ to each; given such a signature, we write $f : n$ to indicate that f is an operation symbol of arity n. A *(metric space) quantitative* Σ*-algebra* $(X, f_X\ (f \in \Sigma))$ is then a metric space X equipped with a nonexpansive function $f_X : X^n \to X$ for each operation symbol $f : n$. We often omit the suffix on the operation symbol and also confuse the metric space with the algebra. A *homomorphism* $h : X \to Y$ of Σ-algebras is a nonexpansive map $h : X \to Y$ such that for all $f : n$ and $x_1, \ldots, x_n \in X$ we have:

$$h(f_X(x_1, \ldots, x_n)) = f_Y(h(x_1), \ldots, h(x_n))$$

This defines a category of quantitative Σ-algebras and homomorphisms; it has an evident subcategory of complete quantitative Σ-algebras.

Wasserstein algebras have a probabilistic choice operation over metric spaces; barycentric algebras form a basic class of algebras with such an operation. A *barycentric algebra* (or *abstract convex set*) is a set X equipped with binary operations $+_r$ for every real number

$r \in [0,1]$ such that the following equational laws hold:

(B1) $$x +_1 y = x$$

(B2) $$x +_r x = x$$

(SC) $$x +_r y = y +_{1-r} x$$

(SA) $$(x +_p y) +_r z = x +_{pr} (y +_{\frac{r-pr}{1-pr}} z) \quad \text{provided } r < 1, p < 1$$

SC stands for *skew commutativity* and SA for *skew associativity*. Homomorphisms of barycentric algebras are termed *affine*.

One can inductively define finite convex sums in a barycentric algebra by:

$$\sum_{i=1,n} r_i x_i = x_1 +_{r_1} \sum_{i=2,n} \frac{r_i}{1-r_1} x_i$$

for $n \geq 2$, $r_1 \neq 1$, with the other cases being evident. With this definition, one can prove all the expected laws; further, affine maps preserve such finite convex sums. We should remark that finite convex sums have been axiomatised as *convex spaces* where the sums are required to obey the equations:

$$\sum_{i=1}^{n} \delta_{ik} x_i = x_k$$

where δ_{ik} is the Kronecker symbol, and

$$\sum_{i=1}^{n} p_i (\sum_{k=1}^{m} q_{ik} x_k) = \sum_{k=1}^{m} (\sum_{i=1}^{n} p_i q_{ik}) x_k$$

For any set X, the barycentric algebra $P_f(X)$ of probability distributions over X with finite support is the free barycentric algebra over X, with universal algebra the Dirac delta function (see [8, 10]); if $f : X \to B$ is a map to a barycentric algebra B, then the unique extension of $f : X \to B$ along δ is given by:

$$\overline{f}(\mu) = \sum_{s \in \mathrm{supp}(\mu)} \mu(s) f(s)$$

For further bibliographic references to barycentric algebras and convex spaces, and historical discussion, see [6].

A *p-Wasserstein algebra* is a quantitative algebra $(X, +_r : X^2 \to X \ (r \in [0,1]))$ forming a barycentric algebra such that for all $x, x', y, y' \in X$ we have:

$$d(x +_r y, x' +_r y')^p \quad \leq \quad r d(x,x')^p + (1-r) d(y,y')^p \quad (*)$$

We remark that the hypothesis of non-expansiveness of the $+_r$ is redundant as, setting

$$m = \max\{d(x,x'), d(y,y')\}$$

we have:

$$\begin{array}{rcl} d(x +_r y, x' +_r y') & \leq & (r d(x,x')^p + (1-r) d(y,y')^p)^{1/p} \\ & \leq & (r m^p + (1-r) m^p)^{1/p} \\ & = & m \end{array}$$

We gather some other remarks about convex combinations into:

Lemma 5. *Let $(X, +_r (r \in [0,1]))$ be a p-Wasserstein quantitative algebra. Then the following hold:*

(1) The functions $+_r$ are $r^{1/p}$-Lipschitz in their first argument and $(1-r)^{1/p}$-Lipschitz in their second argument.
(2) Considered as a function of r, $x +_r y$ is $1/p$-Hölder continuous.
(3) The following generalisation of $(*)$ to finite convex sums:

$$d(\textstyle\sum_i r_i x_i, \sum_i r_i x_i')^p \quad \leq \quad \sum_i r_i d(x_i, x_i')^p$$

holds.

Proof. (1) Taking $y = y'$ $(x = x')$ in $(*)$ we respectively obtain:

$$d(x +_r y, x' +_r y) \leq r^{1/p} d(x, x') \quad \text{and} \quad d(x +_r y, x +_r y') \leq (1-r)^{1/p} d(y, y')$$

as required.

(2) Fix $x, y \in X$ and choose $r, s \in [0,1]$. Suppose that $s \leq r$ and set $e = r - s$. Assuming $e \neq 1$, we have

$$x +_r y = x +_e (x +_{s/1-e} y) \text{ and } x +_s y = y +_e (x +_{s/1-e} y)$$

So by part 1 we have

$$d(x +_r y, x +_s y) \leq e^{1/p} d(x, y)$$

This also holds when $e = 1$ as then $r = 1$ and $s = 0$. As $e = d(r, s)$, we therefore have:

$$d(x +_r y, x +_s y) \leq d(x, y) d(r, s)^{1/p}$$

By symmetry this also holds when $r \leq s$. As $p \geq 1$, $0 < 1/p \leq 1$, so, as a function of r, $x +_r y$ is $1/p$-Hölder continuous, as required.

(3) This is a straightforward induction.

□

Lemma 6. *For any metric space X, $P_p(X)$ equipped with the standard convex combination operations $\mu +_p \nu$, for $p \in [0,1]$ forms a p-Wasserstein algebra.*

Proof. The set of Radon probability measures on X with finite p-th moment is evidently closed under the standard convex combination operations $\mu +_r \nu$, for $r \in [0,1]$, and so forms a barycentric algebra.

It remains to show that for any $\mu, \mu', \nu, \nu' \in P_p(X)$ and $r \in [0,1]$ we have:

$$\mathrm{W}_p(\mu +_r \nu, \mu' +_r \nu')^p \quad \leq \quad r\mathrm{W}_p(\mu, \mu')^p + (1-r)\mathrm{W}_p(\nu, \nu')^p$$

To this end, choose $\mu, \mu', \nu, \nu' \in P_p(X)$. Let α be a coupling between μ and μ' and let β be a coupling between ν and ν'. Then $\alpha +_r \beta$ is a coupling between $\mu +_r \nu$ and $\mu' +_r \nu'$, and we have:

$$\textstyle\int d_X^p \mathrm{d}(\alpha +_r \beta) \quad = \quad \int d_X^p \mathrm{d}\alpha +_r \int d_X^p \mathrm{d}\beta$$

The conclusion then follows, as we have:

$$\begin{array}{rcl} \mathrm{W}_p(\mu +_r \nu, \mu' +_r \nu')^p & = & \inf_\gamma \left(\int d_X^p \mathrm{d}\gamma\right) \\ & \leq & \inf_{\alpha,\beta} \left(\int d_X^p \mathrm{d}(\alpha +_r \beta)\right) \\ & = & \inf_{\alpha,\beta} \left(\int d_X^p \mathrm{d}\alpha +_r \int d_X^p \mathrm{d}\beta\right) \\ & = & \inf_\alpha \left(\int d_X^p \mathrm{d}\alpha\right) +_r \inf_\beta \left(\int d_X^p \mathrm{d}\beta\right) \\ & = & \mathrm{W}_p(\mu, \nu)^p + \mathrm{W}_p(\mu', \nu')^p \end{array}$$

□

It follows from Lemmas 5 and 3 that Theorem 2 implies the theorem in that the metric space of all probability measures with finite p-moment over a complete and separable sapce X is also complete and separable, being generated by the rational measures with finite support in a countable basis of X. For, given the joint nonexpansiveness of $x +_r y$ as a function of x and y and its Hölder continuity as a function of r, it follows that every finitely supported probability measure is a limit of rational convex combinations of finitely supported probability measures with support included in a generating set of X.

We next turn to characterising the free p-Wasserstein algebras, For any metric space X, let $P_{f,p}(X)$ be the sub-p-Wasserstein algebra of P_p of the finitely supported probability measures.

Theorem 3. *For any metric space X, $P_{f,p}(X)$ is the free p-Wasserstein algebra over X, with universal arrow the Dirac delta function $\delta : X \to P_{f,p}(X)$.*

Proof. First, $\delta : X \to P_{f,p}(X)$ is an isometric embedding. For, for any $x, y \in X$, the only coupling between $\delta(x)$ and $\delta(y)$ is $\delta(\langle x, y\rangle)$, and so we have:

$$W_p(\delta(x)), \delta(y)) = (d(x,y)^p)^{1/p} = d(x,y)$$

Next, let $f : X \to W$ be any nonexpansive affine map to a p-Wasserstein algebra B. We know from the above discussion of free barycentric algebras that there is a unique affine map $\overline{f} : P_{f,p}(X) \to W$, given by the formula:

$$\overline{f}(\mu) = \sum_{s \in \mathrm{supp}(\mu)} \mu(s) f(s)$$

It only remains to show that $\overline{f}$ is nonexpansive. Let μ, ν be finitely supported probability measures with respective finite non-empty supports S and T. Let γ be a coupling between them. Then γ has support $\subseteq S \times T$. Consequently we can write it as a convex sum, as follows:

$$\gamma = \sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)\delta(\langle s, t\rangle)$$

As μ and ν are the marginals of γ, we have:

$$\mu = \sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)\delta(s) \quad \text{and} \quad \nu = \sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)\delta(t)$$

We can now calculate:

$$\begin{array}{rcl} d(\overline{f}(\mu), \overline{f}(\nu)) & = & d(\overline{f}(\sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)\delta(s)), \overline{f}(\sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)\delta(t))) \\ & = & d(\sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)\overline{f}(\delta(s)), \sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)\overline{f}(\delta(t))) \\ & = & d(\sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)f(s), (\sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)f(t)) \\ & & \qquad\qquad\qquad\qquad\qquad\qquad (\text{as } f = \overline{f}\delta) \\ & \leq & \left(\sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)d(f(s), f(t))^p\right)^{1/p} \\ & & \qquad\qquad\qquad\qquad\qquad\qquad (\text{by Lemma 5.3}) \\ & \leq & \left(\sum_{s \in S, t \in T} \gamma(\langle s, t\rangle)d(s, t)^p\right)^{1/p} \\ & & \qquad\qquad\qquad\qquad\qquad\qquad (\text{as } f \text{ is nonexpansive}) \\ & = & \left(\int d_X^p \mathrm{d}\gamma\right)^{1/p} \end{array}$$

As γ was an arbitrary coupling between μ and ν we therefore have

$$d(\overline{f}(\mu), \overline{f}(\nu)) \leq W_p(\mu, \nu)$$

as required. □

To characterise the free complete p-Wasserstein algebras, we need two lemmas, one on the completion of metric spaces, and the other on the completion of quantitative Σ-algebras. The usual completion $\overline{X}$ of a metric space X by equivalence classes $[\mathbf{x}]$ of Cauchy sequences is the free complete metric space over X, with universal arrow the isometric embedding $c_X : X \to \overline{X}$, where $c_X(x) = [\langle x\rangle_{n \in \mathbb{N}}]$. That is, for any complete metric space Y and nonexpansive function $f : X \to Y$ there is a unique nonexpansive function $\overline{f} : \overline{X} \to Y$ extending f along c_X. This function is given by the formula:

$$\overline{f}([\mathbf{x}]) = \lim_n f(\mathbf{x}_n)$$

There is a simple criterion for when an isometric embedding of a metric space in another is a universal arrow:

Lemma 7. *Let $\theta : X \to Y$ be an isometric embedding of metric spaces with Y complete, and $\theta(X)$ generating Y. Then Y is the free complete metric space over X, with universal arrow θ.*

Proof. We can assume without loss of generality that θ is an inclusion mapping. We show that $\overline{\theta}$ is an isometry. To show it is onto, choose $y \in Y$. As X generates Y, y is a limit of some Cauchy sequence $\mathbf{x}$ in X and we have

$$\overline{\theta}([\mathbf{x}]) = \lim_n \theta(\mathbf{x}_n) = \lim_n \mathbf{x}_n = y$$

as required.

To show $\overline{\theta}$ preserves distances, we calculate:

$$d(\overline{\theta}([\mathbf{x}]), \overline{\theta}([\mathbf{y}]) = d(\lim_n \mathbf{x}_n, \lim_n \mathbf{y}_n) = \lim_n d(\mathbf{x}_n, \mathbf{y}_n) = d([\mathbf{x}], [\mathbf{y}])$$

So $\overline{\theta}$ is indeed an isometry. It follows that θ is a universal arrow as c_X is, and we are done. □

The criterion of Lemma 7 extends to Σ-algebras:

Lemma 8. *Let X, Y be quantitative Σ-algebras and let $\theta : X \to Y$ be both an algebra homomorphism and an isometric embedding. Suppose that Y is complete as a metric space, and $\theta(X)$ generates Y as a metric space. Then Y is the free complete quantitative Σ-algebra over X, with universal arrow θ.*

Proof. We can assume without loss of generality that θ is an inclusion mapping. Let $h : X \to Z$ be any nonexpansive homomorphism from X to a complete quantitative algebra Z. By Lemma 1, h extends to a unique nonexpansive map $\overline{h} : Y \to Z$, and we only have to show that this is a homomorphism. Taking a binary operation f as an example, let x, y be in Y. As X generates Y as a metric space there are Cauchy sequences $\mathbf{x}$ and $\mathbf{y}$ in X with respective limits x and y, and we calculate:

$$\begin{array}{rcl} \overline{h}(f_Y(x,y)) & = & \overline{h}(f_Y(\lim_n \mathbf{x}_n, \lim_n \mathbf{y}_n)) \\ & = & \overline{h}(\lim_n f_Y(\mathbf{x}_n, \mathbf{y}_n)) \\ & = & \lim_n \overline{h}(f_Y(\mathbf{x}_n, \mathbf{y}_n)) \\ & = & \lim_n \overline{h}(f_X(\mathbf{x}_n, \mathbf{y}_n)) \\ & = & \lim_n h(f_X(\mathbf{x}_n, \mathbf{y}_n)) \\ & = & \lim_n f_Z(h(\mathbf{x}_n), h(\mathbf{y}_n)) \\ & = & f_Z(\lim_n h(\mathbf{x}_n), \lim_n h(\mathbf{y}_n)) \\ & = & f_Z(\lim_n \overline{h}(\mathbf{x}_n), \lim_n \overline{h}(\mathbf{y}_n)) \\ & = & f_Z(\overline{h}(\lim_n \mathbf{x}_n), \overline{h}(\lim_n \mathbf{y}_n)) \\ & = & f_Z(\overline{h}(x), \overline{h}(y)) \end{array}$$

□

We can now prove our main theorem on free complete p-Wasserstein algebras:

Theorem 4. *For any complete metric space X, $P_p(X)$ is the free complete p-Wasserstein algebra over X, with universal arrow the Dirac delta function.*

Proof. Theorem 2 and Lemma 6 tell us that $P_p(X)$ is a complete p-Wasserstein algebra. Further, from Theorem 2 we also have that $P_p(X)$ is generated by $P_{f,p}(X)$, and that is both a subalgebra and a submetric space of $P_p(X)$. By Lemma 8 we then have that $P_p(X)$ is the free quantitative algebra with signature the $+_r$ over $P_{f,p}(X)$, with universal arrow the inclusion. It is therefore, in particular, the free p-Wasserstein algebra over $P_p(X)$, with universal arrow the inclusion.

In turn, Theorem 3 tells us that $P_{f,p}(X)$ is the free p-Wasserstein algebra over X, with universal arrow the Dirac function. Combining these two free algebra assertions, we see that, as required, $P_p(X)$ is the free p-Wasserstein algebra over X, with universal arrow the Dirac delta function. □

There is a disconnect between the present work and that in [7] where extended metrics are used. We sketch the straightforward extension of the above freeness results to extended metric spaces. An extended metric is a function $d : X^2 \to \overline{\mathbb{R}}_+$ to the extended reals, defined in the usual way using the natural extension of addition to $\overline{\mathbb{R}}_+$, where $x + \infty = \infty + x = \infty$

and the natural extension of the order where $x \leq \infty$; see [12] for information on such spaces. The topology on an extended metric space (X, d) is defined in the usual way using open balls. Nonexpansive functions are defined in the usual way, and are continuous. Cauchy sequences are defined in the usual way, as are then complete extended metric spaces. The completion of an extended metric space is defined as usual and the function $c_X : X \to \overline{X}$ so obtained is an isometry and is universal. The analogue of Lemma 7 goes through, with the same proof, as does that of Lemma 8, with the evident definitions of quantitative Σ-algebras over extended metric spaces and their homomorphisms. Quantitative p-Wasserstein algebras over an extended metric space are also defined as before. In [7] these p-Wasserstein algebras are axiomatised by the following axiom-scheme:

$$\{x =_{q_1} y, x' =_{q_2} y'\} \vdash x +_e x' =_r y +_e y'$$

where q_1, q_2, q, r range over rationals in $[0, 1]$ such that $rq_1^p + (1-r)q_2^p \leq q^p$.

The extension of Lemma 6 to extended metric spaces then goes through, with the same proofs, as do the following extensions of Theorems 3 and 4:

Theorem 5. *For any metric space X, $P_{f,p}(X)$ is the free extended metric p-Wasserstein algebra over X, with universal arrow the Dirac delta function $\delta : X \to P_{f,p}(X)$.*

Theorem 6. *For any complete metric space X, $P_p(X)$ is the free complete extended metric p-Wasserstein algebra over X, with universal arrow the Dirac delta function.*

References

[1] G. Basso, A Hitchhiker's guide to Wasserstein distances, available at: `http://n.ethz.ch`, 2015.

[2] P. Clement and W. Desch, An elementary proof of the triangle inequality for the Wasserstein metric, *Proceedings of the American Mathemtical Society*, **136**(1) 333–339, 2008.

[3] D. E. Edwards, A simple proof in Monge-Kantorovich duality theory, *Studia Math.*, **200**(1), 67–77, 2010.

[4] T. Fritz and P. Perrone, A Probability Monad as the Colimit of Finite Powers, arXiv: 1712.05363, 2017.

[5] D. H. Fremlin, *Measure Theory, Volume 4: Topological Measure Spaces*, Torres Fremlin, 2006.

[6] K. Keimel and G. D. Plotkin, Mixed powerdomains for probability and nondeterminism, *Logical Methods in Computer Science* **13**(1:2), 1–84, 2017.

[7] R. Mardare, P. Panangaden, and G. D. Plotkin, Quantitative Algebraic Reasoning, *Proc. 31st. LICS* (eds. Martin Grohe, Eric Koskinen, and Natarajan Shankar), 700–709, 2016.

[8] W. D. Neumann. On the quasivariety of convex subsets of affine spaces. *Archiv dxer Mathematik*, 21:11–16, 1970.

[9] K. R. Parthasarathy, *Probability measures on metric spaces*, Academic Press, 1967.

[10] A. B. Romanowska and J. D. H. Smith, *Modes*, World Scientific, Singapore, 2002.

[11] C. Villani, *Optimal transport: old and new*, Vol. 338, Springer Science & Business Media, 2008.

[12] D. Burago, Y. D. Burago, and S. Ivanov, *A course in metric geometry*, Vol. 33, American Mathematical Soc., 2001.