

## Free complete Wasserstein algebras

Mardare, Radu; Panangaden, Prakash; Plotkin, Gordon D.

*Published in:*  
Logical Methods in Computer Science

*DOI (link to publication from Publisher):*  
[10.23638/LMCS-14\(3:19\)2018](https://doi.org/10.23638/LMCS-14(3:19)2018)

*Creative Commons License*  
CC BY 4.0

*Publication date:*  
2018

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Mardare, R., Panangaden, P., & Plotkin, G. D. (2018). Free complete Wasserstein algebras. *Logical Methods in Computer Science*, 14(3), Article 19. [https://doi.org/10.23638/LMCS-14\(3:19\)2018](https://doi.org/10.23638/LMCS-14(3:19)2018)

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

## FREE COMPLETE WASSERSTEIN ALGEBRAS

RADU MARDARE, PRAKASH PANANGADEN, AND GORDON D. PLOTKIN

Department of Computer Science, Aalborg University  
*e-mail address:* mardare@cs.aau.dk

School of Computer Science, McGill University  
*e-mail address:* prakash@cs.mcgill.ca

Laboratory for Foundations of Computer Science, School of Informatics, University of Edinburgh  
*e-mail address:* gdp@inf.ed.ac.uk

---

**ABSTRACT.** We present an algebraic account of the Wasserstein distances  $W_p$  on complete metric spaces, for  $p \geq 1$ . 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 with finite moments of order  $p$ , equipped with the Wasserstein distance as metric and with the usual binary convex sums as operations.

### 1. INTRODUCTION

The denotational semantics of probabilistic programs generally makes use of one or another monad of probability measures. In [16, 8], Lawvere and Giry provided the first example of such a monad, over the category of measurable spaces; Giry also gave a second example, over Polish spaces. Later, in [13, 12], Jones and Plotkin provided a monad of evaluations over directed complete posets. More recently, in [10], Heunen, Kammar, et al provided such a monad over quasi-Borel spaces.

Metric spaces (and nonexpansive maps) form another natural category to consider. The question then arises as to what metric to take over probability measures, and the Kantorovich-Wasserstein, or, more generally, the Wasserstein distances of order  $p > 1$  (generalising from  $p = 1$ ), provide natural choices. A first result of this kind was given in [2], where it was shown that the Kantorovich distance provides a monad over the subcategory of 1-bounded complete separable metric spaces.

One purpose of such monads is to make probabilistic choice operations available, and so an algebraic account of the monads in terms of these operations is of interest; we focus here on such results in the metric context. Barycentric algebras axiomatise probabilistic choice. They have binary convex combination operations  $x +_r y$ , for  $r \in [0, 1]$ ; one can think



of  $x +_r y$  as choosing to continue with  $x$  with probability  $r$  and with  $y$  with probability  $1 - r$ . The operations are required to obey appropriate laws.

In this paper we show that the Wasserstein distance of order  $p$  yields a monad  $P_p$  on the category of complete metric spaces and nonexpansive maps and that this monad can be characterised as the free algebra monad for barycentric algebras over complete metric spaces where the barycentric operations are required to be nonexpansive and are subject to an appropriate Wasserstein condition. The probability measures in  $P_p(X)$  are required to be Radon (equivalently, tight) and to have finite  $p$ -moment. The important Kantorovich case of this result, where  $p = 1$ , was essentially already established in [6], as it is an immediate consequence of Theorem 5.2.1 there.

In [17] a somewhat different direction was taken where monads on (extended) metric spaces are defined algebraically, using a quantitative analogue 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. One then seeks to characterise the action of the monad on as wide a variety of spaces as possible.

This is part of a program begun in [17] to establish a quantitative algebraic theory of effects in programming languages. In particular, in [17] 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. Comparing this result to the results in this paper and in [6], it is worth remarking that probability measures on 1-bounded spaces automatically have finite moments, and probability measures on separable complete spaces are automatically Radon.

In Section 2, we discuss the Wasserstein distance,  $W_p$  of order  $p$  between probability measures on a metric space  $X$ . In Theorem 2.6 we show that the distance is a metric on the Radon probability measures with finite moments of order  $p$ . To do so, we make use of a result in [4] that the triangle inequality holds for all probability measures if  $X$  is separable. Here, and throughout the paper, we reduce the general case to the separable case using a lemma, due to Basso, which states that the support of any probability measure on a metric space is separable.

Next, in Theorem 2.7 we show that if  $X$  is complete then the Wasserstein metric of order  $p$  on the Radon probability measures with finite moments of order  $p$  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 Wasserstein metric of order  $p$  on all probability measures over  $X$  with finite moments of order  $p$  is also complete and separable, being generated by the rational measures with finite support in a countable basis of  $X$  (see, e.g., [24, Theorem 6.18] and the bibliographic discussion there). The section concludes with a side-excursion, a discussion of weak convergence: Theorem 2.12 generalises the characterisation of the topology induced by the Wasserstein metrics in terms of weak convergence given by [24, Theorem 6.9] from complete separable spaces to all complete spaces.

In Section 3 we discuss the algebraic aspect of these spaces. In particular, in Theorem 3.6 we show that the Radon probability measures with finite moments of order  $p$  on a complete metric space  $X$ , equipped with the  $W_p$  metric and binary convex sums, form the free complete Wasserstein algebra of order  $p$  over  $X$ . To do this, we first characterise the free Wasserstein algebra of order  $p$  over a metric space; the characterisation follows straightforwardly from the well-known characterisation of the free barycentric algebra over a set as the natural such

algebra on the finite probability measures on the set. Having done so, we pass to the case of complete metric spaces using a general theorem on completions of metric algebras, combined with Theorem 2.7.

Finally, we discuss some loose ends. First, there is a certain disconnect between our work and that of [17] in that here we use standard metric spaces, whereas there the more general framework of extended metric spaces is employed where distances can be infinite. We sketch how to bridge this disconnect at the end of Section 3. Second, there are other natural algebraic approaches to probability. Convex spaces are an algebraic formulation of finite convex combinations  $\sum_{i=1}^n r_i x_i$ . Midpoint algebras have a midpoint operation  $x \oplus y$ : one can think of  $x \oplus y$  as providing a fair choice between  $x$  and  $y$ , and so being equivalent to  $x +_{0.5} y$ . We sketch how our algebraic characterisation of the Wasserstein monads on complete metric spaces can be rephrased in terms of either of these alternative approaches.

We affectionately dedicate this paper to Furio Honsell on the occasion of his 60th birthday.

## 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 [7, 19]. By probability measure we mean a Borel probability measure. Given such a probability measure  $\mu$  on a Hausdorff space  $X$ , we say that a Borel set  $B$  is *compact inner regular* (for  $\mu$ ) if:

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

Then  $\mu$  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  $\mu(B \setminus C) < \varepsilon$ ). Every Radon measure is tight, 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  $\mu$  on a topological space  $X$  is

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

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

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

(writing  $\mu(s)$  instead of  $\mu(\{s\})$ ). We say that  $\mu$  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 2.1.** *Every probability measure on a metric space has separable support.*

*Proof.* Let  $\mu$  be a probability measure on a metric space. For every  $\varepsilon > 0$  let  $C_\varepsilon$  be a maximal set of points in  $\text{supp}(\mu)$  with the property that all points are at distance  $\geq \varepsilon$  apart. As any two open balls with centre in  $C_\varepsilon$  and radius  $\frac{1}{2}\varepsilon$  are disjoint and each such ball has  $\mu$ -measure  $> 0$ ,  $C_\varepsilon$  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  $C_\varepsilon$ , any point in  $\text{supp}(\mu)$  is at distance  $< \varepsilon$  from some point in  $C_\varepsilon$ . It follows that the countable set  $\bigcup_{n \geq 0} C_{2^{-n}}$  is dense in  $\text{supp}(\mu)$ .  $\square$

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

**Lemma 2.2.** *Let  $\gamma$  be a coupling between two probability measures  $\mu$  and  $\nu$  on a topological space  $X$ . Then:*

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

*Further, if  $\mu$  and  $\nu$  are tight, so is  $\gamma$ . Moreover, in the case that  $X$  is a metric space, if  $\mu$  and  $\nu$  are Radon, so is  $\gamma$ .*

*Proof.* For the first part, suppose that  $(x, y) \in \text{supp}(\gamma)$  and let  $U$  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  $x \in \text{supp}(\mu)$ . Similarly  $y \in \text{supp}(\nu)$ .

For the second part, choose  $\varepsilon > 0$ . As  $\mu$  and  $\nu$  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  $\mu(X \setminus C) < \frac{1}{2}\varepsilon$  and  $\nu(Y \setminus D) < \frac{1}{2}\varepsilon$ . Then we have:

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

So, as  $\varepsilon$  was arbitrary,  $X \times Y$  is compact inner regular for  $\gamma$ , as required. The last part is immediate as all tight probability measures on metric spaces are Radon.  $\square$

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

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

is finite. To see that the existence of the  $p$ -th moment does not depend on the choice of  $x_0$ , recollect the inequality  $(a + b)^p \leq 2^{p-1}(a^p + b^p)$  for  $a, b \in \mathbb{R}_+$  and  $p \geq 1$ . Then, for any  $x, y, z \in X$ , we have:

$$d(x, y)^p \leq 2^{p-1}(d(x, z)^p + d(z, y)^p) \quad (2.1)$$

and the conclusion follows, taking  $z = x_0$  and  $y$  to be any other choice. Note that finite probability measures with finite support have finite moments of all orders.

**Example 2.3.** One can obtain examples of measures with countable support, and with or without finite moments, by using the fact that the Dirichlet series  $\zeta(s) = \sum_{n \geq 1} \frac{1}{n^s}$  converges for reals  $s > 1$  and diverges for  $s = 1$ . Taking the natural numbers as a metric space with the usual Euclidean metric  $d(m, n) = |m - n|$ , one sees that, for  $q \geq 1$ , the discrete probability measure  $D_q$  where  $D_q(n) = \zeta(q+1)^{-1} \frac{1}{n^{q+1}}$  for  $n \geq 1$  and  $D_q(0) = 0$ , has finite  $p$ -moment if, and only if,  $p < q$ .

The Wasserstein distance of order  $p$ ,  $W_p$  is defined between probability measures with finite moments of order  $p$  by:

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

where  $\gamma$  runs over the couplings between  $\mu$  and  $\nu$ . To see that it is well-defined for probability measures with finite  $p$ -moment, one again invokes (1), this time to get an integrable upper bound on  $d_X^p$  as follows:

$$\int d_X^p d\gamma \leq 2^{p-1} \left( \int d(-, x_0)^p d\gamma + \int d(x_0, -)^p d\gamma \right) = 2^{p-1} \left( \int d(-, x_0)^p d\mu + \int d(x_0, -)^p d\nu \right)$$

The Wasserstein distance is monotonic in  $p$ , that is,  $W_p(\mu, \nu) \leq W_q(\mu, \nu)$  for  $1 \leq p \leq q$  and  $\mu, \nu$  with finite moments of order  $q$ . This is an immediate consequence of the inequality  $(\int f^p d\mu)^{1/p} \leq (\int f^q d\mu)^{1/q}$ , which holds for any probability measure  $\mu$  and any  $f \geq 0$  with  $f^q$  integrable w.r.t.  $\mu$  (this inequality is itself a straightforward consequence of Hölder's inequality).

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  $i_*(\mu)$  for the pushforward of a finite measure  $\mu$  on  $C$  along the inclusion  $i : C \rightarrow X$  of  $C$  in  $X$ , so  $i_*(\mu)(B) = \mu(B \cap C)$ ; and we write  $r(\nu)$  for the restriction of a finite measure  $\nu$  on  $X$  to a finite measure on  $C$ , so  $r(\nu)(B) = \nu(B)$ . Note that  $r(i_*(\mu)) = \mu$ .

**Lemma 2.4.** *Let  $C$  be a closed subset of a metric space  $X$ .*

- (1) *If  $\mu$  is a Radon probability measure on  $C$ , then  $i_*(\mu)$  is a Radon probability measure on  $X$  with the same support as  $\mu$ . Further,  $i_*(\mu)$  has finite moment of order  $p$  if  $\mu$  does.*
- (2) *If  $\nu$  is a Radon probability measure on  $X$  with support included in  $C$  then  $r(\nu)$  is a Radon probability measure on  $C$ , and we have:*

$$\nu = i_*(r(\nu))$$

*If, further,  $\nu$  has finite moment of order  $p$ , so does  $r(\nu)$ .*

*Proof.* (1) Let  $\mu$  be a Radon probability measure on  $C$ , and note that  $C$  must then be non-empty. It is straightforward to check that  $i_*(\mu)$  is a Radon probability measure on  $X$  with the same support as  $\mu$ . Next, choosing  $x_0 \in C$ , as

$$\int d_X(x_0, -)^p d(i_*\mu) = \int d_C(x_0, -)^p d\mu$$

we see that  $i_*(\mu)$  has finite moment of order  $p$  if  $\mu$  does.

- (2) Let  $\nu$  be a Radon probability measure on  $X$  with support included in  $C$ . It is straightforward to check that  $r(\nu)$  is a Radon probability measure on  $C$ . Regarding the equality, for any Borel set  $B$  of  $X$  we have:

$$i_*(r(\nu))(B) = r(\nu)(B \cap C) = \nu(B \cap C) = \nu(B)$$

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

$$\int d_C(x_0, -)^p d(r(\nu)) = \int d_X(x_0, -)^p d(i_*(r(\nu))) = \int d_X(x_0, -)^p d\nu \quad \square$$

**Lemma 2.5.** *Let  $C$  be a closed subset of a metric space  $X$ . Then*

(1) For any Radon probability measures  $\mu, \nu$  on  $X$  with finite moments of order  $p$  we have:

$$W_p(\mu, \nu) = W_p(i_*(\mu), i_*(\nu))$$

(2) For any Radon probability measures  $\mu, \nu$  on  $X$  with finite moments of order  $p$  whose support is included in  $C$ , we have:

$$W_p(\mu, \nu) = W_p(r(\mu), r(\nu))$$

*Proof.* (1) Let  $\gamma$  be a coupling between  $\mu$  and  $\nu$ . Then  $(i \times i)_*(\gamma)$  is a coupling between  $i_*(\mu)$  and  $i_*(\nu)$ , as  $(\pi_0)_*((i \times i)_*(\gamma)) = i_*((\pi_0)_*(\gamma)) = i_*(\mu)$ , and similarly for  $\nu$ . We also have:

$$\int d_X^p d((i \times i)_*\gamma) = \int d_C^p d\gamma$$

As  $\gamma$  was chosen arbitrarily, we therefore have:

$$W_p(\mu, \nu) \geq W_p(i_*(\mu), i_*(\nu))$$

For the reverse inequality, let  $\gamma$  be a coupling between  $i_*(\mu)$  and  $i_*(\nu)$ . By Lemma 2.4.1,  $i_*(\mu)$  and  $i_*(\nu)$  are Radon. So, by Lemma 2.2,  $\gamma$  is also Radon. Also,  $\text{supp}(\gamma) \subseteq C \times C$ , since  $\text{supp}(i_*(\mu)) \subseteq C$  and  $\text{supp}(i_*(\nu)) \subseteq C$ . So, by Lemma 2.4.2,  $r(\gamma)$  is a Radon probability measure on  $C^2$  and  $\gamma = (i \times i)_*(r(\gamma))$ .

Further,  $r(\gamma)$  is a coupling between  $\mu$  and  $\nu$ , for:

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

and similarly for  $\nu$ . We then have:

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

As  $\gamma$  was chosen arbitrarily, we therefore have, as required:

$$W_p(i_*(\mu), i_*(\nu)) \geq W_p(\mu, \nu)$$

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

$$W_p(\mu, \nu) = W_p(i_*(r(\mu)), i_*(r(\nu))) = W_p(r(\mu), r(\nu)) \quad \square$$

With these technical lemmas established, we can now prove two theorems on spaces of Radon probability measures. For any metric space  $X$  and  $p \geq 1$ , define  $P_p(X)$  to be the set of Radon probability measures on  $X$  with finite moments of order  $p$ , equipped with the  $W_p$  distance. For the first theorem we use the result in [4] that the triangle inequality holds for all probability measures if  $X$  is separable.

**Theorem 2.6.** *Let  $X$  be a metric space. Then  $P_p(X)$  is a metric space.*

*Proof.* First  $W_p(\mu, \mu) = 0$  for any  $\mu \in P_p(X)$  as  $\Delta_*\mu$ , the pushforward of  $\mu$  along the diagonal  $\Delta : X \rightarrow X \times X$ , is a coupling between  $\mu$  and itself. For the converse, choose  $\mu, \nu \in P_p(X)$ , and suppose  $W_p(\mu, \nu) = 0$ . Then  $W_1(\mu, \nu) = 0$ , as  $W_q$  is monotonic in  $q$ . It follows that  $\mu = \nu$  as  $W_1$  is a metric on  $P_1(X)$  (see, e.g., [5, 1]).

Symmetry is evident. To show the triangle inequality, suppose that  $\mu, \nu, \omega \in P_p(X)$ . Let  $C$  be the closed set  $\text{supp}(\mu) \cup \text{supp}(\nu) \cup \text{supp}(\omega)$ , separable by Lemma 2.1. Then  $r(\mu)$ ,

$r(\nu)$  and  $r(\omega)$  are probability measures on the separable space  $C$ , and so by [4], we have  $W_p(r(\mu), r(\omega)) \leq W_p(r(\mu), r(\nu)) + W_p(r(\nu), r(\omega))$ . Then, by Lemmas 2.5.2 and 2.4.2, we see that  $W_p(\mu, \omega) \leq W_p(\nu, \omega) + W_p(\nu, \mu)$ , as required.  $\square$

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

**Theorem 2.7.** *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  $\bigcup_i \text{supp}(\mu_i)$ . By Lemma 2.1, each set  $\text{supp}(\mu_i)$  is separable, and so  $C$  is itself separable, being the closure of a countable union of separable sets. Further,  $C$  is complete as it is a closed subset of a complete space. So  $P_p(C)$  is complete. Applying Lemmas 2.4.2 and 2.5.2, we see that  $\langle r(\mu_i) \rangle_i$  is a Cauchy sequence in  $P_p(C)$ . Let  $\mu$  be its limit there. As Lemma 2.5.1 shows that  $i_*$  is an isometric embedding, we see that  $\langle i_*(r(\mu_i)) \rangle_i$  is a Cauchy sequence in  $P_p(C)$  with limit  $i_*(\mu)$ . But, by Lemma 2.4.2,  $\langle i_*(r(\mu_i)) \rangle_i$  is  $\langle \mu_i \rangle_i$ .

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  $\text{supp}(\mu)$ , we have  $r(\mu) \in P_p(C)$ . Then, as  $P_p(C)$  is generated by the rational measures with finite support, there is a finitely supported probability measure  $\alpha \in P_p(C)$  at distance  $\leq \varepsilon$  from  $r(\mu)$ , and so we see that  $i_*(\alpha)$  is at distance  $\leq \varepsilon$  from  $\mu$ . Finally,  $i_*(\alpha)$  is finitely supported as, by Lemma 2.4.1, it has the same support as  $\alpha$ .  $\square$

It is interesting to consider how the different Wasserstein metrics generate the various  $P_p(X)$  starting from the same basis, i.e., the probability measures with finite support. As  $W_p$  is monotonic in  $p$  when  $p \leq q$ , any sequence Cauchy in the  $W_q$  metric is also Cauchy in the  $W_p$  metric. So the difference must be that sequences of probability measures with finite support can be Cauchy in the  $W_p$  metric, but not in the  $W_q$  metric when  $p < q$ . Examples of this phenomenon can be found by building on Example 2.3:

**Example 2.8.** Set

$$D_{q,m}(n) = \begin{cases} 0 & (n = 0) \\ \zeta(q+1)^{-1} \frac{1}{n^{q+1}} & (1 \leq n \leq m) \\ 1 - \zeta(q+1)^{-1} \sum_{i=1}^m \frac{1}{i^{q+1}} & (n = m+1) \\ 0 & (n > m+1) \end{cases}$$

Note that  $D_{q,m}$  and  $D_q$  agree for  $1 \leq n \leq m$ , but the rest of the mass of  $D_q$  is concentrated on  $D_{q,m}$  at  $m+1$ . Then the sequence  $D_{q,m}$  converges to  $D_q$  in the  $W_p$  metric when  $p < q$ , but is not Cauchy w.r.t.  $W_p$  when  $p = q$ .

We pause our development to examine the topologies induced by the Wasserstein metrics. In the case of separable such spaces it is known that these metrics topologise a suitable notion of weak convergence, see [23, Theorem 7.12]. It turns out that we can again use our techniques and generalise these results to all complete metric spaces.

Given a metric space  $X$  and probability measures  $\mu_i$  ( $i \geq 0$ ) and  $\mu$  on  $X$ , we say that the  $\mu_i$  converge weakly to  $\mu$ , and write  $\mu_i \rightarrow \mu$  if, for all continuous bounded  $f : X \rightarrow \mathbb{R}$ , we have  $\int f d\mu_i \rightarrow \int f d\mu$ ; an equivalent formulation is that  $\underline{\lim}_i \mu_i(U) \geq \mu(U)$  for every open set  $U$  (see [19, Theorem 6]).



**Lemma 2.9.** *Let  $C \subseteq X$  be a closed subset of a metric space and let  $\mu_i$  ( $i \geq 0$ ) and  $\mu$  be probability measures on  $C$ . Then we have:*

$$\mu_i \longrightarrow \mu \quad \Longleftrightarrow \quad i_*(\mu_i) \longrightarrow i_*(\mu)$$

where  $i : C \rightarrow X$  is the inclusion map.

*Proof.* We use the equivalent formulation of weak convergence in terms of measures of open sets. In one direction, suppose that  $\mu_i \longrightarrow \mu$ . Then  $i_*(\mu_i) \longrightarrow i_*(\mu)$ , since for every open subset  $U \subseteq X$ , and any probability measure  $\nu$  on  $C$ , we have  $i_*(\nu)(U) = \nu(U \cap C)$ . In the other direction, suppose that  $i_*(\mu_i) \longrightarrow i_*(\mu)$ . Then  $\mu_i \longrightarrow \mu$  since for any open set  $U$  of  $C$  there is an open set  $V$  of  $X$  such that  $U = V \cap C$  and then, for any probability measure  $\nu$  on  $C$ , we have  $\nu(U) = i_*(\nu)(V)$ .  $\square$

**Lemma 2.10.** *Let  $C$  be a separable closed subset of a complete metric space  $X$ , and let  $i : C \rightarrow X$  be the inclusion map. Then, for any probability measures  $\alpha_i$  ( $i \geq 0$ ) and  $\alpha$  on  $C$ , and any  $x_0 \in C$ , the following are equivalent, and hold independently of the choice of  $x_0 \in C$ :*

- (1)  $i_*(\alpha_i) \longrightarrow i_*(\alpha)$  and  $\int d_X(x_0, -)^p di_*(\alpha_i) \longrightarrow \int d_X(x_0, -)^p di_*(\alpha)$ .
- (2)  $\alpha_i \longrightarrow \alpha$  and  $\int d_C(x_0, -)^p d\alpha_i \longrightarrow \int d_C(x_0, -)^p d\alpha$ .
- (3) For all continuous functions  $f : C \rightarrow \mathbb{R}$  such that  $|f(x)| \leq c(1 + d_C(x_0, x)^p)$  for all  $x \in C$ , for some  $c \in \mathbb{R}$ , one has  $\int f d\alpha_i \longrightarrow \int f d\alpha$ .
- (4) For all continuous functions  $f : X \rightarrow \mathbb{R}$  such that  $|f(x)| \leq c(1 + d_X(x_0, x)^p)$  for all  $x \in X$ , for some  $c \in \mathbb{R}$ , one has  $\int f di_*(\alpha_i) \longrightarrow \int f di_*(\alpha)$ .

*Proof.* As  $C$  is separable and complete, by [23, Theorem 7.12], (2) and (3) are equivalent. and hold independently of the choice of  $x_0 \in C$ . As, further, (4) trivially implies (1), it suffices to prove that (1) implies (2) and (3) implies (4), for any  $x_0 \in C$ . The first of these implications follows using Lemma 2.9 and the fact that  $\int d_C(x_0, -)^p d\alpha_i = \int d_X(x_0, -)^p di_*(\alpha_i)$ , for all  $i \geq 0$ , and similarly for  $\alpha$ . For the second of these implications, one again uses Lemma 2.9 and notes that if  $f : X \rightarrow \mathbb{R}$  is a continuous function such that  $|f(x)| \leq c(1 + d_X(x_0, x)^p)$  for all  $x \in X$ , for some  $c \in \mathbb{R}$ , then  $fi : C \rightarrow \mathbb{R}$  is a continuous function such that  $|(fi)(x)| \leq c(1 + d_C(x_0, x)^p)$ , for all  $x \in C$  and that  $\int f di_*(\alpha_i) = \int (fi) d\alpha_i$ , and similarly for  $\alpha$ .  $\square$

**Theorem 2.11.** *Let  $X$  be a complete metric space and choose  $\mu_i$  ( $i \geq 0$ ) and  $\mu$  in  $P_p(X)$ . Then the following two conditions hold independently of the choice of  $x_0 \in X$  and are equivalent:*

- (1)  $\mu_i \longrightarrow \mu$  and  $\int d(x_0, -)^p d\mu_i \longrightarrow \int d(x_0, -)^p d\mu$ ,
- (2) For all continuous functions  $f : X \rightarrow \mathbb{R}$  such that  $|f(x)| \leq c(1 + d(x_0, x)^p)$  for all  $x \in X$ , for some  $c \in \mathbb{R}$ , one has  $\int f d\mu_i \longrightarrow \int f d\mu$ .

*Proof.* First, choose  $a, b \in X$ . Next, choose a separable closed set  $C \subseteq X$  containing  $a$  and  $b$  and including the supports of the  $\mu_i$  ( $i \geq 0$ ) and  $\mu$ . Then, taking  $\alpha_i = r(\mu_i)$  ( $i \geq 0$ ) and  $\alpha = r(\mu)$  in Lemma 2.10, we see that (1) and (2) hold independently of the choice of  $x_0 \in C$  (e.g., whether choosing  $a$  or  $b$ ) and are equivalent for any such choice as  $\mu_i = i_*(\alpha_i)$ , for all  $i \geq 0$ , and  $\mu = i_*(\alpha)$ . As  $a$  and  $b$  were chosen arbitrarily from  $X$ , the conclusion follows.  $\square$

If either of the two equivalent conditions of Theorem 2.11 hold for  $\mu_i$  ( $i \geq 0$ ) and  $\mu$  in  $P_p(X)$  and any choice of  $x_0 \in X$ , where  $X$  is a complete metric space, we say that the  $\mu_i$  converge  $p$ -weakly to  $\mu$ , and write  $\mu_i \longrightarrow_p \mu$ .

**Theorem 2.12.** *For any complete metric space  $X$ , the Wasserstein distance  $W_p$  metricises  $p$ -weak convergence in  $P_p(X)$ .*

*Proof.* Choose  $\mu_i$  ( $i \geq 0$ ) and  $\mu$  in  $P_p(X)$  to show that  $\mu_i \rightarrow_p \mu$  iff  $\mu_i$  converges to  $\mu$  w.r.t. the Wasserstein distance  $W_p$  in  $X$ . To this end, let  $C \subseteq X$  be a closed separable set containing the supports of the  $\mu_i$  and  $\mu$  (and so necessarily nonempty), and set  $\alpha_i = r(\mu_i)$  and  $\alpha = r(\mu)$  and let  $x_0$  be an element of  $C$ . By [24, Theorem 6.9] we have that  $\alpha_i \rightarrow_p \alpha$  iff  $\alpha_i$  converges to  $\alpha$  w.r.t. the Wasserstein distance  $W_p$  in  $C$ . So it suffices to show (i) that  $\mu_i \rightarrow_p \mu$  iff  $\alpha_i \rightarrow_p \alpha$ , and (ii) that  $\mu_i$  converges to  $\mu$  w.r.t. the Wasserstein distance  $W_p$  in  $X$  iff  $\alpha_i$  converges to  $\alpha$  w.r.t. the Wasserstein distance  $W_p$  in  $C$ .

That (i) holds follows from the definition of  $p$ -weak convergence and the equivalence between parts (1) and (2) of Lemma 2.10 applied to the  $\alpha_i$ ,  $\alpha$ , and  $x_0$ . That (ii) holds follows from the fact that, by Lemma 2.5.1, the inclusion of  $C$  in  $X$  is an isometric embedding.  $\square$

### 3. WASSERSTEIN ALGEBRAS

We begin with an account of barycentric algebras. We then move on to quantitative algebras, by which we mean algebras over metric spaces. After some general considerations on these we move on to Wasserstein algebras of order  $p$ , and our main theorem, characterising the Wasserstein monads  $P_p$ .

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:

$$x +_1 y = x \tag{B1}$$

$$x +_r x = x \tag{B2}$$

$$x +_r y = y +_{1-r} x \tag{SC}$$

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

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  $n > 2$ ,  $r_1 \neq 1$ , with the other cases being evident. Note that affine maps preserve such finite convex sums. Finite convex sums have been axiomatised as *convex spaces* (or *convex algebras*) where the sums are required to obey the *projection* axioms:

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

where  $\delta_{ik}$  is the Kronecker symbol, and the *barycentre* axioms:

$$\sum_{i=1}^n r_i \left( \sum_{k=1}^m s_{ik} x_k \right) = \sum_{k=1}^m \left( \sum_{i=1}^n r_i s_{ik} \right) x_k$$

With the above definition of finite sums, barycentric algebras form convex spaces, and indeed the two categories of algebras are equivalent under this correspondence.

Probability measures on a measurable space, and subclasses of them, naturally form barycentric algebras under the standard binary pointwise convex combination operations:

$$(\mu +_r \nu)(B) = r\mu(B) + (1 - r)\nu(B)$$

In particular, for any set  $X$ , the barycentric algebra  $P_f(X)$  of probability measures over  $X$  with finite support is the free barycentric algebra over  $X$ , with universal arrow the Dirac delta function (see [18, 20]); if  $f : X \rightarrow B$  is a map to a barycentric algebra  $B$ , then the unique extension of  $f : X \rightarrow B$  along  $\delta$  is given by:

$$\bar{f}(\mu) = \sum_{s \in \text{supp}(\mu)} \mu(s)f(s)$$

Barycentric algebras originate with the work of M.H. Stone in [21], and convex spaces with that of T. Świrszcz in [22]. For further bibliographic references and historical discussion, see, e.g., [14].

Turning to quantitative algebras, 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*  $\Sigma$  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  $\Sigma$ -algebra*  $(X, f_X (f \in \Sigma))$  is then a metric space  $X$  equipped with a nonexpansive function  $f_X : X^n \rightarrow X$  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*  $h : X \rightarrow Y$  of  $\Sigma$ -algebras is a nonexpansive map  $h : X \rightarrow Y$  such that for all  $f : n$  and  $x_1, \dots, x_n \in X$  we have:

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

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

For any  $p \geq 1$ , a *Wasserstein (barycentric) algebra of order  $p$*  is a quantitative algebra  $(X, +_r : X^2 \rightarrow X (r \in [0, 1]))$  forming a barycentric algebra such that for all  $x, x', y, y' \in X$  we have:

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

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:

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

We gather some other remarks about convex combinations. Recollect that a function  $f : X \rightarrow Y$  between metric spaces is  $\alpha$ -Hölder continuous if, for all  $x, y \in X$ , we have  $d(fx, fy) \leq Md(x, y)^\alpha$  for some constant  $M$ .

**Lemma 3.1.** *Let  $(X, +_r (r \in [0, 1]))$  be a Wasserstein algebra of order  $p$ . Then:*

- (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  $r$ ,  $x+_ry$  is  $1/p$ -Hölder continuous.
- (3) The following generalisation of equation (\*) to finite convex sums holds:

$$d(\sum_{i=1}^n r_i x_i, \sum_{i=1}^n r_i x'_i)^p \leq \sum_{i=1}^n r_i d(x_i, x'_i)^p$$

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

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

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

$$x+_ry = x+_e(x+_s y) \quad \text{and} \quad x+_sy = y+_e(x+_s y)$$

So by part 1 we have

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

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

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

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

- (3) This is a straightforward induction. □

For any metric space  $X$ , we turn  $P_p(X)$  into a barycentric algebra by equipping it with the standard convex combination operations. It is evident that  $\mu+_r\nu$  has finite moment of order  $p$  if  $\mu$  and  $\nu$  do.

**Lemma 3.2.** *For any metric space  $X$ ,  $P_p(X)$  forms a Wasserstein algebra of order  $p$ .*

*Proof.* We need only show that for any  $\mu, \mu', \nu, \nu' \in P_p(X)$  and  $r \in [0, 1]$  we have:

$$W_p(\mu+_r\nu, \mu'+_r\nu')^p \leq r W_p(\mu, \mu')^p + (1-r) W_p(\nu, \nu')^p$$

(recalling that the Wasserstein condition implies nonexpansiveness of the operations). To this end, choose  $\mu, \mu', \nu, \nu' \in P_p(X)$ . Let  $\alpha$  be a coupling between  $\mu$  and  $\mu'$  and let  $\beta$  be a coupling between  $\nu$  and  $\nu'$ . Then  $\alpha+_r\beta$  is a coupling between  $\mu+_r\nu$  and  $\mu'+_r\nu'$ , and we have:

$$\int d_X^p d(\alpha+_r\beta) = \int d_X^p d\alpha +_r \int d_X^p d\beta$$

The conclusion then follows, as we have:

$$\begin{aligned} W_p(\mu+_r\nu, \mu'+_r\nu')^p &= \inf_{\gamma} \left( \int d_X^p d\gamma \right) \\ &\leq \inf_{\alpha, \beta} \left( \int d_X^p d(\alpha+_r\beta) \right) \\ &= \inf_{\alpha, \beta} \left( \int d_X^p d\alpha +_r \int d_X^p d\beta \right) \\ &= \inf_{\alpha} \left( \int d_X^p d\alpha \right) +_r \inf_{\beta} \left( \int d_X^p d\beta \right) \\ &= W_p(\mu, \mu')^p +_r W_p(\nu, \nu')^p \end{aligned} \quad \square$$

We next turn to characterising the free Wasserstein algebras of order  $p$ . For any metric space  $X$ , let  $F_p(X)$  be the sub-Wasserstein algebra of order  $p$  of  $P_p$  that consists of the finitely supported probability measures.

**Theorem 3.3.** *For any metric space  $X$ ,  $F_p(X)$  is the free Wasserstein algebra of order  $p$  over  $X$ , with universal arrow the Dirac delta function  $\delta : X \rightarrow F_p(X)$ .*

*Proof.* First,  $\delta : X \rightarrow F_p(X)$  is an isometric embedding, since, 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  $f : X \rightarrow W$  be any nonexpansive affine map to a Wasserstein algebra  $B$  of order  $p$ . As a set,  $F_p(X)$  is the free barycentric algebra over  $X$  with unit the Dirac function, and the unique affine map  $\bar{f} : F_p(X) \rightarrow W$  extending  $f$  along the unit is given by the formula:

$$\bar{f}(\mu) = \sum_{s \in \text{supp}(\mu)} \mu(s) f(s)$$

So it only remains to show that  $\bar{f}$  is nonexpansive. Let  $\mu, \nu$  be finitely supported probability measures with respective finite non-empty supports  $S$  and  $T$ . Let  $\gamma$  be a coupling between them. Then  $\gamma$  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  $\mu$  and  $\nu$  are the marginals of  $\gamma$ , 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:

$$\begin{aligned} d(\bar{f}(\mu), \bar{f}(\nu)) &= d(\bar{f}(\sum_{s \in S, t \in T} \gamma(\langle s, t \rangle) \delta(s)), \bar{f}(\sum_{s \in S, t \in T} \gamma(\langle s, t \rangle) \delta(t))) \\ &= d(\sum_{s \in S, t \in T} \gamma(\langle s, t \rangle) \bar{f}(\delta(s)), \sum_{s \in S, t \in T} \gamma(\langle s, t \rangle) \bar{f}(\delta(t))) \\ &= d(\sum_{s \in S, t \in T} \gamma(\langle s, t \rangle) f(s), (\sum_{s \in S, t \in T} \gamma(\langle s, t \rangle) f(t)) \\ &\quad \text{(as } f = \bar{f}\delta) \\ &\leq \left( \sum_{s \in S, t \in T} \gamma(\langle s, t \rangle) d(f(s), f(t))^p \right)^{1/p} \\ &\quad \text{(by Lemma 3.1.3)} \\ &\leq \left( \sum_{s \in S, t \in T} \gamma(\langle s, t \rangle) d(s, t)^p \right)^{1/p} \\ &\quad \text{(as } f \text{ is nonexpansive)} \\ &= \left( \int d_X^p d\gamma \right)^{1/p} \end{aligned}$$

As  $\gamma$  was an arbitrary coupling between  $\mu$  and  $\nu$  we therefore have

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

as required. □

To characterise the free complete Wasserstein algebras of order  $p$ , we need two lemmas, one on the completion of metric spaces, and the other on the completion of quantitative  $\Sigma$ -algebras. The usual completion  $\bar{X}$  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  $c_X : X \rightarrow \bar{X}$ , where  $c_X(x) = [\langle x \rangle_{n \in \mathbb{N}}]$ . That is, for any complete metric space  $Y$  and nonexpansive function  $f : X \rightarrow Y$  there is a unique nonexpansive function  $\bar{f} : \bar{X} \rightarrow Y$  extending  $f$  along  $c_X$ . This function is given by the formula:

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

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

**Lemma 3.4.** *Let  $\theta : X \rightarrow Y$  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  $\theta$ .*

*Proof.* We can assume without loss of generality that  $\theta$  is an inclusion mapping. We show that  $\bar{\theta}$  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

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

as required.

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

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

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

The criterion of Lemma 3.4 extends to  $\Sigma$ -algebras:

**Lemma 3.5.** *Let  $X, Y$  be quantitative  $\Sigma$ -algebras and let  $\theta : X \rightarrow Y$  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  $\Sigma$ -algebra over  $X$ , with universal arrow  $\theta$ .*

*Proof.* We can assume without loss of generality that  $\theta$  is an inclusion mapping. Let  $h : X \rightarrow Z$  be any nonexpansive homomorphism from  $X$  to a complete quantitative algebra  $Z$ . By Lemma 1,  $h$  extends to a unique nonexpansive map  $\bar{h} : Y \rightarrow Z$ , 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:

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

$\square$

We can now prove our main theorem on free complete Wasserstein algebras of order  $p$ :

**Theorem 3.6.** *For any complete metric space  $X$ ,  $P_p(X)$  is the free complete Wasserstein algebra of order  $p$  over  $X$ , with universal arrow the Dirac delta function.*

*Proof.* Theorem 2.7 and Lemma 3.2 tell us that  $P_p(X)$  is a complete Wasserstein algebra of order  $p$ . It therefore remains to prove it is the free one over  $X$ , with unit the Dirac delta function.

First, by Theorem 3.3,  $F_p(X)$  is the free Wasserstein algebra of order  $p$  over  $X$ , with universal arrow the Dirac function.

Second,  $P_p(X)$  is the free complete Wasserstein algebra of order  $p$  over  $F_p(X)$ , with universal arrow the inclusion. To see this, first note that  $F_p(X)$  is a sub-quantitative algebra of  $P_p(X)$  that, by Theorem 2.7, generates  $P_p(X)$ . So, by Lemma 3.5,  $P_p(X)$  is the free complete quantitative algebra with signature the  $+_r$  over  $F_p(X)$ , with universal arrow the inclusion. It is therefore, in particular, the free complete Wasserstein algebra of order  $p$  over  $F_p(X)$ , with universal arrow the inclusion.

Combining these two free algebra assertions, we see that, as required,  $P_p(X)$  is the free complete Wasserstein algebra of order  $p$  over  $X$ , with universal arrow the Dirac delta function.  $\square$

There is a disconnect between the present work and that in [17] 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  $d : X^2 \rightarrow \overline{\mathbb{R}}_+$  to the extended reals, defined in the usual way, but 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 [3] 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  $c_X : X \rightarrow \overline{X}$  so obtained is an isometry and is universal.

The analogue of Lemma 3.4 then goes through, with the same proof, as does that of Lemma 3.5, with the evident definitions of quantitative  $\Sigma$ -algebras over extended metric spaces and their homomorphisms. Next, the Wasserstein algebras of order  $p$  over an extended metric space are also defined as before, and the extension of Lemma 3.2 to extended metric spaces then goes through, with the same proofs, as do the following extensions of Theorems 3.3 and 3.6:

**Theorem 3.7.** *For any metric space  $X$ ,  $F_p(X)$  is the free extended metric Wasserstein algebra of order  $p$  over  $X$ , with universal arrow the Dirac delta function  $\delta : X \rightarrow F_p(X)$ .*

**Theorem 3.8.** *For any complete metric space  $X$ ,  $P_p(X)$  is the free complete extended metric Wasserstein algebra of order  $p$  over  $X$ , with universal arrow the Dirac delta function.*

We remark that in [17] the Wasserstein algebras of order  $p$  are axiomatised using quantitative equational logic. This logic makes use of quantitative equations, which have the form  $t =_r u$ , where  $t$  and  $u$  are terms built as usual from a finitary signature, and  $r \in [0, 1]$ . If  $t$  and  $u$  denote elements  $a$  and  $b$  of an extended metric space  $X$ , then  $t =_r u$  holds if, and only if,  $d_X(a, b) \leq r$ . The following axiom scheme of quantitative equational logic axiomatises the Wasserstein algebras of order  $p$ :

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

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

Our discussion has been in terms of barycentric algebras, but, as common in algebra, other axiomatisations are possible. As an immediate example, noting that the categories of barycentric algebras and convex spaces are equivalent, let us define Wasserstein convex spaces of order  $p$  to be quantitative convex spaces which obey the natural Wasserstein

condition of order  $p$  for convex spaces, viz that, for all  $n \geq 1$  we have:

$$d(\sum_{i=1}^n r_i x_i, \sum_{i=1}^n r_i x'_i)^p \leq \sum_{i=1}^n r_i d(x_i, x'_i)^p$$

Then the categories of Wasserstein barycentric algebras of order of  $p$  and Wasserstein convex spaces of order of  $p$  are equivalent, as are their subcategories of algebras over complete metric spaces. One then immediately obtains from the above work characterising free Wasserstein barycentric algebras of order of  $p$  corresponding results characterising free Wasserstein convex spaces of order of  $p$ .

The situation is more complex if we consider instead midpoint (or mean-value or medial mean) algebras, see, e.g., [15, 9]. These algebras are the commutative, idempotent, medial groupoids. That is, they have a binary operation, which we write as  $\oplus$ , subject to the following laws:

$$x \oplus y = y \oplus x \tag{C}$$

$$x \oplus x = x \tag{I}$$

$$(x \oplus u) \oplus (v \oplus z) = (x \oplus v) \oplus (u \oplus z) \tag{M}$$

One can define binary convex combinations  $x +_r y$  in these algebras where  $r \leq 1$  is a dyadic fraction (i.e., one of the form  $\frac{m}{2^n}$ ), and they obey analogues of the barycentric axioms. The free such algebra over a set consists of all finite probability distributions with dyadic weights, see [11, Theorems 2.14 and 2.15].

Every barycentric algebra evidently yields a midpoint algebra by setting  $x \oplus y = x +_{0.5} y$ , and we then have a full and faithful functor to the category of midpoint algebras, which acts as the identity on morphisms; however it is not an equivalence of categories, as not every midpoint algebra can be extended to a barycentric algebra with the same carrier <sup>1</sup>.

The situation changes when we move to quantitative algebras. Define Wasserstein midpoint algebras of order  $p$  to be quantitative midpoint algebras which satisfy the condition

$$d(x \oplus y, x' \oplus y')^p \leq d(x, x')^p \oplus d(y, y')^p$$

Then, following the argument of Theorem 3.3, we see that, for any metric space  $X$ ,  $F_p(X)$ , with its natural midpoint algebra structure, is the free Wasserstein midpoint algebra of order  $p$  over  $X$ . Next, note that, for complete  $X$ ,  $P_p(X)$  is even generated by the finite probability distributions with dyadic fractional weights, as, by Lemma,  $x +_r y$  is Hölder continuous in  $r$ , and so, if a sequence  $r_n$  in  $[0, 1]$  converges to  $r$ , then the sequence  $x +_{r_n} y$  is Cauchy and converges to  $x +_r y$ . With that, now following the argument of Theorem 3.6, we see that  $P_p(X)$  forms the free complete Wasserstein midpoint algebra of order  $p$  over  $X$ .

The construction of midpoint algebras from barycentric algebras yields a full and faithful functor from the category of Wasserstein barycentric algebras of order  $p$  over complete metric spaces to Wasserstein midpoint algebras of order  $p$  over complete metric spaces. We now sketch a proof that this is in fact an equivalence of categories. It suffices to show that every complete Wasserstein midpoint algebra of order  $p$  can be extended to a complete Wasserstein barycentric algebra of order  $p$  with the same carrier. Let  $A$  be such a midpoint algebra.

---

<sup>1</sup>The free midpoint algebra over two generators ( $a$  and  $b$ , say) cannot be so extended. It consists of all formal dyadic fractional convex combinations of  $a$  and  $b$ . If, for the sake of contradiction, there was such an extension then we would have  $a +_r b = a +_s b$  for some  $r$  and  $s$  with  $r$  not a dyadic fraction and  $s$  a dyadic fraction. But, according to a lemma of Neumann (see [18, Lemma 4]), if the equation  $a +_r b = a +_s b$  holds for two different  $r, s \in (0, 1)$  in a barycentric algebra, then it holds for all different  $r, s \in (0, 1)$  in that algebra. So, in our case, taking  $r, s \in (0, 1)$  to be two different dyadic fractions we obtain the desired contradiction.



There is a unique nonexpansive morphism  $h : P_p(A) \rightarrow A$  of Wasserstein midpoint algebras of order  $p$  extending the identity on  $A$  along the unit. Using this, one inherits the Hölder continuity of  $+_r$  on  $A$ , for dyadic  $r$  from that on  $P_p(A)$ . Then one can define  $x +_r y \in A$  for  $x, y \in A$  as the limit of the sequence  $x +_{r_n} y$  where  $r_n$  is a chosen sequence of dyadic fractions converging to  $r$ , noting that the sequence  $x +_{r_n} y$  is Cauchy by Hölder continuity.

## REFERENCES

- [1] G. Basso, A Hitchhiker's guide to Wasserstein distances, available at: <http://n.ethz.ch/~gbasso/>, 2015.
- [2] F. van Breugel, The metric monad for probabilistic nondeterminism, Unpublished note, available at <http://www.cse.yorku.ca/~franck/research/drafts/monad.pdf>, 2005.
- [3] D. Burago, Y.D. Burago, and S. Ivanov, *A course in metric geometry*, Vol. 33, American Mathematical Society, 2001.
- [4] P. Clement and W. Desch, An elementary proof of the triangle inequality for the Wasserstein metric, *Proceedings of the American Mathematical Society*, **136**(1), 333–339, 2008.
- [5] D.E. Edwards, A simple proof in Monge-Kantorovich duality theory, *Studia Math.*, **200**(1), 67–77, 2010.
- [6] T. Fritz and P. Perrone, A Probability Monad as the Colimit of Finite Powers, arXiv: 1712.05363, 2017.
- [7] D.H. Fremlin, *Measure Theory, Volume 4: Topological Measure Spaces*, Torres Fremlin, 2006.
- [8] M. Giry, A categorical approach to probability theory, *Categorical aspects of topology and analysis*, Lecture Notes in Mathematics, Vol. 915, 68–85, Springer, 1982.
- [9] R. Heckmann, Probabilistic domains, *Proc. 19th. International Colloquium on Trees in Algebra and Programming*, Lecture Notes in Computer Science, Vol. 787, 142–156, Springer, 1994.
- [10] C. Heunen, O. Kammar, S. Staton, and H. Yang, A convenient category for higher-order probability theory, *Proc. 32nd. Annual Symposium on Logic in Computer Science*, 1–12, IEEE, 2017.
- [11] J. Ježek and T. Kepka, Free commutative idempotent abelian groupoids and quasigroups, *Acta Universitatis Carolinae, Mathematica et Physica*, **17**(2), 13–19, 1976.
- [12] C. Jones, *Probabilistic non-determinism*, Ph.D. Thesis, Department of Computer Science, The University of Edinburgh, 197 pages, available at <https://www.era.lib.ed.ac.uk/handle/1842/413>, 1990.
- [13] C. Jones and G.D. Plotkin, A probabilistic powerdomain of evaluations, *Proc. 4th. Annual Symposium on Logic in Computer Science*, pp. 186–195, IEEE, 1989.
- [14] K. Keimel and G.D. Plotkin, Mixed powerdomains for probability and nondeterminism, *Logical Methods in Computer Science*, **13**(1:2), 1–84, 2017.
- [15] S. Kermit, Cancellative medial means are arithmetic, *Duke Math J.*, **37**, 439–445, 1970.
- [16] F.W. Lawvere, The category of probabilistic mappings, Preprint, available at <https://ncatlab.org/nlab/files/lawvereprobability1962.pdf>, 1962.
- [17] R. Mardare, P. Panangaden, and G.D. Plotkin, Quantitative Algebraic Reasoning, *Proc. 31st Annual Symposium on Logic in Computer Science*, 700–709, 2016.
- [18] W.D. Neumann, On the quasivariety of convex subsets of affine spaces, *Archiv der Mathematik*, 21:11–16, 1970.
- [19] K.R. Parthasarathy, *Probability measures on metric spaces*, Academic Press, 1967.
- [20] A.B. Romanowska and J.D.H. Smith, *Modes*, World Scientific, Singapore, 2002.
- [21] M.H. Stone, Postulates for the barycentric calculus, *Annali di Matematica Pura ed Applicada*, Serie IV, **29**, 25–30, 1942.
- [22] T. Świrszcz, Monadic functors and convexity, *emphBulletin de l'Académie Polonaise des Sciences, Ser. Sci. Math. Astronom. Phys.*, **22**, 39–42, 1974.
- [23] C. Villani, *Topics in optimal transportation*, Vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, 2003.
- [24] C. Villani, *Optimal transport: old and new*, Vol. 338, Springer Science & Business Media, 2008.