

A PROBABILITY MONAD AS THE COLIMIT OF SPACES OF FINITE SAMPLES

TOBIAS FRITZ AND PAOLO PERRONE

ABSTRACT. Abstract We define and study a probability monad on the category of complete metric spaces and short maps. It assigns to each space the space of Radon probability measures on it with finite first moment, equipped with the Kantorovich–Wasserstein distance. This monad is analogous to the Giry monad on the category of Polish spaces, and it extends a construction due to van Breugel for compact and for 1-bounded complete metric spaces.

We prove that this *Kantorovich monad* arises from a colimit construction on finite power-like constructions, which formalizes the intuition that probability measures are limits of finite samples. The proof relies on a criterion for when an ordinary left Kan extension of lax monoidal functors is a monoidal Kan extension. The colimit characterization allows the development of integration theory and the treatment of measures on spaces of measures, without measure theory.

We also show that the category of algebras of the Kantorovich monad is equivalent to the category of closed convex subsets of Banach spaces with short affine maps as morphisms.

Contents

1	Introduction	171
2	Wasserstein spaces	173
3	The Wasserstein space as a colimit	179
4	Monad structure	189
5	Algebras of the Kantorovich monad	200
A	Colimits of (complete) metric spaces	206
B	Kan extensions of lax monoidal functors	208
C	Convex combinations of metric spaces	212
	Bibliography	218

We thank Rory Lucyshyn-Wright, Rostislav Matveev, Prakash Panangaden, Sharwin Rezagholi, Patrick Schultz and David Spivak for helpful discussions. We also thank Prakash Panangaden and David Spivak for helpful comments on a draft of this work, and Roald Koudenburg and Mark Weber for pointers to the literature. Moreover, we thank the anonymous reviewer for very helpful comments.

Received by the editors 2017-12-22 and, in final form, 2019-03-07.

Transmitted by Anders Kock. Published on 2019-03-11.

2010 Mathematics Subject Classification: 60A05, 18C15, 52A01.

Key words and phrases: Categorical probability, Giry monad, graded monad, optimal transport, Wasserstein spaces, Kantorovich-Rubinstein distance, monoidal Kan extension.

© Tobias Fritz and Paolo Perrone, 2019. Permission to copy for private use granted.

1. Introduction

In existing categorical approaches to probability theory, one works with a suitable category of measurable spaces and equips it with a monad, which associates to every space X the space of probability measures on X [Law62]. This applies e.g. to the Giry monad [Gir82] and to the Radon monad, discovered in [Sem73, Świ74] and named such in [FJ15]; see also [Jac17] for a general overview of probability monads and [LW17] for a more general setup. These monads constitute an additional piece of structure on the underlying category. Here, we introduce another such monad—the *Kantorovich monad*—which lives on the category of complete metric spaces and extends the analogous monads studied by van Breugel [vB05] on the full subcategories of compact metric spaces and of 1-bounded complete metric spaces. An extension to suitable non-bounded spaces is necessary for our goal to redevelop basic probability theory categorically, because generic distributions of random variables in probability theory may not have bounded support – the Gaussian is a prominent example.

We prove that the monad structure of the Kantorovich monad naturally arises from a colimit construction on the underlying category, which is motivated by the operational interpretation of a probability measure as a formal limit of finite samples. This allows to approach some elements of probability measure theory, such as integration, in terms of simpler considerations based on finite sets. Among other benefits, we hope that this may help to make probability theory more constructive, perhaps in a way that allows for straightforward implementation in a functional programming language or proof assistant.

Besides this colimit characterization, another reason for using probability monads on *metric* spaces is the following. Most results in probability theory are concerned with approximations (in some sense or another), often in a quantitative manner. Therefore we expect that working with metric spaces will allow us to find categorical formulations, proofs, or perhaps even *generalizations* of such approximation results, such as the law of large numbers, or the Glivenko-Cantelli theorem on the convergence of empirical distributions.

In algebra, theoretical computer science, and other fields, monads often arise from equational theories [PP02]. Categorically, this is formalized by presenting a monad in terms of an associated Lawvere theory, operad, or generalized operad, via a suitable coend or more general colimit. From this perspective, our colimit characterization formalizes the idea that this probability monad models a kind of algebraic theory presented by the operations of taking convex combinations with uniform weights. However, our way of presenting the Kantorovich monad does not involve a Lawvere theory or an operad, but rather a *graded monad* [FKM16].

This theme has also been pursued in the work of van Breugel on the Kantorovich monad for 1-bounded complete metric spaces [vB05], in particular with the consideration of metric mean-value algebras [vBHMW05, Definition 6]. A similar idea underlies recent ongoing work of Mardare, Panangaden and Plotkin. In [MPP16, Theorem 10.9]¹, they

¹ However, in order for their Theorem 10.9 to be correct, their definition of the p -Wasserstein space

consider the underlying functor of the Kantorovich monad on complete separable metric spaces as the free algebras of an \mathbb{R}_+ -enriched Lawvere theory, which is closely related to our Theorem 5.4(d).

Summary. In Section 2 we introduce the main mathematical constructions that we use in this work: the categories \mathbf{Met} and \mathbf{CMet} of (complete) metric spaces and short maps, and the Radon measures on such spaces with finite first moment. We prove (Theorem 2.6) that such measures are equivalently linear, positive and τ -smooth functionals on the space of Lipschitz functions. In Section 2.8 we introduce the Wasserstein metric, and show the functoriality of the Wasserstein space construction (Lemma 2.13), resulting in the *Kantorovich functor* P .

In Section 3 we prove (Theorem 3.21 and Corollary 3.22) that the Wasserstein spaces and the Kantorovich functor can be obtained as the colimit of the *power functors*, defined in 3.1, and that the universal arrow is given by the empirical distribution map, which we define in 3.9.

In Section 4 we prove that P has a monad structure (Theorem 4.13), which arises naturally from the colimit characterization, given the particular monoidal structure of the power functors (Theorems 4.2 and 4.5). This can be interpreted as a Kan extension in the 2-category \mathbf{MonCat} of monoidal categories and lax monoidal functors (Theorem 4.6).

In Section 5 we study the algebras of P . We show (Theorem 5.4) that the algebras are convex metric spaces whose convex structure is compatible with the metric. This implies in turn that the algebras are equivalently closed convex subsets of Banach spaces (Theorem 5.6).

In Appendix A we study colimits in the category of metric spaces, and prove (Proposition A.4) that the tensor product preserves colimits. This is used in Section 4 to define the monad structure of P .

In Appendix B we give the 2-categorical details (Theorem B.1) of why the Kan extensions used in this paper are Kan extensions in \mathbf{MonCat} , or *algebraic Kan extensions*.

In Appendix C we define the operad of convex spaces, and show that \mathbf{Met} is a pseudoalgebra for this operad, giving further motivation for the power functor construction. We also show that, in agreement with the *microcosm principle*, P -algebras in the form of convex spaces with metric compatibility are particular internal algebras in \mathbf{Met} , i.e. they form a full subcategory.

$\Delta[M]$ needs to be restricted to the measures of finite p -th moment, as we do in Section 2.3 for $p = 1$. The reason is that a probability measure of infinite p -th moment has infinite p -Wasserstein distance from any finitely supported probability measure, e.g. because it has infinite distance from any Dirac measure. The error in the proof (as pointed out to us by Prakash Panangaden) is in the claim that the p -Wasserstein distance metrizes weak convergence, which is true only for finite p -th moment.

2. Wasserstein spaces

2.1. CATEGORICAL SETTING. Two categories are of primary interest to us. The first one is the monoidal category \mathbf{Met} , where:

- Objects are metric spaces, with the classical notion of metric as a distance function $d : X \times X \rightarrow [0, \infty)$ satisfying identity of indiscernibles, symmetry, and the triangle inequality;
- Morphisms are short maps (also called 1-Lipschitz maps), i.e. functions $f : X \rightarrow Y$ such that for all $x, x' \in X$:

$$d_Y(f(x), f(x')) \leq d_X(x, x') ; \tag{2.1}$$

- As monoidal structure, we define $X \otimes Y$ to be the set $X \times Y$, equipped with the ℓ^1 -product metric:

$$d_{X \otimes Y}((x, y), (x', y')) := d_X(x, x') + d_Y(y, y'). \tag{2.2}$$

The second one is its full subcategory \mathbf{CMet} whose objects are complete metric spaces.

It is useful to think of metric spaces and short maps as enriched categories and enriched functors [Law73, Law86], although in that context one typically works with a more relaxed notion of metric space, such as allowing infinite distances (or even Lawvere metric spaces), in order to guarantee cocompleteness. For \mathbf{Met} and \mathbf{CMet} , we investigate the existence of particular colimits and their preservation by the monoidal product in Appendix A.

In our considerations, the concept of *isometric embedding* will often come up. It is worth noting that together with the bijective short maps, the isometric embeddings in \mathbf{Met} form an orthogonal factorization system, which is, in many respects, analogous to the (bo,ff) factorization system on \mathbf{Cat} . Isometric embeddings are very close to being characterized 1-categorically as the extremal monomorphisms: every extremal monomorphism in \mathbf{Met} is an isometric embedding, but only the isometric embeddings of *closed* subspaces are extremal monomorphisms, since the isometric embedding of a subspace into its closure is an epimorphism. Since a subspace of a complete metric space is complete if and only if it is closed, it follows that in \mathbf{CMet} the class of extremal monomorphisms coincides with the isometric embeddings.

2.2. ANALYTIC SETTING. The following definitions of an analytic nature will only be needed in this section and in Section 3, where we prove our colimit characterization; all subsequent developments will use the latter and therefore do not require any measure theory.

Every metric space is a topological space, and so also a measurable space with its σ -algebra. We will always suppose probability measures to be Borel, and Radon, i.e. inner regular with respect to compacts.

For $X \in \mathbf{Met}$, we write $\text{Lip}(X)$ for the space of Lipschitz functions $X \rightarrow \mathbb{R}$, where \mathbb{R} carries its usual Euclidean metric. Every Lipschitz function is a scalar multiple of an

element of $\mathbf{Met}(X, \mathbb{R})$, i.e. of a short map $X \rightarrow \mathbb{R}$. We expect that working with the latter space, or even just with $\mathbf{Met}(X, \mathbb{R}_+)$, would be useful for achieving further abstraction. However, currently we prefer to work with $\mathbf{Lip}(X)$, which has the added benefit of being a vector space.

2.3. FINITE FIRST MOMENTS AND A REPRESENTATION THEOREM. In order to define Wasserstein spaces, we first have to define probability measures of finite first moment, which are those for which every Lipschitz function has an expectation value.

2.4. DEFINITION. *Let $X \in \mathbf{Met}$ and p be a probability measure on X . We say that p has finite first moment if the expected distance between two random points is finite, i.e. if*

$$\int d(x, y) dp(x) dp(y) < +\infty.$$

We have borrowed this elegant formulation from Goubault-Larrecq [GL17, Section 1], who attributes it to Fernique.

2.5. LEMMA. *The following statements are equivalent for a probability measure p on $X \in \mathbf{CMet}$:*

(a) *p has finite first moment.*

(b) *There is $y \in X$ such that the expected distance from y is finite,*

$$\int d(y, x) dp(x) < +\infty.$$

(c) *For all $z \in X$, the expected distance from z is finite,*

$$\int d(z, x) dp(x) < +\infty.$$

(d) *Every $f \in \mathbf{Lip}(X)$ has finite expectation value,*

$$\int f(x) dp(x) < +\infty.$$

PROOF. Since p is a probability measure, we know that X is nonempty and thus we can always choose a point whenever we need one.

- (a) \Rightarrow (b): if the integral of a nonnegative function is finite, then the integrand is finite at (at least) one point.

- (b) \Rightarrow (c): For all $z \in X$, and for y as in (b), we have:

$$\begin{aligned} \int d(z, x) dp(x) &\leq \int (d(z, y) + d(y, x)) dp(x) \\ &= d(z, y) + \int d(y, x) dp(x), \end{aligned}$$

where the first term is finite for every z , and the second term is finite by hypothesis.

- (c) \Rightarrow (d): Since f is integrable if and only if $|f|$ is, it is enough to consider the case $f \geq 0$. Then for an arbitrary $z \in X$,

$$\begin{aligned} \int f(x) dp(x) &= \int (f(x) - f(z) + f(z)) dp(x) \\ &\leq f(z) + \int |f(x) - f(z)| dp(x) \\ &\leq f(z) + L_f \int d(x, z) dp(x) < +\infty, \end{aligned}$$

where L_f is the Lipschitz constant of f , which is a finite number.

- (d) \Rightarrow (a): Since the distance is short in each argument, the function

$$x \mapsto \int_X d(x, y) dp(y)$$

is finite by assumption and automatically short. Therefore its expectation is again finite by hypothesis, which implies the finite first moment condition. ■

From now on, we write PX for the set of probability measures on X with finite first moment. Later, we will equip this set with a metric. As we discuss in more detail below, pushing forward measures along a short map $f : X \rightarrow Y$ defines a function $Pf : PX \rightarrow PY$ which makes P into a functor.

Often measures are specified by how they act on functions by integration, such as in the definition of the Daniell integral or in the Riesz representation theorem. We will now state an analogous result for PX . Concretely, every $p \in PX$ defines a linear functional $\mathbb{E}_p : \text{Lip}(X) \rightarrow \mathbb{R}$ given by mapping every function to its expectation value,

$$f \longmapsto \mathbb{E}_p(f) := \int f(x) dp(x). \tag{2.3}$$

We can thus consider \mathbb{E} as a map $\mathbb{E} : PX \rightarrow \text{Lip}(X)^*$ into the algebraic dual. Each functional \mathbb{E}_p has a number of characteristic properties: it is linear, positive, and satisfies a certain continuity property. To define the latter, we consider $\text{Lip}(X)$ as a partially

ordered vector space with respect to the pointwise ordering. A *monotone net* of functions is a family $(f_\alpha)_{\alpha \in I}$ in $\text{Lip}(X)$ indexed by a directed set I , such that $f_\alpha \leq f_\beta$ if $\alpha \leq \beta$. If the supremum $\sup_\alpha f_\alpha$ exists in $\text{Lip}(X)$, we say that this supremum is *pointwise* if $(\sup_\alpha f_\alpha)(x) = \sup_\alpha f_\alpha(x)$ for every $x \in X$. For example on $X = [0, 1]$, the sequence of functions

$$f_n(x) := \min(nx, 1) \tag{2.4}$$

with Lipschitz constant $n \in \mathbb{N}$ is a monotone sequence in $\text{Lip}([0, 1])$ whose supremum is the constant function 1, but this supremum is not pointwise, since $(\sup_n f_n)(0) = 1$, while $\sup_n f_n(0) = 0$.

The following representation theorem is similar to [Edg98, Theorem 2.4.12] and essentially a special case of [Fre06, Theorem 436H].

2.6. THEOREM. *Let $X \in \text{Met}$. Mapping every probability measure to its expectation value functional, $p \mapsto \mathbb{E}_p$, establishes a bijective correspondence between probability measures on X with finite first moment, and linear functionals $\phi : \text{Lip}(X) \rightarrow \mathbb{R}$ with the following properties:*

- *Positivity: $f \geq 0$ implies $\phi(f) \geq 0$;*
- *τ -smoothness: if $(f_\alpha)_{\alpha \in I}$ is a monotone net in $\text{Lip}(X)$ with pointwise supremum $\sup_\alpha f_\alpha \in \text{Lip}(X)$, then*

$$\phi\left(\sup_\alpha f_\alpha\right) = \sup_\alpha \phi(f_\alpha). \tag{2.5}$$

- *Normalization: $\phi(1) = 1$.*

The concept of τ -smoothness is similar to *Scott continuity* in the context of domain theory and to *normality* in the context of von Neumann algebras, but the important difference is that the preservation of suprema only applies to pointwise suprema: the pointwiseness expresses exactly the condition that integration against delta measures must preserve the supremum. For example, integrating (2.4) against δ_0 does not preserve the supremum.

PROOF. The fact that the map $p \mapsto \mathbb{E}_p$ is surjective onto functionals satisfying the above conditions is an instance of [Fre06, Theorem 436H]. It remains to be shown that the representing measure p is unique. If $\mathbb{E}_p = \mathbb{E}_q$, then by [Fre06, Proposition 416E], it is enough to show that $p(U) = q(U)$ for every open $U \subseteq X$. But now the sequence (f_n) of Lipschitz functions

$$f_n(x) := \min(1, n \cdot d(x, X \setminus U))$$

monotonically converges pointwise to the indicator function of U . Together with Lebesgue's monotone convergence theorem, the equality $\mathbb{E}_p = \mathbb{E}_q$ therefore implies $p(U) = q(U)$, as was to be shown. ■

We collect another property for future use, which relies crucially on the non-negativity of a measure:

2.7. LEMMA. *Let $p \in PX$ and $f : X \rightarrow Y$ continuous such that the pushforward measure f_*p is supported on some subset $Y' \subseteq Y$. Then p is supported on $f^{-1}(Y')$.*

PROOF. For $x \in X \setminus f^{-1}(Y')$, by assumption there is a neighborhood $U \ni f(x)$ to which f_*p assigns zero measure. Therefore $(f_*p)(U) = p(f^{-1}(U)) = 0$, and $f^{-1}(U)$ is a neighborhood of x . ■

2.8. CONSTRUCTION OF THE WASSERSTEIN SPACE.

2.9. DEFINITION. *Let $X \in \text{Met}$. The Wasserstein space PX is the set of probability measures on X with finite first moment, with metric given by the Wasserstein distance, or Kantorovich-Rubinstein distance, or earth mover's distance²:*

$$d_{PX}(p, q) := \inf_{r \in \Gamma(p, q)} \int_{X \times X} d_X(x, y) dr(x, y) \tag{2.6}$$

where $\Gamma(p, q)$ is the set of probability measures on $X \times X$ with marginals p and q , respectively.

In terms of duality, one can also characterize the Wasserstein metric as

$$d_{PX}(p, q) = \sup_{f: X \rightarrow \mathbb{R}} \left| \int f(x) d(p - q)(x) \right| = \sup_{f: X \rightarrow \mathbb{R}} (\mathbb{E}_p[f] - \mathbb{E}_q[f]), \tag{2.7}$$

where the sup is taken over all short maps [Vil09, AGS05], which we think of as well-behaved random variables. This duality formula provides one way to see that d_{PX} is in fact a metric.

A simple special case of the Wasserstein distance is:

2.10. LEMMA. *Let $\delta(x_0)$ be the Dirac measure at some $x_0 \in X$.³ Then*

$$d(\delta(x_0), p) = \int d(x_0, x) dp(x). \tag{2.8}$$

PROOF. The only joint that has $\delta(x_0)$ as its first marginal and p as its second marginal is the product measure $\delta(x_0) \otimes p$. Therefore,

$$\begin{aligned} d(\delta(x_0), p) &= \int_{X \times X} d(y, x) d(\delta(x_0) \otimes p)(x, y) \\ &= \int_{X \times X} d(y, x) d(\delta(x_0))(y) dp(x) \\ &= \int_X d(x_0, x) dp(x). \end{aligned}$$

■

²For the different names, see the bibliographical notes at the end of Chapter 6 in [Vil09].

³The notation $\delta(x_0)$ suggests a map $\delta : X \rightarrow PX$. This is indeed the case, as we will see in 4.7.

2.11. THEOREM. [Bas15, Theorem 1.8] *Let $X \in \mathbf{CMet}$. Then PX is also a complete metric space.*

Moreover, if X is separable (resp. compact), then PX is also separable (resp. compact), as proven for example in [Vil09, Theorem 6.18].

2.12. LEMMA. *If $f : X \rightarrow Y$ is an isometric embedding, then so is $Pf : PX \rightarrow PY$.*

PROOF. This follows from the duality formula (2.7) and the fact that, for $X \subseteq Y$, every 1-Lipschitz function $g : X \rightarrow \mathbb{R}$ can be extended to Y , for example via

$$y \mapsto \sup_{x \in X} (g(x) - d(x, y)).$$

■

We would like the construction $X \mapsto PX$ to be functorial in X , and this indeed turns out to be the case. For $f : X \rightarrow Y$, we define $Pf : PX \rightarrow PY$ to be the map which takes every measure to its pushforward $f_*p \in PY$. In the dual picture, in terms of functionals, f_*p is characterized by the substitution formula: for every $g : Y \rightarrow \mathbb{R}$,

$$\mathbb{E}_{f_*p}(g) = \int_Y g(y) d(f_*p)(y) = \int_X g(f(x)) dp(x) = \mathbb{E}_p(g \circ f). \quad (2.9)$$

While preservation of composition and identities are clear, there are still two small things to check in order to establish functoriality:

2.13. LEMMA. *Let $f : X \rightarrow Y$ be short, and $p \in PX$. Then,*

- (a) *f_*p has finite first moment as well;*
- (b) *$f_* : PX \rightarrow PY$ is short.*

PROOF.

- (a) Let $g : Y \rightarrow \mathbb{R}$ be a Lipschitz map, we have $\mathbb{E}_{f_*p}(g) = \mathbb{E}_p(g \circ f) < \infty$ by (2.9) and by the assumption that p has finite first moment.
- (b)

$$\begin{aligned} d_{PY}(f_*p, f_*q) &= \sup_{g:Y \rightarrow \mathbb{R}} (\mathbb{E}_{f_*p}(g) - \mathbb{E}_{f_*q}(g)) = \sup_{g:Y \rightarrow \mathbb{R}} (\mathbb{E}_p(g \circ f) - \mathbb{E}_q(g \circ f)) \\ &\leq \sup_{h:X \rightarrow \mathbb{R}} (\mathbb{E}_p(h) - \mathbb{E}_q(h)) = d_{PX}(p, q). \end{aligned}$$

■

Thus we have a functor $P : \mathbf{Met} \rightarrow \mathbf{Met}$. By Theorem 2.11, P restricts to an endofunctor of \mathbf{CMet} , which we also denote by P . This is the functor we study in this paper. We call it the *Kantorovich functor*, in accordance with [vB05].

3. The Wasserstein space as a colimit

Thanks to the metric structure, it turns out that for $X \in \mathbf{CMet}$, the Wasserstein space PX also arises as the colimit of a diagram involving certain powers of X . The intuition behind this colimit is very operational and formalizes the idea that a probability measure is an idealized version of a finite ensemble of elements of X , sampled randomly via repeated trials. In the next section, we will exploit this colimit characterization in order to equip P with a monad structure.

3.1. POWER FUNCTORS. For $X \in \mathbf{Met}$ and $n \in \mathbb{N}$, let X^n be the metric space whose underlying set is the cartesian power, as in the case of $X^{\otimes n}$, but whose distances are renormalized,

$$d_{X^n}((x_1, \dots, x_n), (y_1, \dots, y_n)) := \frac{d_X(x_1, y_1) + \dots + d_X(x_n, y_n)}{n}. \tag{3.1}$$

One way to motivate this renormalization is that the diagonal map $X \rightarrow X^{\otimes n}$ is not short⁴, while the diagonal map $X \rightarrow X^n$ is an isometric embedding which we call the *n-copy embedding*. Another motivation is given in Appendix C.1, where we show that \mathbf{Met} is a pseudoalgebra of the simplex operad in such a way that the power X^n is the uniform n -ary “convex combination” of X with itself.

Let X_n be the quotient of X^n under the equivalence relation $(x_1, \dots, x_n) \sim (x_{\sigma(1)}, \dots, x_{\sigma(n)})$ for any permutation $\sigma \in S_n$. The elements of X_n are therefore multisets $\{x_1, \dots, x_n\}$. The quotient metric is explicitly given by

$$d_{X_n}(\{x_1 \dots x_n\}, \{y_1 \dots y_n\}) := \min_{\sigma \in S_n} \frac{1}{n} \sum_{i=1}^n d_X(x_i, y_{\sigma(i)}), \tag{3.2}$$

since this is exactly the minimal distance between the two relevant fibers of the quotient map $q_n : X^n \rightarrow X_n$, and these distances already satisfy the triangle inequality. Due to this formula, the composite $X \rightarrow X^n \rightarrow X_n$ is also an isometric embedding, which we call the *symmetrized n-copy embedding* $\delta_n : X \rightarrow X_n$. It is clear that the assignments $X \mapsto X^n$ and $X \mapsto X_n$ are functorial in $X \in \mathbf{Met}$, so that we have functors $(-)^n : \mathbf{Met} \rightarrow \mathbf{Met}$ and $(-)_n : \mathbf{Met} \rightarrow \mathbf{Met}$. The quotient map is a natural transformation $q_n : (-)^n \Rightarrow (-)_n$.

There is a simple alternative way to write the metric (3.2) that connects to the Wasserstein distance (2.6):

⁴This is related to the fact that the symmetric monoidal category (\mathbf{Met}, \otimes) is semicartesian, but not cartesian.

3.2. LEMMA.

$$d_{X_n}(\{x_i\}, \{y_i\}) = \min_A \frac{1}{n} \sum_{i,j} A_{ij} d(x_i, y_j), \quad (3.3)$$

where A ranges over all bistochastic matrices⁵.

PROOF. The right-hand side of (3.2) is upper bounded by (3.2) since every permutation matrix is bistochastic. Conversely, every bistochastic matrix is a convex combination of permutation matrices: according to the Birkhoff-von Neumann theorem, the bistochastic matrices of fixed dimension form a polytope whose vertices are precisely the permutation matrices. Therefore the linear optimization of (3.3) attains the minimum on a permutation matrix. ■

It is not hard to see that if X is complete, then so is every X^S . Since every X_n is a coequalizer of X^n via (3.2) by the action of a finite group, it follows that X_n is complete as well: if $(x_k)_{k \in \mathbb{N}}$ is a Cauchy sequence in X_n , then we can assume without loss of generality that $d(x_k, x_{k+1}) \leq 2^{-k}$ after passing to a subsequence. Then we can lift every $x_k \in X_n$ to $\hat{x}_k \in X^n$, in such a way that $d(\hat{x}_k, \hat{x}_{k+1}) \leq 2^{-k}$ as well, which implies that $(\hat{x}_k)_{k \in \mathbb{N}}$ is also Cauchy and therefore convergent. It follows that $\lim_{k \rightarrow \infty} x_k = q(\lim_{k \rightarrow \infty} \hat{x}_k)$, so that X_n is complete.

3.3. LEMMA. *If $f : X \rightarrow Y$ is an isometric embedding, then so are $f^n : X^n \rightarrow Y^n$ and $f_n : X_n \rightarrow Y_n$.*

Categorically, it is more natural to consider the powers X^S for nonempty finite sets S , where X^S is the metric space whose elements are functions $x_{(-)} : S \rightarrow X$ equipped with the rescaled ℓ^1 -metric,

$$d_{X^S}(x_{(-)}, y_{(-)}) := \frac{1}{|S|} \sum_{s \in S} d_X(x_s, y_s).$$

The idea is that the points of X^S are finite samples indexed by a set of observations S , and a function $x_{(-)} : S \rightarrow X$ assigns to every observation s its outcome x_s . Then it is natural to define the distance between two finite sets of observations as the average distance between the outcomes.

It is clear that X^S is functorial in X , but how about functoriality in S ? Without the rescaling, we would have functoriality $X^T \rightarrow X^S$ for arbitrary injective $S \rightarrow T$, corresponding to semicartesianness of (\mathbf{Met}, \otimes) . But due to the rescaling by $\frac{1}{|S|}$, the functoriality now is quite different:

3.4. LEMMA. *Whenever $\phi : S \rightarrow T$ has fibers of uniform cardinality, i.e. when the cardinality of $\phi^{-1}(t)$ does not depend on $t \in T$, we have an isometric embedding $- \circ \phi : X^T \rightarrow X^S$.*

We also denote this map $- \circ \phi$ by X^ϕ .

⁵We recall that a bistochastic matrix is a square matrix with non-negative entries, whose row and columns all sum to one.

PROOF. Let $x_{(-)}, y_{(-)} \in X^T$. Then:

$$\begin{aligned}
 d_{X^S}(X^\phi(x_{(-)}), X^\phi(y_{(-)})) &= d_{X^S}((x_{\phi(-)}), (y_{\phi(-)})) \\
 &= \frac{1}{|S|} \sum_{s \in S} d_X(x_{\phi(s)}, y_{\phi(s)}) \\
 &= \frac{1}{|S|} \sum_{t \in T} |\phi^{-1}(t)| d_X(x_t, y_t) \\
 &= \frac{1}{|S|} \frac{|S|}{|T|} \sum_{t \in T} d_X(x_t, y_t) \\
 &= d_{X^T}(x_{(-)}, y_{(-)}).
 \end{aligned}$$

■

3.5. DEFINITION. Let $\mathbf{FinUnif}$ be the monoidal category where:

- Objects are nonempty finite sets;
- Morphisms are functions $\phi : S \rightarrow T$ with fibers of uniform cardinality,

$$|\phi^{-1}(t)| = |S|/|T| \quad \forall t \in T. \tag{3.4}$$

- The monoidal structure is given by the cartesian product⁶.

In particular, $\mathbf{FinUnif}$ contains all bijections between nonempty finite sets, and all its morphisms are surjective maps. If we think of every finite set as carrying the uniform probability measure, then $\mathbf{FinUnif}$ is precisely the subcategory of \mathbf{FinSet} which contains the measure-preserving maps.

In the following, we either use the powers X^S for finite sets $S \in \mathbf{FinUnif}$, or equivalently the X^n . In the latter case, we take the n to be the objects of a skeleton of $\mathbf{FinUnif}$ indexed by positive natural numbers n . By equivalence of categories, the choice between the two approaches is only a matter of notation.

We write $X^{(-)} : \mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{CMet}$ for the *power functor* corresponding to Lemma 3.4.

3.6. DEFINITION. Let \mathbf{N} be the monoidal poset of positive natural numbers $\mathbb{N} \setminus \{0\}$ ordered by reverse divisibility, so that a unique morphism $n \rightarrow m$ exists if and only if $m|n$, and monoidal structure given by multiplication.

\mathbf{N} is the posetification of $\mathbf{FinUnif}$, in the sense that the canonical functor $|-| : \mathbf{FinUnif} \rightarrow \mathbf{N}$ which maps every S to its cardinality is the initial functor from $\mathbf{FinUnif}$ to a poset. Since $|S \times T| = |S| \cdot |T|$, this functor is strict monoidal.

In analogy with the power functor $X^{(-)} : \mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{CMet}$, we can also consider the *symmetrized power functor* $X_{(-)} : \mathbf{N}^{\text{op}} \rightarrow \mathbf{CMet}$ which takes $n \in \mathbf{N}$ to X_n , and the unique

⁶This is not the categorical product. In fact, $\mathbf{FinUnif}$ does not have any nontrivial products, but it is semicartesian monoidal.

morphism $m \rightarrow mn$, or $m|mn$, is mapped to the embedding $X_{m|mn} : X_m \rightarrow X_{mn}$ given by n -fold repetition on multisets,

$$\{x_1, \dots, x_m\} \longmapsto \{x_1, \dots, x_m, \dots, x_1, \dots, x_m\}. \quad (3.5)$$

which is clearly natural in X . One can also consider this map as the bottom arrow of a diagram of the form

$$\begin{array}{ccc} X^T & \xrightarrow{X^\phi} & X^S \\ \downarrow & & \downarrow \\ X_{|T|} & \xrightarrow{X_{|T||S|}} & X_{|S|} \end{array} \quad (3.6)$$

The bottom arrow is determined uniquely by the universal property of the quotient map on the left.

3.7. LEMMA. $X_{m|mn} : X_m \rightarrow X_{mn}$ is an isometric embedding.

PROOF. Let $\{x_i\}, \{y_i\} \in X_m$. Then using Lemma 3.2, we can write

$$d_{X_{mn}}(X_{m|mn}(\{x_i\}), X_{m|mn}(\{y_i\})) = \frac{1}{mn} \min_A \sum_{i,j,\alpha,\beta} A_{(i,\alpha),(j,\beta)} d_X(x_i, y_j),$$

where A ranges over all bistochastic matrices of size $mn \times mn$ with rows and columns indexed by pairs (i, α) with $i = 1, \dots, m$ and $\alpha = 1, \dots, n$. Similarly,

$$d_{X_m}(\{x_i\}, \{y_i\}) = \frac{1}{m} \min_B \sum_{i,j} B_{ij} d_X(x_i, y_j).$$

For given B , one can achieve the same value of the first optimization by putting $A_{ij}^{\alpha\beta} := \frac{1}{n} B_{ij}$ for all values of the indices. Conversely, we can put $B_{ij} := \frac{1}{n} \sum_{\alpha,\beta} A_{(i,\alpha),(j,\beta)}$ in order to achieve the same value in the other direction. \blacksquare

Thus the symmetrized power functor $X_{(-)} : \mathbf{N}^{\text{op}} \rightarrow \mathbf{CMet}$ lands in the subcategory of complete metric spaces and isometric embeddings.

Again we have a quotient map $q_S : X^S \rightarrow X_{|S|}$ given by “forgetting the labeling” of particular outcomes and only remembering the multiset of values of the given function $x_{(-)} : S \rightarrow X$,

$$q_S(x_{(-)}) = \{x_s : s \in S\} \in X_{|S|}. \quad (3.7)$$

It is the universal morphism which coequalizes all automorphisms of X^S of the form X^σ , where σ ranges over all bijections $\sigma : S \rightarrow S$.

In this way, we obtain a natural transformation $q : X^{(-)} \Rightarrow X_{|-|}$ of functors $\mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{CMet}$.

3.8. LEMMA. Via the map q appearing in the formula (3.7), the functor $X_{(-)} : \mathbf{N}^{\text{op}} \rightarrow \mathbf{Met}$ is the left Kan extension of $X^{(-)} : \mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{Met}$ along $|-|^{\text{op}}$. Likewise for \mathbf{CMet} in place of \mathbf{Met} .

PROOF. Again because $\mathbf{CMet} \subseteq \mathbf{Met}$ is reflective, it is enough to prove the claim for \mathbf{Met} . There it follows from the universal property of the quotient map q . In more detail, we have the diagram

$$\begin{array}{ccc} \mathbf{FinUnif}^{\text{op}} & \xrightarrow{X(-)} & \mathbf{Met} \\ & \searrow \text{||-||} & \nearrow X_{(-)} \\ & \mathbf{N}^{\text{op}} & \end{array}$$

Consider now another functor K and a natural transformation α as in

$$\begin{array}{ccc} \mathbf{FinUnif}^{\text{op}} & \xrightarrow{X(-)} & \mathbf{Met} \\ & \searrow \text{||-||} & \nearrow K \\ & \mathbf{N}^{\text{op}} & \end{array}$$

Unraveling the definition, this means that for each $S \in \mathbf{FinUnif}$ we have a map

$$\alpha_S : X^S \rightarrow K(|S|),$$

and we need to find a factorization

$$\begin{array}{ccc} X^S & \xrightarrow{q} & X_{|S|} \\ & \searrow \alpha_S & \downarrow u_{|S|} \\ & & K(|S|) \end{array} \tag{3.8}$$

for some $u : X_{(-)} \Rightarrow K$. By naturality of α with respect to automorphisms $\sigma : S \rightarrow S$, we know that α_S is invariant under precomposing by X^σ . Therefore it factors uniquely through q and this defines $u_{|S|}$, which is enough since $|-|$ is (essentially) bijective on objects. It remains to prove naturality of u , which means that for all $m, n \in \mathbf{N}$, the diagram

$$\begin{array}{ccc} X_m & \xrightarrow{X_{m|mn}} & X_{mn} \\ \downarrow u_m & & \downarrow u_{mn} \\ K(m) & \xrightarrow{K(m|mn)} & K(mn) \end{array}$$

commutes. This follows from the fact that $|-| : \mathbf{FinUnif} \rightarrow \mathbf{N}$ is also full, so that the morphism $X_{m|mn}$ is the image of some morphism in $\mathbf{FinUnif}$, together with naturality of α and the definition (3.8). ■

Equivalently, we could have obtained this result from the usual coend formula for left Kan extensions: although general copowers do not exist in \mathbf{Met} or \mathbf{CMet} , since we require all distances to be finite, they do exist in this particular case since \mathbf{N}^{op} is a poset, so that the coend formula only requires the existence of the trivial copowers by the empty set and by singleton sets.

In conclusion, we also have endofunctors $(-)^S : \mathbf{CMet} \rightarrow \mathbf{CMet}$ and $(-)_n : \mathbf{CMet} \rightarrow \mathbf{CMet}$.

3.9. EMPIRICAL DISTRIBUTION MAPS.

3.10. DEFINITION. Let $X \in \text{Met}$. For $S \in \text{FinUnif}$, the empirical distribution map is the map $\iota^S : X^S \rightarrow PX$ which assigns to each S -indexed family $x_{(-)} \in X^S$ the uniform probability measure,

$$\iota^S(x_{(-)}) := \frac{1}{|S|} \sum_{s \in S} \delta(x_s). \quad (3.9)$$

This map is clearly permutation-invariant, so it uniquely determines a map on symmetrized powers as well:

3.11. DEFINITION. For $n \in \mathbb{N}$, the symmetric empirical distribution map is the map $\iota_n : X_n \rightarrow PX$ given by assigning to each multiset $\{x_1, \dots, x_n\} \in X_n$ the corresponding uniform probability measure,

$$\iota_n(\{x_1 \dots x_n\}) := \frac{\delta(x_1) + \dots + \delta(x_n)}{n}. \quad (3.10)$$

The empirical distribution carries less information than the original sequence. The information lost is precisely the ordering, as the following proposition shows.

3.12. PROPOSITION. $\iota_n : X_n \rightarrow PX$ is an isometric embedding for each X and n .

PROOF. For $\{x_i\}, \{y_i\} \in X_n$, consider the finite set

$$N_{xy} := \{1_x, \dots, n_x\} \amalg \{1_y, \dots, n_y\},$$

and equip it with the pseudometric which makes the canonical map $N_{xy} \rightarrow X$ into an isometric embedding, which means in particular that $d(i_x, j_y) = d_X(x_i, y_j)$. In the commutative square

$$\begin{array}{ccc} N_{xy,n} & \xrightarrow{\iota_n} & PN_{xy} \\ \downarrow & & \downarrow \\ X_n & \xrightarrow{\iota_n} & PX \end{array}$$

both vertical arrows are isometric embeddings by Lemmas 3.3 and 2.12, where both P and the lemmas generalize to pseudometric spaces in the obvious way. It is therefore enough to prove that in PN_{xy} , the distance between the uniform distribution on the points $\{1_x, \dots, n_x\}$ and $\{1_y, \dots, n_y\}$ is equal to the distance between these two sets as elements of $N_{xy,n}$. This is indeed the case, since the latter distance is given by (3.3),

$$d(\{i_x\}, \{j_y\}) = \min_A \frac{1}{n} \sum_{ij} A_{ij} d(x_i, y_j),$$

where A ranges over all bistochastic matrices, which means exactly that $\frac{1}{n}A$ ranges over all couplings between the two uniform marginals as in the definition of the Wasserstein distance (2.6). ■

It is clear that ι^S is natural in X , so we consider it as a transformation $\iota^S : (-)^S \Rightarrow P$ between the power functor at S and the Kantorovich functor. Similarly, $\iota_n : (-)_n \Rightarrow P$.

3.13. LEMMA. *Let $n, m \in \mathbf{N}$, and $X \in \mathbf{CMet}$. Then the following diagram commutes:*

$$\begin{array}{ccc}
 X_m & \xrightarrow{X_{m|mn}} & X_{mn} \\
 & \searrow \iota_m & \swarrow \iota_{mn} \\
 & & PX
 \end{array} \tag{3.11}$$

PROOF. For $\{x_1, \dots, x_m\} \in X_m$,

$$\begin{aligned}
 \iota_{mn} \circ X_{m|mn}(\{x_1 \dots x_m\}) &= \iota_{mn}(\{x_1 \dots x_m, \dots, x_1, \dots, x_m\}) \\
 &= \frac{\delta(x_1) + \dots + \delta(x_m) + \dots + \delta(x_1) + \dots + \delta(x_m)}{mn} \\
 &= \frac{\delta(x_1) + \dots + \delta(x_m)}{m} = \iota_m(x_1 \dots x_m).
 \end{aligned}$$

■

Therefore the symmetrized empirical distribution map ι_n is natural in n . It follows that the empirical distribution map ι^S is natural in S .

3.14. UNIVERSAL PROPERTY.

3.15. DEFINITION. *Let X be a complete metric space, and consider the symmetrized empirical distribution maps $\iota_n : X_n \rightarrow PX$, which are embeddings for each $n \in \mathbf{N}$. We write $I(X)$ for the union of their images,*

$$I(X) := \bigcup_{n \in \mathbf{N}} \iota_n(X_n) \subseteq PX. \tag{3.12}$$

3.16. LEMMA. *$I(X)$ is the colimit of the functor $X_{(-)} : \mathbf{N}^{\text{op}} \rightarrow \mathbf{Met}$, and also of the functor $X^{(-)} : \mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{Met}$, with the ι_n and the ι^S forming the universal cocones.*

PROOF. By Lemma 3.8 and composition of Kan extensions, it is enough to prove this for $X_{(-)}$. So let the $\{g_n : X_n \rightarrow Y\}$ form a cocone, i.e. a family of short maps such that $g_m = g_{mn}X_{m|mn}$. Since the $\iota_n : X_n \rightarrow I(X)$ are jointly epic by definition of $I(X)$, there can be at most one map $I(X) \rightarrow Y$ that is a morphism of cocones, which establishes uniqueness. Concerning existence, every point of $I(X)$ is of the form $\iota_n(\{x_i\})$ for some n and some $\{x_i\} \in X_n$, and we therefore define its image in Y to be $g_n(\{x_i\})$. This is well-defined: if $\iota_n(\{x_i\}) = \iota_m(\{x'_j\})$, then the relative frequencies of all points of X in the multiset $\{x_i\}$ must coincide with those in $\{x'_j\}$. In particular this implies $X_{m|mn}(\{x_i\}) = X_{n|mn}(\{x'_j\})$, which is enough by the assumed naturality of the $\{g_m\}$. Finally, the resulting map is still short since any two points in $I(X)$ come from some common X_n , and $\iota_n : X_n \rightarrow I(X)$ is an isometric embedding. ■

$I(X)$ is not complete unless $|X| \leq 1$. The following result is essentially proven in [Bas15, Proposition 1.9] by reduction to the separable case treated in [Vil09]. We give here an alternative proof that works without mentioning separability.

3.17. THEOREM. *Let X be a metric space. Then $I(X)$ is dense in PX .*

We prove this in several steps, starting with the compact case.

3.18. LEMMA. *Let X be a compact metric space. Then $I(X)$ is dense in PX .*

PROOF. First of all, let $p \in PX$ be finitely supported, but not necessarily with rational coefficients. The measure p is of the form $p = p_1 \delta_{x_1} + \dots + p_n \delta_{x_n}$ for some fixed finite n and positive coefficients p_i . For every $\varepsilon > 0$ and $i = 1, \dots, n-1$, we can find rational numbers q_i such that $q_i \leq p_i$ and $p_i - q_i < \varepsilon$. Define also $q_n := 1 - \sum_{i=1}^{n-1} q_i$, which is rational as well. The measure $q := q_1 \delta_{x_1} + \dots + q_n \delta_{x_n}$ is then an element of $I(X)$. Moreover, denoting by D the (finite) diameter of X ,

$$\begin{aligned} d(p, q) &\leq \sum_{i=1}^{n-1} (p_i - q_i) \cdot d(x_i, x_n) \\ &\leq \sum_{i=1}^{n-1} (p_i - q_i) \cdot D \\ &\leq (n-1) \cdot \varepsilon \cdot D. \end{aligned}$$

Since this can be made arbitrarily small, $I(X)$ is dense in the space of (arbitrary) finitely supported probability measures. It remains to be shown that finitely supported probability measures are dense in PX .

For given $\varepsilon > 0$, the open sets of diameter at most ε cover X . By compactness, already finitely many of these, say U_1, \dots, U_n , cover X . Consider the Boolean algebra generated by the U_i , its atoms are measurable sets A_1, \dots, A_k of diameter at most ε which partition X .

$\{A_i\}$ is then a finite family of measurable subsets, mutually disjoint, which cover X . Choosing arbitrary $y_i \in A_i$, we have $d(x_i, y_i) < \varepsilon$ for every $x_i \in A_i$. For given $p \in PX$, the probability measure

$$p_\varepsilon := \sum_{i=1}^k p(A_i) \delta(y_i). \quad (3.13)$$

is finitely supported. To see that it is close to p , we choose a convenient joint,

$$m := \sum_{i=1}^k p|_{A_i} \otimes \delta(y_i), \quad (3.14)$$

where $p|_{A_i}$ is the measure with $p|_{A_i}(B) := p(B \cap A_i)$. Therefore

$$d_{PX}(p, p_\varepsilon) \leq \int_{X \times X} d_X(x, y) dm(x, y) = \sum_i \int_{A_i \times X} d_X(x, y) dp(x) \delta(y_i)(y) dy$$

$$= \sum_i \int_{A_i} d_X(x, y_i) dp(x) \leq \sum_i \int_{A_i} \varepsilon dp(x) = \varepsilon \sum_i p(A_i) = \varepsilon,$$

as was to be shown. ■

Before getting to the general case, we record another useful fact.

3.19. LEMMA. *Let $p, q_1, q_2 \in PX$ and $\lambda \in [0, 1]$. Then*

$$d_{PX}(\lambda q_1 + (1 - \lambda)p, \lambda q_2 + (1 - \lambda)p) = \lambda d_{PX}(q_1, q_2). \tag{3.15}$$

This follows immediately from the duality (2.7), but it is instructive to derive the inequality ‘ \leq ’ directly by using the fact that any coupling $r \in \Gamma(q_1, q_2)$ gives a coupling

$$\lambda r + (1 - \lambda)(P\Delta)(p) \in \Gamma(\lambda q_1 + (1 - \lambda)p, \lambda q_2 + (1 - \lambda)p)$$

where $\Delta : X \rightarrow X \times X$ is the diagonal embedding, and the second term does not contribute to the expected distance as it is supported on the diagonal.

3.20. LEMMA. *Let X be a metric space. Then the set of compactly supported probability measures is dense in PX .*

PROOF. We first show that boundedly supported measures are dense in PX by finite first moment, and then that compactly supported measures are dense in boundedly supported measures by tightness.

For the first part, let $p \in PX$ and $x_0 \in X$ be given. Let $B(x_0, \rho)$ be the closed ball of radius $\rho > 0$ around x_0 . We would like to approximate p by the boundedly supported measure $p|_{B(x_0, \rho)}$, but this is not normalized. The most convenient way to fix this is to use

$$p' := p|_{B(x_0, \rho)} + p(X \setminus B(x_0, \rho)) \delta(x_0)$$

By decomposing

$$p = p|_{B(x_0, \rho)} + p|_{X \setminus B(x_0, \rho)} \tag{3.16}$$

we can compute

$$\begin{aligned} d_{PX}(p, p') &\stackrel{(3.15)}{=} p(X \setminus B(x_0, \rho)) d_{PX}\left(\delta(x_0), \frac{p|_{X \setminus B(x_0, \rho)}}{p(X \setminus B(x_0, \rho))}\right) \stackrel{(2.8)}{=} \int_{X \setminus B(x_0, \rho)} d(x, x_0) dp(x) \\ &= \int_X d(x, x_0) dp(x) - \int_{B(x_0, \rho)} d(x, x_0) dp(x). \end{aligned}$$

The second term on the right-hand side is the expectation value of the function

$$f_\rho(x) := \begin{cases} d(x, x_0) & \text{if } d(x, x_0) \leq \rho, \\ 0 & \text{otherwise.} \end{cases} \tag{3.17}$$

which converges pointwise to $d(-, x_0)$ as $\rho \rightarrow \infty$. By monotone convergence, this term converges to the first term, $\int_X d_X(x, x_0) dp(x)$, which is finite by the assumption of finite

first moment. Hence $d_{PX}(p, p') \rightarrow 0$ as $\rho \rightarrow \infty$, and the approximating measures p' are boundedly supported.

For the second part, we can then assume that $\text{diam}(X) < \infty$. Let $p \in PX$. For suitably large compact $K \subseteq X$, we would like to approximate p by the compactly supported measure $p|_K$, where $p|_K(A) := p(A \cap K)$, but this is not normalized. The most convenient way to fix this is to choose an arbitrary point $x_0 \in K$, and to use

$$p' := p|_K + p(X \setminus K) \delta(x_0), \quad (3.18)$$

By decomposing

$$p = p|_K + p|_{X \setminus K}, \quad (3.19)$$

we can compute

$$d_{PX}(p, p') \stackrel{(3.15)}{=} p(X \setminus K) d\left(\frac{p|_{X \setminus K}}{p(X \setminus K)}, \delta(x_0)\right) \stackrel{(2.8)}{=} p(X \setminus K) \text{diam}(X),$$

By tightness, this tends to 0 as $K \rightarrow X$. ■

Theorem 3.17 then follows as a corollary.

We now consider what happens in the reflective subcategory of *complete* metric spaces, $\mathbf{CMet} \subseteq \mathbf{Met}$.

3.21. THEOREM. *The space PX is the colimit of the functor $X_{(-)} : \mathbf{N}^{\text{op}} \rightarrow \mathbf{CMet}$, and also of the functor $X^{(-)} : \mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{CMet}$.*

PROOF. Use Lemma 3.16 together with Theorem 3.17, and the fact that if Y is a complete metric space with $X \subseteq Y$ dense, then Y is the completion of X with the inclusion as the universal morphism. ■

Since colimits over $\mathbf{FinUnif}^{\text{op}}$ or \mathbf{N}^{op} in a category of functors into \mathbf{Met} or \mathbf{CMet} are computed pointwise⁷, this implies that the Wasserstein space construction in the form of the object $P \in [\mathbf{CMet}, \mathbf{CMet}]$, is the colimit of the power functor construction:

3.22. COROLLARY. *The empirical distribution maps form two colimiting cocones in the following way:*

- (a) *Consider the functor $(=)_{(-)} : \mathbf{N}^{\text{op}} \rightarrow [\mathbf{CMet}, \mathbf{CMet}]$ mapping $n \in \mathbf{N}$ to the symmetrized power functor $X \mapsto X_n$. Then $P \in [\mathbf{CMet}, \mathbf{CMet}]$ is the colimit of $(=)_{(-)}$, with the colimit cocone given by the symmetrized empirical distribution maps $\iota_n : (-)_n \Rightarrow P$.*
- (b) *Consider the functor $(=)^{(-)} : \mathbf{FinUnif}^{\text{op}} \rightarrow [\mathbf{CMet}, \mathbf{CMet}]$ mapping $S \in \mathbf{FinUnif}$ to the power functor $X \mapsto X^S$. Then $P \in [\mathbf{CMet}, \mathbf{CMet}]$ is the colimit of $(=)^{(-)}$, with the colimit cocone given by the empirical distribution maps $\iota^S : (-)^S \Rightarrow P$.*

⁷Technically, this relies on the fact that such colimits always exists in \mathbf{Met} and \mathbf{CMet} , per Lemma A.1.

3.23. REMARK. Readers concerned with size issues may find it problematic that the endofunctor category $[\mathbf{CMet}, \mathbf{CMet}]$ is not locally small, so that the above universal properties potentially involve bijections between proper classes (or large sets). One way to see that $[\mathbf{CMet}, \mathbf{CMet}]$ is not locally small is to borrow the fact that the functor category $[\mathbf{CMet}, \mathbf{Set}^{\text{op}}]$ is not small from [FS95], and choose a faithful functor $\mathbf{Set}^{\text{op}} \rightarrow \mathbf{CMet}$, such as the composition of the power set functor $\mathbf{Set}^{\text{op}} \rightarrow \mathbf{Set}$ with the discrete metric space functor $\mathbf{Set} \rightarrow \mathbf{CMet}$ which equips every set with the discrete $\{0, 1\}$ -valued metric. In this way, a large hom-set in $[\mathbf{CMet}, \mathbf{Set}^{\text{op}}]$ embeds into a hom-set of $[\mathbf{CMet}, \mathbf{CMet}]$, which must therefore also be large.

It may be possible to alleviate this problem by uncurrying, using $(=)_{(-)} : \mathbf{N}^{\text{op}} \times \mathbf{CMet} \rightarrow \mathbf{CMet}$ and $(=)^{(-)} : \mathbf{FinUnif}^{\text{op}} \times \mathbf{CMet} \rightarrow \mathbf{CMet}$, as in the theory of graded monads developed in [FKM16].

4. Monad structure

The main result of this section is that the functor P is part of a monad, with units and compositions defined in a way analogous to the Giry monad [Gir82]. It was proven in [vB05] that the restriction of P to compact metric spaces carries a monad structure. In the spirit of categorical probability theory, the monad multiplication is given by averaging a measure on measures to a measure, and the unit by assigning to each point its Dirac measure.

An appealing feature of the Kantorovich functor is that its monad structure can be constructed directly from the colimit characterization in terms of the power functors defined in Section 3. This uses the fact that the power functors carry the structure of a monad *graded* by $\mathbf{FinUnif}^{\text{op}}$, in the sense of a lax monoidal functor⁸ into the endofunctor category $[\mathbf{CMet}, \mathbf{CMet}]$, and similarly for the symmetrized power functors in terms of \mathbf{N}^{op} . This construction of the Kantorovich monad extends the idea that probability measures are formal limits of finite samples to the level of integration.

4.1. THE POWER FUNCTORS FORM A GRADED MONAD. As we will see next, the functor $(=)^{(-)} : \mathbf{FinUnif}^{\text{op}} \rightarrow [\mathbf{CMet}, \mathbf{CMet}]$ has a canonical strong monoidal structure with respect to the monoidal structure on $\mathbf{FinUnif}$ given by the cartesian product. We assume the latter to be strict for notational convenience.

Concerning the unit, there is a canonical transformation $\delta : 1_{\mathbf{CMet}} \Rightarrow (=)^1$ with components given by the identity isomorphisms $X \cong X^1$. For the multiplication, we use the currying maps $E^{S,T} : (X^S)^T \cong X^{S \times T}$. It takes a T -indexed family of S -indexed families $\{\{x_{ij}\}_{i \in S}\}_{j \in T}$ to the $(S \times T)$ -indexed family $\{x_{ij}\}_{i \in S, j \in T}$. A straightforward computation shows that $E^{S,T}$ preserves distances, since distances add up across all components i and j and get rescaled by $|S| \cdot |T|$ in both cases. It is also clear that $E^{S,T}$ is natural in X .

⁸An ordinary monad on a category \mathbf{C} is graded by the terminal category 1 : being a monoid in $[\mathbf{C}, \mathbf{C}]$, it is equivalently a lax monoidal functor $1 \rightarrow [\mathbf{C}, \mathbf{C}]$.

We find it curious that at this stage, both of these structure maps are isomorphisms, resulting in a strong monoidal functor. While the relevant coherence properties are immediate by the universal properties, we state them here for convenient reference.

4.2. **THEOREM.** *The above structure transformations δ and $E^{-,-}$ equip the functor $(=)^{(-)}$ with a strong monoidal structure, i.e. the following diagrams commute for all $X \in \mathbf{CMet}$:*

- *The unit triangles*

$$\begin{array}{ccc} X^S & \xrightarrow{\delta} & (X^S)^1 \\ & \searrow & \downarrow E^{S,1} \\ & & X^{S \times 1} \end{array} \quad \begin{array}{ccc} X^S & \xrightarrow{\delta^S} & (X^1)^S \\ & \searrow & \downarrow E^{1,S} \\ & & X^{1 \times S} \end{array} \quad (4.1)$$

- *The associativity square*

$$\begin{array}{ccc} ((X^R)^S)^T & \xrightarrow{E^{S,T}} & (X^R)^{S \times T} \\ \downarrow (E^{R,S})^T & & \downarrow E^{R,S \times T} \\ (X^{R \times S})^T & \xrightarrow{E^{R \times S, T}} & X^{R \times S \times T} \end{array} \quad (4.2)$$

For the proof, it is enough to verify commutativity at the level of the underlying sets, where these are standard properties of currying which follow from the universal property of exponential objects.

4.3. **THE SYMMETRIZED POWER FUNCTORS FORM A GRADED MONAD.** We now move on to consider the analogous structure on the symmetrized power functors $X \mapsto X_n$. By definition, the quotient map $q_n : X^n \rightarrow X_n$ is the universal map which coequalizes the action of the symmetric group S_n permuting the factors. In order to analyze the graded monad structure, we need to analyze the power of a power. The four ways of forming a power of a power fit into the square

$$\begin{array}{ccc} (X^m)^n & \xrightarrow{q_n} & (X^m)_n \\ \downarrow (q_m)^n & & \downarrow (q_m)_n \\ (X_m)^n & \xrightarrow{q_n} & (X_m)_n \end{array} \quad (4.3)$$

which commutes by naturality of q_n . The left arrow has a universal property as well:

4.4. **LEMMA.** *In \mathbf{CMet} , the morphism $(q_m)^n$ is the universal map out of $(X^m)^n$ which coequalizes the action of $(S_m)^{\times n}$ given by acting on each outer factor separately.*

PROOF. For every space Y , the map $Y \otimes q_m : Y \otimes X^m \rightarrow Y \otimes X_m$ is also the coequalizer of the S_m -action on X^m , thanks to Proposition A.4. Therefore $(q_m)^{\otimes n} : (X^m)^{\otimes n} \rightarrow (X_m)^{\otimes n}$ is the coequalizer of the factor-wise action of $(S_m)^{\times n}$ on $(X^m)^{\otimes}$. Finally, the analogous statement for $(q_m)^n : (X^m)^n \rightarrow (X_m)^n$ follows by rescaling both metrics by $1/n$, which is an automorphism of the category and therefore preserves colimits. \blacksquare

It follows that the diagonal morphism is the universal morphism which coequalizes the action of the wreath product group $S_m \wr S_n$, where S_n acts on $(S_m)^{\times n}$ by permutation of the factors. We are not aware of any description for $(q_m)_n$ other than the factorization across q_n by the universal property of the latter.

We now define $E_{m,n} : (X_m)_n \rightarrow X_{mn}$ by the universal property of the $S_m \wr S_n$ -quotient map $(X^m)^n \rightarrow (X_m)_n$ as the unique morphism which makes the diagram

$$\begin{array}{ccc} (X^m)^n & \xrightarrow{E^{m,n}} & X^{mn} \\ \downarrow & & \downarrow q_{mn} \\ (X_m)_n & \xrightarrow{E_{m,n}} & X_{mn} \end{array} \tag{4.4}$$

commute. Explicitly, $E_{m,n}$ takes a multiset of n multisets of cardinality m and forms the union over the outer layer, resulting in a single multiset of cardinality mn . This is a graded version of the multiplication in the commutative monoid monad; in particular, in contrast to the $E^{m,n}$, the $E_{m,n}$ are not isomorphisms (unless $m = 1$ or $n = 1$). Naturality in X follows directly from the definition. Concerning the unit, we have the composite isomorphism $X \cong X^1 \cong X_1$, which we also denote by δ .

4.5. THEOREM. *The above structure transformations δ and $E_{-, -}$ equip the functor $(=)_{(-)}$ with a lax monoidal structure, meaning that the following diagrams commute for all $X \in \mathbf{CMet}$:*

- *The unit triangles*

$$\begin{array}{ccc} X_m & \xrightarrow{\delta} & (X_m)_1 \\ & \searrow & \downarrow E_{m,1} \\ & & X_{m \times 1} \end{array} \quad \begin{array}{ccc} X_m & \xrightarrow{\delta_m} & (X_1)_m \\ & \searrow & \downarrow E_{1,m} \\ & & X_{1 \times m} \end{array} \tag{4.5}$$

- *The associativity square*

$$\begin{array}{ccc} ((X_\ell)_m)_n & \xrightarrow{E_{m,n}} & (X_\ell)_{mn} \\ \downarrow (E_{\ell,m})_n & & \downarrow E_{\ell,mn} \\ (X_{\ell m})_n & \xrightarrow{E_{\ell m,n}} & X_{\ell mn} \end{array} \tag{4.6}$$

PROOF. We reduce the claim to Theorem 4.2. Only the associativity square is nontrivial.

By reasoning similarly to (4.3), composing the quotient maps results in a unique

epimorphism $((X^\ell)^m)^n \rightarrow ((X_\ell)_m)_n$. In fact, we get a cube:

$$\begin{array}{ccccc}
 & & ((X^\ell)^m)_n & \xrightarrow{(q_m)_n} & ((X^\ell)_m)_n \\
 & \nearrow q_n & \downarrow ((q_l)^m)_n & & \nearrow q_n \\
 ((X^\ell)^m)^n & \xrightarrow{(q_m)^n} & ((X^\ell)_m)^n & & ((X^\ell)_m)_n \\
 \downarrow ((q_l)^m)_n & & \downarrow ((q_l)_m)_n & & \downarrow ((q_l)_m)_n \\
 & \nearrow q_n & ((X_\ell)^m)_n & \xrightarrow{(q_m)_n} & ((X_\ell)_m)_n \\
 & & \downarrow ((q_l)_m)_n & & \nearrow q_n \\
 ((X_\ell)^m)^n & \xrightarrow{(q_m)^n} & ((X_\ell)_m)^n & & ((X_\ell)_m)_n
 \end{array} \tag{4.7}$$

where the top, bottom, right, and left faces commute by naturality of q_n , and the front and back faces commute by the naturality of q_m . Using this, we consider the cube

$$\begin{array}{ccccc}
 & & ((X_\ell)_m)_n & \xrightarrow{E_{m,n}} & (X_\ell)_{mn} \\
 & \nearrow & \downarrow (E_{\ell,m})_n & & \nearrow \\
 ((X^\ell)^m)^n & \xrightarrow{E^{m,n}} & (X^\ell)^{mn} & & (X_\ell)_{mn} \\
 \downarrow (E^{\ell,m})^n & & \downarrow & & \downarrow E_{\ell,mn} \\
 & \nearrow & (X_{\ell m})_n & \xrightarrow{E_{\ell,mn}} & X_{\ell mn} \\
 & & \downarrow & & \nearrow \\
 (X^{\ell m})^n & \xrightarrow{E^{\ell m,n}} & X^{\ell mn} & & X_{\ell mn}
 \end{array} \tag{4.8}$$

where the unlabeled diagonal arrows are the quotient maps discussed previously. We need to show that the back face commutes. The bottom and right faces commute by (4.4). The top face also commutes, thanks to

$$\begin{array}{ccc}
 ((X^\ell)^m)^n & \xrightarrow{E^{m,n}} & (X^\ell)^{mn} \\
 \downarrow & & \downarrow \\
 ((X_\ell)^m)^n & \xrightarrow{E^{m,n}} & (X_\ell)^{mn} \\
 \downarrow & & \downarrow \\
 ((X_\ell)_m)_n & \xrightarrow{E_{m,n}} & ((X_\ell)_m)_n
 \end{array}$$

and similarly for the left face. Finally, commutativity of the front face follows from Theorem 4.2. Therefore, since $((X^\ell)^m)^n \rightarrow ((X_\ell)_m)_n$ is epi, this implies that the back face commutes as well. \blacksquare

We can also consider the \mathbf{N}^{op} -graded monad $(=)_{(-)}$ as the universal \mathbf{N}^{op} -graded monad that one obtains from the $\mathbf{FinUnif}^{\text{op}}$ -graded monad $(=)^{(-)}$ by change of grading along $\mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{N}^{\text{op}}$. In fact, this follows from Lemma 3.8 and Theorem B.1 in Appendix B:

4.6. **THEOREM.** *Let \mathbf{MonCat} be the bicategory of monoidal categories, lax monoidal functors, and monoidal transformations. Then the lax monoidal functor $(=)_{(-)} : \mathbf{N}^{\text{op}} \rightarrow [\mathbf{CMet}, \mathbf{CMet}]$ is the left Kan extension in \mathbf{MonCat} of $(=)^{(-)} : \mathbf{FinUnif}^{\text{op}} \rightarrow [\mathbf{CMet}, \mathbf{CMet}]$ along $\mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{N}^{\text{op}}$.*

PROOF. By Lemma 3.8, this Kan extension works in \mathbf{Cat} , and it is clear that $\mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{N}^{\text{op}}$ is strong monoidal and essentially surjective. In order to apply Theorem B.1, it remains to check two things: first, that the transformation $q : (=)^{(-)} \rightarrow (=)_{(-)}$ is monoidal, which boils down to the diagram

$$\begin{array}{ccc} (X^m)^n & \xrightarrow{E^{m,n}} & X^{mn} \\ \downarrow q \circ q & & \downarrow q \\ (X_m)_n & \xrightarrow{E_{m,n}} & X_{mn} \end{array}$$

which is (4.4) again. Second, that $q \otimes q$ is an epimorphism in the functor category $[\mathbf{FinUnif}^{\text{op}} \times \mathbf{FinUnif}^{\text{op}}, [\mathbf{CMet}, \mathbf{CMet}]]$, which follows from the fact that even every individual double quotient map $(X^m)^n \rightarrow (X_m)_n$ is an epimorphism. ■

4.7. **THE MONAD STRUCTURE ON THE KANTOROVICH FUNCTOR.** Now that we have shifted the graded monad structure from $\mathbf{FinUnif}^{\text{op}}$ to \mathbf{N}^{op} , we shift it one step further to a lax monoidal functor $1 \rightarrow [\mathbf{CMet}, \mathbf{CMet}]$, i.e. to an ungraded monad on \mathbf{CMet} whose underlying functor is P .

We define the unit and multiplication maps in terms of the power functors and the empirical distribution maps.

4.8. **DEFINITION.** *For $X \in \mathbf{CMet}$ and $n \in \mathbf{N}$, The Dirac delta embedding is the composite*

$$X \xrightarrow{\delta} X^1 \xrightarrow{\iota_1} PX,$$

which we also denote by δ .

Proposition 3.12 implies that δ is an isometric embedding. As a composite of natural transformations, we also have naturality $\delta : \mathbb{1} \Rightarrow P$. Before getting to the multiplication, we need another bit of preparation. As we show in Corollary A.2, \mathbf{CMet} has sifted colimits. These colimits are preserved by the power functors:

4.9. **LEMMA.** *Both the power functors $(-)^S$ and the symmetrized power functors $(-)_n$ preserve sifted colimits in \mathbf{CMet} .*

PROOF. Let \mathbf{D} be a sifted category. Since $(-)^S$ is $(-)^{\otimes S}$ composed with a rescaling, it is enough to show that $(-)^{\otimes S}$ preserves \mathbf{D} -colimits. But since the monoidal product preserves colimits in each argument, $(-)^{\otimes S}$ turns a \mathbf{D} -colimit into a $\mathbf{D}^{\times S}$ -colimit. But since the diagonal functor $\mathbf{D} \rightarrow \mathbf{D}^{\times S}$ is final by the siftedness assumption, the claim for $(-)^S$ follows. The claim for $(-)_n$ follows by commutation of colimits with colimits. ■

\mathbf{N}^{op} is trivially sifted thanks to being directed. However, the category $\mathbf{FinUnif}^{\text{op}}$ itself is not sifted: for example, the spans in $\mathbf{FinUnif}$

$$\begin{array}{ccc} & S & \\ // & & \\ S & & S \end{array} \quad \begin{array}{ccc} & S & \\ // & & \searrow \alpha \\ S & & S \end{array}$$

are not connected by any zig-zag, for any $S \in \mathbf{FinUnif}$ with a non-identity automorphism $\alpha : S \rightarrow S$.

Similarly to the quotient maps $(X^m)^n \rightarrow (X_m)_n$ in (4.3), we have a commutative square

$$\begin{array}{ccc} (X_m)_n & \xrightarrow{(\iota_m)_n} & (PX)_n \\ \downarrow \iota_n & & \downarrow \iota_n \\ P(X_m) & \xrightarrow{P\iota_m} & PPX \end{array} \quad (4.9)$$

where now all maps are isometric embeddings. In the following, we use this composite as the map $(X_m)_n \rightarrow PX$.

4.10. PROPOSITION. *PPX is the colimit of both*

- (a) *the $(X_m)_n$ with the colimiting cocone given by the $\iota_n \circ (\iota_m)_n = P\iota_m \circ \iota_n$ for $m, n \in \mathbf{N}^{\text{op}}$;*
- (b) *the subdiagram of this formed by the $(X_n)_n$ for $n \in \mathbf{N}^{\text{op}}$.*

While measures on spaces of measures are often quite delicate to handle, this result gives a concrete way to work with them in terms of finite data only. Although we do not currently have any use for even higher powers of P , the analogous statement holds for any $P^n X$.

PROOF. The second claim follows from the first since \mathbf{N}^{op} is sifted. For the first, the Lemma 4.9 tells us that the $(\iota_m)_n : (X_m)_n \rightarrow (PX)_n$ form a colimiting cocone for each n ; the claim then follows from the construction of a colimit over $\mathbf{N}^{\text{op}} \times \mathbf{N}^{\text{op}}$ by first taking the colimit over the first factor and then over the second. ■

4.11. LEMMA. For $X \in \mathbf{CMet}$, there is a unique morphism $E : PPX \rightarrow PX$ such that

$$\begin{array}{ccc} (X_m)_n & \xrightarrow{E_{m,n}} & X_{mn} \\ \downarrow & & \downarrow \\ PPX & \xrightarrow{E} & PX \end{array} \quad (4.10)$$

commutes for all $m, n \in \mathbf{N}$.

The map $E : PPX \rightarrow PX$ amounts to taking the *expected distribution*.

PROOF. This amounts to showing that the $\iota_{mn} \circ E_{m,n}$ form a cocone to which the universal property of Proposition 4.10 applies. Since every morphism in \mathbf{N} is a divisibility relation, this corresponds to commutativity of the two diagrams

$$\begin{array}{ccc} (X_m)_n & \xrightarrow{(X_m|\ell m)_n} & (X_{\ell m})_n \\ E_{m,n} \downarrow & & \downarrow E_{\ell m,n} \\ X_{mn} & \xrightarrow{X_{mn}|\ell mn} & X_{\ell mn} \\ \iota_{mn} \searrow & & \swarrow \iota_{\ell mn} \\ & PX & \end{array} \quad \begin{array}{ccc} (X_m)_n & \xrightarrow{(X_m)_n|\ell n} & (X_m)_{\ell n} \\ E_{m,n} \downarrow & & \downarrow E_{m,\ell n} \\ X_{mn} & \xrightarrow{X_{mn}|\ell mn} & X_{\ell mn} \\ \iota_{mn} \searrow & & \swarrow \iota_{\ell mn} \\ & PX & \end{array}$$

for every $\ell \in \mathbf{N}$. The upper squares commute by naturality of E in its two arguments in \mathbf{N} , and the triangles by Lemma 3.13. \blacksquare

$E : PPX \rightarrow PX$ is natural in X thanks to the uniqueness, i.e. we have a natural transformation $E : PP \Rightarrow P$.

Let us show why the map E is exactly the integration map taking the expected distribution. Denote for now by \tilde{E} the usual integration map, i.e. for all $\mu \in PPX$, let $\tilde{E}\mu \in PX$ be the measure mapping every Lipschitz function $f : X \rightarrow \mathbb{R}$ to

$$\int_X f d(\tilde{E}\mu) := \int_{PX} \left(\int_X f dp \right) d\mu(p),$$

\tilde{E} is short because the map $p \mapsto \int_X f dp$ is. It furthermore makes diagram (4.10) commute, since for all $\{\{x_{11}, \dots, x_{m1}\}, \dots, \{x_{1n}, \dots, x_{mn}\}\}$ in $(X^M)^N$, by linearity of the integral:

$$\begin{aligned} & \int f d(\tilde{E} \circ \iota_n \circ (\iota_m)_n \{\{x_{11}, \dots, x_{m1}\}, \dots, \{x_{1n}, \dots, x_{mn}\}\}) \\ &= f(x_{11}) + \dots + f(x_{m1}) + \dots + f(x_{1n}) + \dots + f(x_{mn}) \\ &= \int f d(\iota_{mn} \circ E_{m,n} \{\{x_{11}, \dots, x_{m1}\}, \dots, \{x_{1n}, \dots, x_{mn}\}\}). \end{aligned}$$

Therefore, again by uniqueness, $\tilde{E} = E$.

4.12. MONAD AXIOMS. E and δ satisfy the monad axioms. This can be proven using the universal property and the monoidal properties of the power functors described in 4.3.

4.13. THEOREM. (P, δ, E) is a monad on \mathbf{CMet} . In other words, we have commutative diagrams:

$$\begin{array}{ccc}
 P & \xrightarrow{P\delta} & PP & \xleftarrow{\delta P} & P \\
 & \searrow & \downarrow E & \swarrow & \\
 & & P & &
 \end{array} \tag{4.11}$$

and:

$$\begin{array}{ccc}
 PPP & \xrightarrow{PE} & PP \\
 \downarrow EP & & \downarrow E \\
 PP & \xrightarrow{E} & P
 \end{array} \tag{4.12}$$

When equipped with this additional structure, we call the Kantorovich functor P the *Kantorovich monad*.

PROOF. We already know that δ and E are natural. Hence we only need to check the commutativity at each object $X \in \mathbf{CMet}$. Because of the universal property of P , E_n , E and ι , we have the following.

(a) The left unit triangle at X is the back face of the following prism:

$$\begin{array}{ccccc}
 & & PX & \xrightarrow{\delta} & PPX \\
 & \nearrow & \parallel & \nearrow & \downarrow E \\
 X_m & \xrightarrow{(X_m)_1|_n \circ \delta} & (X_m)_n & & PX \\
 & \searrow & \downarrow E_{m,n} & \swarrow & \\
 & & X_{mn} & &
 \end{array} \tag{4.13}$$

Now:

- The front face can be decomposed into the following diagram:

$$\begin{array}{ccccc}
 X_m & \xrightarrow{\delta} & (X_m)_1 & \xrightarrow{(X_m)_1|_n} & (X_m)_n \\
 & \searrow & \downarrow E_{m,1} & & \downarrow E_{m,n} \\
 & & X_m & \xrightarrow{X_m|_n} & X_{mn}
 \end{array} \tag{4.14}$$

which commutes by the left unit diagram of Theorem 4.5, together with naturality of $E_{m,-}$;

- The top face can be decomposed into the following diagram:

$$\begin{array}{ccccc}
X_m & \xrightarrow{\delta} & (X_m)_1 & \xrightarrow{(X_m)_{1|n}} & (X_m)_n \\
\downarrow \iota_m & & \downarrow (\iota_m)_1 & & \downarrow (\iota_m)_n \\
PX & \xrightarrow{\delta} & (PX)_1 & \xrightarrow{(PX)_{1|n}} & (PX)_n \xrightarrow{\iota_n} PPX
\end{array} \tag{4.15}$$

which commutes by naturality of δ and $(-)_1|_n$;

- The right face commutes by Lemma 4.11;
- The left bottom face commutes by the naturality of the empirical distribution map.

The empirical distribution maps are not epic, but across *all* m, n they are *jointly* epic, therefore the back face has to commute as well.

- (b) The right unit triangle at X is the back face of the following prism:

$$\begin{array}{ccccc}
& & PX & \xrightarrow{P\delta} & PPX \\
& \nearrow & & \parallel & \downarrow E \\
X_n & \xrightarrow{(X_{1|m} \circ \delta)_n} & (X_m)_n & & PX \\
& \searrow & \downarrow E_{m,n} & \nearrow & \\
& & X_{mn} & &
\end{array} \tag{4.16}$$

Now:

- The front face can be decomposed into the following diagram:

$$\begin{array}{ccccc}
X_n & \xrightarrow{(\delta)_n} & (X_1)_n & \xrightarrow{(X_{1|m})_n} & (X_m)_n \\
\parallel & & \downarrow E_{1,n} & & \downarrow E_{m,n} \\
& & X_n & \xrightarrow{X_n|_{mn}} & X_{mn}
\end{array} \tag{4.17}$$

which commutes by the right unit diagram of Theorem 4.5, together with naturality of $E_{-,n}$;

- The top face can be decomposed into the following diagram:

$$\begin{array}{ccccc}
X_n & \xrightarrow{(\delta)_n} & (X_1)_n & \xrightarrow{(X_{1|m})_n} & (X_m)_n \\
\downarrow \iota_n & & \downarrow \iota_n & & \downarrow \iota_n \\
PX & \xrightarrow{P\delta} & P(X_1) & \xrightarrow{P(X_{1|m})} & P(X_m) \xrightarrow{P\iota_m} PPX
\end{array} \tag{4.18}$$

which commutes by naturality of ι_n ;

- The right face commutes again by Lemma 4.11;
- The left bottom face commutes again by the naturality of the empirical distribution map.

Again, the empirical distribution maps across all m, n are jointly epic, therefore the back face has to commute as well.

(c) The associativity square at each X is the back face of the following cube:

$$\begin{array}{ccccc}
 & & PPPX & \xrightarrow{PE} & PPX \\
 & \nearrow & \downarrow E & & \nearrow \\
 ((X_\ell)_m)_n & \xrightarrow{(E_{\ell,m})_n} & (X_{\ell m})_n & & PPX \\
 \downarrow E_{m,n} & & \downarrow E_{\ell m,n} & & \downarrow E \\
 & \nearrow & PPX & \xrightarrow{E} & PX \\
 (X_\ell)_{mn} & \xrightarrow{E_{\ell,mn}} & X_{\ell mn} & & \nearrow
 \end{array} \tag{4.19}$$

where the map $((X_\ell)_m)_n \rightarrow PPPX$ is uniquely obtained in the same way as the map $((X^\ell)^m)^n \rightarrow ((X_\ell)_m)_n$ in the proof of Theorem 4.5, using naturality of ι instead of q . Now:

- The front face is just the associativity square of Theorem 4.5;
- The top face can be decomposed into:

$$\begin{array}{ccccc}
 ((X_\ell)_m)_n & \longrightarrow & (PPX)_n & \xrightarrow{\iota_n} & PPPX \\
 \downarrow (E_{\ell,m})_n & & \downarrow (E)_n & & \downarrow PE \\
 (X_{\ell m})_n & \longrightarrow & (PX)_n & \xrightarrow{\iota_n} & PPX
 \end{array} \tag{4.20}$$

which commutes by Lemma 4.11, and by naturality of ι_n ;

- The left, right, and bottom faces commute by Lemma 4.11.

Once again, the empirical distribution maps across all ℓ, m, n are jointly epic, therefore the back face has to commute as well.

It follows that (P, δ, E) is a monad. ■

In analogy with Theorem 4.6, we can now conclude that P is the monad that one obtains by taking the $\mathbf{FinUnif}^{\text{op}}$ -graded monad $(=)^{(-)}$ or the \mathbf{N}^{op} -graded monad $(=)_{(-)}$ and “crushing them down” universally to an ungraded monad:

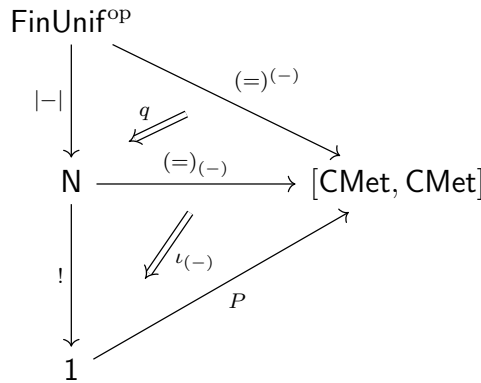
4.14. THEOREM. *As a lax monoidal functor, $P : \mathbf{1} \rightarrow [\mathbf{CMet}, \mathbf{CMet}]$ is the Kan extension in \mathbf{MonCat}*

(a) *of the strong monoidal functor $(=)^{(-)} : \mathbf{FinUnif}^{\text{op}} \rightarrow [\mathbf{CMet}, \mathbf{CMet}]$ along $! : \mathbf{FinUnif}^{\text{op}} \rightarrow \mathbf{1}$, and*

(b) *of the lax monoidal functor $(=)_{(-)} : \mathbf{N}^{\text{op}} \rightarrow [\mathbf{CMet}, \mathbf{CMet}]$ along $! : \mathbf{N}^{\text{op}} \rightarrow \mathbf{1}$,*

with respect to the empirical distribution maps making up the universal transformation.

Together with Corollary 3.22 and Theorem 4.6, this means that we have a diagram



in which all 2-cells are Kan extensions, both in \mathbf{Cat} and in \mathbf{MonCat} .

PROOF. By composition of Kan extensions and Theorem 4.6, it is enough to prove the second item. In order to apply Theorem B.1, it remains to check two things: first, that the transformation $\iota_{(-)} : (=)_{(-)} \Rightarrow P$ is monoidal, which boils down to the diagram

$$\begin{array}{ccc}
 (X_m)_n & \xrightarrow{E_{m,n}} & X_{mn} \\
 \iota_n \circ (\iota_m)_n \downarrow & & \downarrow \iota_{mn} \\
 PPX & \xrightarrow{E} & PX
 \end{array}$$

which equals (4.10) again. Second, that $i \otimes i$ is an epimorphism in the functor category $[\mathbf{N}^{\text{op}} \times \mathbf{N}^{\text{op}}, [\mathbf{CMet}, \mathbf{CMet}]]$, which follows from the fact that for every X , the maps $(X_m)_n \rightarrow PPX$ are jointly epic. ■

The uniqueness statement in Theorem B.1 also shows that the monad structure on P is the only one which makes the empirical distribution maps into a morphism of graded monads.

5. Algebras of the Kantorovich monad

A P -algebra for the Kantorovich monad P consists of $A \in \mathbf{CMet}$ together with a map $e : PA \rightarrow A$ such that the following diagrams commute:

$$\begin{array}{ccc} A & \xrightarrow{\delta} & PA \\ & \searrow & \downarrow e \\ & & A \end{array} \qquad \begin{array}{ccc} PPA & \xrightarrow{Pe} & PA \\ \downarrow E & & \downarrow e \\ PA & \xrightarrow{e} & A \end{array}$$

The expectation value of a real- or vector-valued random variable is of central importance in probability theory. Indeed, as is generally the case in categorical probability, this is precisely what the map e encodes: $P\mathbb{R}$ is the space of distributions of real-valued random variables with well-defined expectation value, and the operation of taking the expectation is a morphism $P\mathbb{R} \rightarrow \mathbb{R}$ which makes \mathbb{R} into a P -algebra. Replacing \mathbb{R} by other P -algebras results in a definition of integral or expectation for A -valued random variables $X \rightarrow A$ by composing the induced map $PX \rightarrow PA$ with the algebra map $PA \rightarrow A$. In this sense, the characterization of P -algebras of Theorem 5.4 together with our purely categorical construction of the Kantorovich monad P via Theorem 4.14 below can be regarded as a definition of integral which does not require any measure theory.

A morphism of P -algebras $e_A : PA \rightarrow A$ and $e_B : PB \rightarrow B$ is a short map $f : A \rightarrow B$ such that

$$\begin{array}{ccc} PA & \xrightarrow{Pf} & PB \\ \downarrow e_A & & \downarrow e_B \\ A & \xrightarrow{f} & B \end{array}$$

commutes. We also say that f is P -affine. The Eilenberg-Moore category \mathbf{CMet}^P is then the category of P -algebras and P -affine maps. Any Wasserstein space PX is a free P -algebra, with structure map $e = E : PPA \rightarrow PA$. The Kleisli category \mathbf{CMet}_P is the full subcategory of \mathbf{CMet}^P on the free algebras. Its morphisms are the short maps $X \rightarrow PY$ for complete metric spaces X, Y , which correspond bijectively and naturally to P -affine maps $PX \rightarrow PY$, so that it naturally contains \mathbf{CMet} as a subcategory.

As usual in categorical probability, the Kleisli morphisms should be thought of as *stochastic maps* or *Markov kernels* [Gir82]. An important difference between other approaches to categorical probability theory and the one developed by van Breugel [vB05] and in this work is that these stochastic maps are also required to be short. This leads to the unpleasant phenomenon that conditional expectations do not always exist: for given $p \in PX$ and $f : X \rightarrow Y$, it is generally not possible to write p as the image of the pushforward $(Pf)(p)$ under a Kleisli morphism $PY \rightarrow PX$.

In this section, we will give equivalent characterizations of the P -algebras and their category. In the context of compact and of 1-bounded complete metric spaces, it seems to be known that the Kantorovich monad captures the operations of taking formal binary midpoints [vBHMW05, Section 4]. We develop similar ideas for all complete metric spaces.

By evaluating the structure map on a finitely supported measure, one assigns to every formal convex combination of points another point. In this way, a P -algebra looks like a convex set in which the convex structure interacts well with the metric. And indeed, we will show that the category of P -algebras is equivalent to the category of closed convex subsets of Banach spaces with short affine maps. A similar characterization of the category of algebras of the Radon monad exists, as the category of compact convex sets in locally convex spaces [Świ74]; see also [Kei08] for a more recent exposition. A similarly simple characterization of the algebras of the Giry monad is apparently not known; even for the Giry monad on the category of Polish spaces with continuous maps, the existing characterization is significantly more complicated [Dob06].

5.1. CONVEX SPACES. A set together with an abstract notion of convex combination which satisfies the same equations as convex combinations in a vector space is a *convex space*. This is a notion which has been discovered many times over in various forms, e.g. in [Sto49, Gud73, Świ74]. A convex space can be defined as an algebra of the convex combinations monad on **Set**. This monad assigns to every set M the set of finitely supported probability measures on M , where the unit is again given by the Dirac delta embedding and the multiplication by the formation of the expected measure,

$$\sum_i \alpha_i \delta \left(\sum_j \beta_{ij} \delta(x_{ij}) \right) \mapsto \sum_{i,j} \alpha_i \beta_{ij} \delta(x_{ij})$$

Equivalently, a convex space is a model of the Lawvere theory opposite to the category of stochastic matrices, **FinStoch** [Fri09]. An axiomatization in terms of binary operations is as follows:

5.2. DEFINITION. A convex space is a set A equipped with a family of binary operations $c : [0, 1] \times A \times A \rightarrow A$, such that the following properties hold for all $x, y, z \in A$ and $\lambda, \mu \in [0, 1]$:

- (a) *Unitality*: $c_0(x, y) = x$;
- (b) *Idempotency*: $c_\lambda(x, x) = x$;
- (c) *Parametric commutativity*: $c_\lambda(x, y) = c_{1-\lambda}(y, x)$;
- (d) *Parametric associativity*: $c_\lambda(c_\mu(x, y), z) = c_{\lambda\mu}(x, c_\nu(y, z))$, where:

$$\nu = \begin{cases} \frac{\lambda(1-\mu)}{1-\lambda\mu} & \text{if } \lambda, \mu \neq 1; \\ \text{any number in } [0, 1] & \text{if } \lambda = \mu = 1. \end{cases} \tag{5.1}$$

The category of convex spaces has as morphisms the maps $f : A \rightarrow B$ for which

$$\begin{array}{ccc} A \times A & \xrightarrow{f \times f} & B \times B \\ \downarrow c_\lambda & & \downarrow c_\lambda \\ A & \xrightarrow{f} & B \end{array} \tag{5.2}$$

commutes for every $\lambda \in [0, 1]$.

In the following, we will make use of the standard equivalence between this definition of convex spaces and algebras of the convex combinations monad $C : \mathbf{Set} \rightarrow \mathbf{Set}$.

5.3. EQUIVALENT CHARACTERIZATIONS OF P -ALGEBRAS.

5.4. THEOREM. *The following structures are equivalent on a complete metric space A , in the sense that there is an isomorphism of categories over \mathbf{CMet} :*

- (a) A P -algebra structure;
- (b) A short map $e_n : A_n \rightarrow A$ for each $n \in \mathbf{N}$, such that $e_1 = \delta^{-1}$, and such that the diagrams

$$\begin{array}{ccc}
 A_m & \xrightarrow{A_{m|mn}} & A_{mn} \\
 & \searrow e_m & \swarrow e_{mn} \\
 & & A
 \end{array}
 \qquad
 \begin{array}{ccc}
 (A_m)_n & \xrightarrow{(e_m)_n} & A_n \\
 \downarrow E_{m,n} & & \downarrow e_n \\
 A_{mn} & \xrightarrow{e_{mn}} & A
 \end{array}
 \tag{5.3}$$

commute. Structure-preserving maps are those $f : A \rightarrow B$ for which the diagrams

$$\begin{array}{ccc}
 A_n & \xrightarrow{f_n} & B_n \\
 \downarrow e_n & & \downarrow e_n \\
 A & \xrightarrow{f} & B
 \end{array}
 \tag{5.4}$$

commute for all $n \in \mathbf{N}$.

- (c) A short map $e^S : A^S \rightarrow A$ for each $S \in \mathbf{FinUnif}$, such that $e^1 = \delta^{-1}$, and such that the diagrams

$$\begin{array}{ccc}
 A^T & \xrightarrow{A^\phi} & A^S \\
 & \searrow e^T & \swarrow e^S \\
 & & A
 \end{array}
 \qquad
 \begin{array}{ccc}
 (A^S)^T & \xrightarrow{(e^S)^T} & A^T \\
 \downarrow E^{S,T} & & \downarrow e^T \\
 A^{S \times T} & \xrightarrow{e^{S \times T}} & A
 \end{array}
 \tag{5.5}$$

commute for every $S, T \in \mathbf{FinUnif}$ and $\phi \in \mathbf{FinUnif}(S, T)$. Structure-preserving maps are those $f : A \rightarrow B$ for which the diagrams

$$\begin{array}{ccc}
 A^S & \xrightarrow{f^S} & B^S \\
 \downarrow e^S & & \downarrow e^S \\
 A & \xrightarrow{f} & B
 \end{array}
 \tag{5.6}$$

commute for all $S \in \mathbf{FinUnif}$.

(d) A structure of convex space satisfying a compatibility inequality with the metric,

$$d(c_\lambda(x, z), c_\lambda(y, z)) \leq \lambda d(x, y), \tag{5.7}$$

where the morphisms are the short maps that are also morphisms of convex spaces.

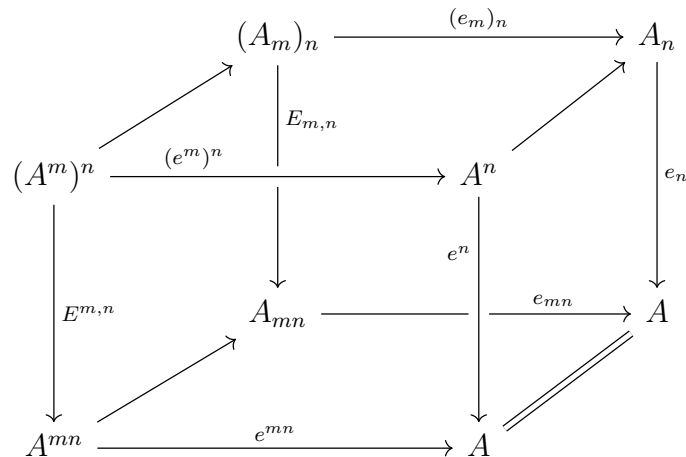
We make two remarks on related literature. First, in the special case of complete separable metric spaces, [MPP16, Theorem 10.9]⁹ can also be interpreted as establishing the equivalence between (a) and (d). Second, the structures in (c) and (b) are different from the *graded algebras* in the sense of [FKM16, Definition 1]: for a graded algebra, the algebra morphisms would have to be of type $(A_m)_n \rightarrow A_{mn}$ and $(A^S)^T \rightarrow A^{S \times T}$, respectively.

It follows from Theorem 5.6 that in structures of type (d), the inequality (5.7) necessarily holds with equality.

PROOF. We first show that the structures of type (a), (c) and (b) are equivalent, using the universal property proven in 3.14.

- (b) ⇔ (c): By composing with the quotient maps $A^S \rightarrow A_{|S|}$, the $(e_n)_{n \in \mathbb{N}}$ determine morphisms $e^S : A^S \rightarrow A$, and conversely by the universal property. The equivalence between the triangles in (5.3) and (5.5) follows from $e_{|S|} = e^S \circ q$ and the diagram (3.6). For the same reason, (5.4) and (5.6) are equivalent.

It remains to verify the equivalence of the squares in (5.3) and (5.5). This follows by a cube similar to (4.8),



where the front face commutes if and only if the back face commutes, since all other faces commute, the quotient map on the upper left is epic, and the identity on the lower right is monic.

⁹Modulo the earlier footnote 1.

- (a) \Leftrightarrow (b): This works similarly. By the universal property of PA , the cocone defined by the first diagram in (5.3) is equivalent to a short map $e : PA \rightarrow A$. The equivalence between the square in (5.3) and the multiplication square of a P -algebra follow by considering the cube

$$\begin{array}{ccccc}
 & & PPA & \xrightarrow{Pe} & PA \\
 & \nearrow & \downarrow E & & \downarrow e \\
 (A_m)_n & \xrightarrow{\quad} & A_n & \xrightarrow{\quad} & A \\
 \downarrow E_{m,n} & & \downarrow e_n & \searrow e & \downarrow \\
 A_{mn} & \xrightarrow{\epsilon_{mn}} & A & \xrightarrow{\quad} & A
 \end{array} \tag{5.8}$$

and using that the upper left diagonals are jointly epic as m and n vary.

- (a) \Rightarrow (d): Finite convex combinations with real coefficients are a special case of Radon measures, and therefore every P -algebra $e : PA \rightarrow A$ also is a convex space in a natural way. Technically, this is based on the morphism of monads

$$\begin{array}{ccc}
 \mathbf{CMet} & \xrightarrow{P} & \mathbf{CMet} \\
 \downarrow U & \nearrow \eta & \downarrow U \\
 \mathbf{Set} & \xrightarrow{C} & \mathbf{Set}
 \end{array}$$

where U is the forgetful functor, and η is the natural transformation with $\eta : CUX \rightarrow UPX$ given by the map which reinterprets a finitely supported measure on UX as a finitely supported measure on X , considered as an element of the underlying set of PX . It is straightforward to check that this is a morphism of monads. Thus we have a functor from P -algebras in \mathbf{CMet} to C -algebras in \mathbf{Set} . In other words, every P -algebra is a convex space in a canonical way.

Let us now check compatibility with the metric. Since e is short, we get

$$\begin{aligned}
 & d(c_\lambda(x, z), c_\lambda(y, z)) \\
 &= d(e(\lambda\delta(x) + (1 - \lambda)\delta(z)), e(\lambda\delta(y) + (1 - \lambda)\delta(z))) \\
 &\leq d(\lambda\delta(x) + (1 - \lambda)\delta(z), \lambda\delta(y) + (1 - \lambda)\delta(z)) \\
 &\stackrel{\text{Lemma 3.19}}{=} \lambda d(\delta(x), \delta(y)) = \lambda d(x, y).
 \end{aligned}$$

- (d) \Rightarrow (c): Intuitively, the e^S correspond to taking convex combinations with equal weights, and commutativity of (5.5) follow from the equations satisfied by taking convex combinations in any convex space. To make this precise, it is most convenient

to consider a convex space as a model of the Lawvere theory $\mathbf{FinStoch}^{\text{op}}$. Considering $\mathbf{FinUnif}$ as a subcategory $\mathbf{FinUnif} \subseteq \mathbf{FinSet} \subseteq \mathbf{FinStoch}$, defining maps $u_S : 1 \rightarrow S$ in $\mathbf{FinStoch}$ which select the uniform distribution on each finite set S results in commutativity of the two diagrams

$$\begin{array}{ccc} & 1 & \\ u_S \swarrow & & \searrow u_T \\ S & \xrightarrow{\phi} & T \end{array} \qquad \begin{array}{ccc} 1 & \xrightarrow{u_{S \times T}} & S \times T \\ \downarrow u_T & & \parallel \\ T & \xrightarrow{u_{S \times T}} & S \times T \end{array}$$

for every $S, T \in \mathbf{FinUnif}$ and $\phi \in \mathbf{FinUnif}(S, T)$. Thus, given a convex space A as a model of $\mathbf{FinStoch}^{\text{op}}$, the u_S become maps $e^S : A^S \rightarrow A$ satisfying the required equations, and every affine map between convex spaces will make (5.6) commute. To see that the e^S are short apply (5.7) twice, to get

$$d(c_\lambda(x, z), c_\lambda(y, w)) \leq \lambda d(x, y) + (1 - \lambda) d(z, w),$$

which generalizes to

$$d\left(e\left(\sum_i \lambda_i \delta(x_i)\right), e\left(\sum_i \lambda_i \delta(y_i)\right)\right) \leq \sum_i \lambda_i d(x_i, y_i) \tag{5.9}$$

by decomposing a general convex combination into a sequence of binary ones and by induction. Shortness of e^S is the special case where the λ_i 's are uniform and equal to $1/|S|$.

It is clear that, starting with a P -algebra A and applying the constructions (a) \Rightarrow (d) \Rightarrow (c), one recovers the underlying (c)-structure of A . To see that the composite functor (d) \Rightarrow (b) \Leftrightarrow (a) \Rightarrow (d) is the identity as well, we claim that two convex space structures c and c' which satisfy the metric compatibility inequality and coincide for convex combinations with rational weights must be equal. Indeed, we prove $d(c_\lambda(x, y), c'_\lambda(x, y)) = 0$ for all $\lambda \in (0, 1)$: first, as λ varies, this distance is bounded because $d(c_\lambda(x, y), y) = d(c_\lambda(x, y), c_\lambda(y, y)) \leq \lambda d(x, y) \leq d(x, y)$, and similarly for c' , so that we get an upper bound of $2d(x, y)$,

$$d(c_\mu(x, y), c'_\mu(x, y)) \leq 2d(x, y) \quad \forall \mu \in [0, 1].$$

We choose a sufficiently small rational $\varepsilon > 0$, as well as a rational $\nu \in (\frac{\lambda - \varepsilon}{1 - \varepsilon}, \frac{\lambda}{1 - \varepsilon})$, and put $z := c_\nu(x, y) = c'_\nu(x, y)$. Then

$$c_\lambda(x, y) = c_\varepsilon(c_\mu(x, y), z), \quad c'_\lambda(x, y) = c'_\varepsilon(c'_\mu(x, y), z),$$

where $\mu := \frac{\lambda - (1 - \varepsilon)\nu}{\varepsilon}$ is in $[0, 1]$ due to the assumed bounds on ν . Now since ε is rational, we can bound the distance between these two points by

$$\begin{aligned} d(c_\lambda(x, y), c'_\lambda(x, y)) &= d(c_\varepsilon(c_\mu(x, y), z), c'_\varepsilon(c'_\mu(x, y), z)) \\ &\leq \varepsilon d(c_\mu(x, y), c'_\mu(x, y)) \leq 2\varepsilon d(x, y), \end{aligned}$$

from which the claim follows as $\varepsilon \rightarrow 0$. ■

We suspect that P -algebras also coincide with the *metric mean-value algebras* of [vBHMW05, Definition 6], if the requirement of 1-boundedness is dropped.

5.5. ALGEBRAS AS CLOSED CONVEX SUBSETS OF BANACH SPACES. If $(E, \|\cdot\|)$ is a Banach space and $A \subseteq E$ is a closed convex subset, then A is a convex space which carries the metric

$$d(x, y) := \|x - y\|$$

with respect to which it is complete. These two structures fulfill the metric compatibility inequality (5.7),

$$\|(\lambda x + (1 - \lambda)z) - (\lambda y + (1 - \lambda)z)\| = \|\lambda x - \lambda y\| = \lambda\|x - y\|,$$

which even holds with equality. Therefore, by Theorem 5.3(d), A is a P -algebra $e : PA \rightarrow A$ in a canonical way. In particular, we can *define* the expectation value $\int_A x dp(x)$ of any $p \in PA$ (which has finite first moment) to be $e(p)$. By functoriality of P , this also defines the expectation value of any Banach-space valued random variable with finite first moment on any (other) complete metric space.

Let $\mathbf{ConvBan}$ be the category whose objects are closed convex subsets of Banach spaces $A \subseteq E$, and whose morphisms $f : (A \subseteq E) \rightarrow (B \subseteq F)$ are the short affine maps $f : A \rightarrow B$.¹⁰ We have a canonical functor $\mathbf{ConvBan} \rightarrow \mathbf{CMet}^P$ which is fully faithful.

Moreover, it was shown in [CF13] that this functor is essentially surjective, meaning that every P -algebra in the form (d) is isomorphic both as a convex space and as a metric space (or equivalently as a P -algebra) to a closed convex subset of a Banach space. We therefore conclude that P -algebras and closed convex subsets of Banach spaces are the same concept:

5.6. THEOREM. *The functor $\mathbf{CBan} \rightarrow \mathbf{CMet}^P$ is an equivalence of categories.*

A. Colimits of (complete) metric spaces

The main goal of this section is to prove that the monoidal structures of \mathbf{Met} and \mathbf{CMet} preserve colimits (Proposition A.4).

To this end, it is useful to consider \mathbf{Met}_∞ , by which we mean the category defined just as \mathbf{Met} above, with the only difference that we now allow distances in $[0, \infty]$. Given $(X, d) \in \mathbf{Met}_\infty$, the “having finite distance” relation partitions the set X into equivalence classes, each of which is a metric space with finite distances and therefore an object of \mathbf{Met} . The morphisms of \mathbf{Met}_∞ must respect this equivalence relation. By this line of reasoning, it is straightforward to verify that \mathbf{Met}_∞ is a free coproduct completion of \mathbf{Met} , with the inclusion functor $\mathbf{Met} \rightarrow \mathbf{Met}_\infty$ as the universal arrow. The advantage of \mathbf{Met}_∞ over \mathbf{Met} is

¹⁰One might be tempted to define morphisms to be equivalence classes of short affine maps $f : E \rightarrow F$ which satisfy $f(A) \subseteq B$, where two such maps are identified whenever they are equal on A . This is not equivalent, since a short affine map $A \rightarrow F$ can in general not be extended to a short (or even merely continuous) affine map $E \rightarrow F$.

that \mathbf{Met}_∞ is cocomplete, with colimits of the following form. Let $U : \mathbf{Met}_\infty \rightarrow \mathbf{Set}$ be the underlying set functor and $F : \mathbf{C} \rightarrow \mathbf{Met}_\infty$ be a diagram indexed by a small category \mathbf{C} . Equip the set colim $(U \circ F)$ with the largest pseudometric (with possibly infinite values) which makes the colimiting cocone components $\ell_c : U(F(c)) \rightarrow \text{colim}(U \circ F)$ short; this pseudometric can be computed by starting with the distance function

$$\hat{d}(x, y) := \inf \{ d_{F(c)}(x', y') : c \in \mathbf{C}, \ell_c(x') = x, \ell_c(y') = y \}, \tag{A.1}$$

and then by shortest path optimization on \hat{d} in order to enforce the triangle inequality, resulting in a pseudometric on colim $(U \circ F)$ (with possibly infinite values). Identifying all pairs of points with distance zero results in an object of \mathbf{Met}_∞ , and it is not hard to see that this is the colimit of F in \mathbf{Met}_∞ .

The very same statements apply to the category \mathbf{CMet}_∞ , defined as the full subcategory of \mathbf{CMet} on the complete spaces, and its full subcategory \mathbf{CMet} . Since $\mathbf{CMet}_\infty \subseteq \mathbf{Met}_\infty$ is also a reflective subcategory of \mathbf{Met}_∞ with the completion as the reflector, colimits in \mathbf{CMet}_∞ can be computed as colimits in \mathbf{Met}_∞ using the procedure above and then taking the completion.

A.1. LEMMA. *A functor $F : \mathbf{C} \rightarrow \mathbf{Met}$ (resp. $F : \mathbf{C} \rightarrow \mathbf{CMet}$) for a small category \mathbf{C} has a colimit if and only if the full subcategory on those objects $c \in \mathbf{C}$ with $F(c) \neq \emptyset$ is connected.*

Intuitively, the only obstruction to the existence of colimits is the lack of coproducts.

PROOF. We only treat the case of \mathbf{Met} , since the case of \mathbf{CMet} is completely analogous.

If $F : \mathbf{C} \rightarrow \mathbf{Met}$ satisfies the connectedness assumption, then the shortest path optimization for (A.1) will clearly have a finite result, so that the resulting colimit in \mathbf{Met}_∞ is actually an object in \mathbf{Met} , and is a colimit in there as well since $\mathbf{Met} \subseteq \mathbf{Met}_\infty$ is full.

Conversely, suppose that $F : \mathbf{C} \rightarrow \mathbf{Met}$ does not satisfy the connectedness assumption. By restricting F to the full subcategory \mathbf{C}' on those $c \in \mathbf{C}$ with $F(c) \neq \emptyset$, we obtain $F' : \mathbf{C}' \rightarrow \mathbf{Met}$ such that F -cocones and F' -cocones are in canonical bijection, thanks to initiality of \emptyset . We can therefore assume $F = F'$ without loss of generality, or equivalently $F(c) \neq \emptyset$ for all c . Then the non-connectedness assumption also means that the category \mathbf{C} is not connected, and therefore can be partitioned into disjoint subcategories \mathbf{C}_1 and \mathbf{C}_2 without any morphisms between them, and inhabited by objects $c_1 \in \mathbf{C}_1$ and $c_2 \in \mathbf{C}_2$ with points $x_1 \in F(c_1)$ and $x_2 \in F(c_2)$. Considering the space $Y := \{1, 2\}$ with $d(1, 2) = r \in \mathbb{R}_+$ a parameter, there is a cocone from F to Y which takes all points in the spaces assigned to \mathbf{C}_1 to $1 \in Y$, and all points in the spaces assigned to \mathbf{C}_2 to $2 \in Y$. Hence if the colimit existed in \mathbf{Met} , then the distance of the images of x_1 and x_2 in the colimit space would have to be lower bounded by $d(1, 2) = r$. This is absurd since r was arbitrary. ■

The characterization of sifted categories as inhabited categories with connected categories of cospans [AR01, Theorem 1.6] now implies:

A.2. COROLLARY. *Sifted colimits exist both in \mathbf{Met} and in \mathbf{CMet} .*

In addition, we obtain:

A.3. PROPOSITION. *The inclusion functor $\mathbf{Met} \subseteq \mathbf{Met}_\infty$ (resp. $\mathbf{CMet} \subseteq \mathbf{CMet}_\infty$) preserves all colimits which exist in \mathbf{Met} .*

PROOF. As we saw in the proof of Lemma A.1, the colimits which exist in \mathbf{Met} are all also colimits in \mathbf{Met}_∞ , and likewise in the complete case. ■

A.4. PROPOSITION. *For every $X \in \mathbf{Met}$ (resp. $X \in \mathbf{CMet}$), the functor $X \otimes - : \mathbf{Met} \rightarrow \mathbf{Met}$ (resp. $X \otimes - : \mathbf{CMet} \rightarrow \mathbf{CMet}$) preserves colimits.*

PROOF. We first treat the case of \mathbf{Met} . We note that \mathbf{Met}_∞ is monoidal closed under the ℓ^1 -product, using the supremum distance for turning the hom-sets $\mathbf{Met}(X, Y)$ into metric spaces; this is essentially [Law73, Section 2]. Hence $X \otimes - : \mathbf{Met}_\infty \rightarrow \mathbf{Met}_\infty$ is a left adjoint for every $X \in \mathbf{Met}$, and as such it preserves colimits, including those of \mathbf{Met} .

The complete case is again analogous, noting that the hom-space $\mathbf{CMet}_\infty(X, Y) := \mathbf{Met}_\infty(X, Y)$ with the sup-distance is already complete (by the existence of pointwise limits). This makes \mathbf{CMet}_∞ also closed, and the inclusion $\mathbf{CMet}_\infty \subseteq \mathbf{Met}_\infty$ preserves exponential objects. ■

B. Kan extensions of lax monoidal functors

In this section we give a criterion for when a Kan extension of lax monoidal functors is a Kan extension in the bicategory \mathbf{MonCat} of monoidal categories, lax monoidal functors and monoidal natural transformations. This in particular comprises the Kan extensions used to define the monad structure of P in Section 4.

Kan extensions in \mathbf{MonCat} , or more generally, Kan extensions preserving some given algebraic structure, are known in the literature as *algebraic Kan extensions* [Web16, Kou15]. There are some results in the literature on when a left Kan extension in \mathbf{Cat} of lax or strong monoidal functors is again monoidal (see [MT08, Theorem 1] and [Pat12, Proposition 4]) in such a way that the Kan extension also holds in \mathbf{MonCat} . There are also general results on when a Kan extension on a 2-category or double category can be lifted to a Kan extension in the 2-category of pseudoalgebras of a 2-monad (see [Kou15, Theorem 1.1b], [Kou, Theorem 7.7] and [Web16, Theorem 2.4.4]), which can be applied to the monoidal category 2-monad. In [FP18], we also derive a statement of this type, which is very close to an instance of Koudenburg’s result [Kou, Theorem 7.7]. For a more detailed discussion of the literature, and for additional context on algebraic Kan extensions, we refer to the discussion in [FP18, Section 1.2].

None of the results cited above applies verbatim to the situation of Section 4. Therefore, here we derive a result of this type, tailored to our needs.

For a monoidal category \mathbf{C} , we denote its unit $e : \mathbf{1} \rightarrow \mathbf{C}$ and multiplication $\otimes : \mathbf{C} \times \mathbf{C} \rightarrow \mathbf{C}$ without explicit reference to the category. For a lax monoidal functor F , we denote its unit by η_F and its multiplication by μ_F .

B.1. THEOREM. *Let the following hypotheses be satisfied:*

- In MonCat , we have a diagram

$$\begin{array}{ccc}
 \mathbf{C} & \xrightarrow{F} & \mathbf{D} \\
 \searrow G & \Downarrow \lambda & \nearrow L \\
 & \mathbf{C}' &
 \end{array}
 \tag{B.1}$$

- λ makes L into the left Kan extension of F along G in Cat .
- $G : \mathbf{C} \rightarrow \mathbf{C}'$ is strong monoidal and essentially surjective.
- The natural transformation $\lambda(-) \otimes \lambda(-)$, by which we mean

$$\begin{array}{ccc}
 \mathbf{C} \times \mathbf{C} & \xrightarrow{F \times F} & \mathbf{D} \times \mathbf{D} \xrightarrow{\otimes} \mathbf{D} \\
 \searrow G \times G & \Downarrow \lambda \times \lambda & \nearrow L \times L \\
 & \mathbf{C}' \times \mathbf{C}' &
 \end{array}
 \tag{B.2}$$

is an epimorphism in the functor category $\text{Cat}(\mathbf{C} \times \mathbf{C}, \mathbf{D})$.

Then λ makes L into the left Kan extension of F along G also in MonCat . Moreover, the monoidal structure of L is the only monoidal structure that can be put on L such that λ is monoidal.

In comparison to previous results, this is closest to [Kou15, Theorem 1.1b]. In fact, Koudenburg’s theorem could alternatively be used for the proof of Theorem 4.14, but not for the proof of Theorem 4.6, for which we really need Theorem B.1.

PROOF. Given a lax monoidal functor $X : \mathbf{C}' \rightarrow \mathbf{D}$ and a monoidal transformation $\chi : F \Rightarrow X \circ G$, we can apply the Kan extension property in Cat , so that there exists a unique $u : L \Rightarrow X$ such that

$$\begin{array}{ccc}
 \begin{array}{ccc}
 \mathbf{C} & \xrightarrow{F} & \mathbf{D} \\
 \searrow G & \Downarrow \lambda & \nearrow L \\
 & \mathbf{C}' & \\
 & \nearrow u & \\
 & \mathbf{D} &
 \end{array}
 & \equiv &
 \begin{array}{ccc}
 \mathbf{C} & \xrightarrow{F} & \mathbf{D} \\
 \searrow G & \Downarrow \chi & \nearrow X \\
 & \mathbf{C}' &
 \end{array}
 \end{array}
 \tag{B.3}$$

What we need to show is that this u is automatically monoidal. We first prove that it respects the units,

$$\begin{array}{ccc}
 \begin{array}{ccc}
 & \mathbf{C}' & \\
 e \nearrow & & \searrow e \\
 \mathbf{1} & \xrightarrow{\eta_L} & \mathbf{C}' \\
 \eta_L \nearrow & \xrightarrow{u} & \mathbf{D} \\
 e \searrow & & \nearrow X \\
 & \mathbf{D} &
 \end{array}
 & \equiv &
 \begin{array}{ccc}
 & \mathbf{C}' & \\
 e \nearrow & & \searrow e \\
 \mathbf{1} & \xrightarrow{\eta_X} & \mathbf{C}' \\
 \eta_X \nearrow & & \mathbf{D} \\
 e \searrow & & \nearrow X \\
 & \mathbf{D} &
 \end{array}
 \end{array}
 \tag{B.4}$$

To obtain this, we use that λ respects units, which means

$$\begin{array}{ccc}
 \begin{array}{c}
 \begin{array}{ccc}
 & C & \\
 e \nearrow & & \searrow G \\
 1 & \xrightarrow{e} & C' \\
 e \searrow & & \nearrow L \\
 & D &
 \end{array}
 & \equiv &
 \begin{array}{ccc}
 & C & \\
 e \nearrow & & \searrow G \\
 1 & \xrightarrow{\eta_F} & C' \\
 e \searrow & & \nearrow L \\
 & D &
 \end{array}
 \end{array}
 \quad (B.5)$$

and similarly for χ . Since η_G is an isomorphism, (B.4) follows if we can prove it after postcomposing with η_G ,

$$\begin{array}{ccc}
 \begin{array}{c}
 \begin{array}{ccc}
 & C & \\
 e \nearrow & & \searrow G \\
 1 & \xrightarrow{e} & C' \\
 e \searrow & & \nearrow L \\
 & D &
 \end{array}
 & \equiv &
 \begin{array}{ccc}
 & C & \\
 e \nearrow & & \searrow G \\
 1 & \xrightarrow{\eta_F} & C' \\
 e \searrow & & \nearrow L \\
 & D &
 \end{array}
 \end{array}
 \quad \equiv \quad
 \begin{array}{c}
 \begin{array}{ccc}
 & C & \\
 e \nearrow & & \searrow G \\
 1 & \xrightarrow{\eta_F} & C' \\
 e \searrow & & \nearrow L \\
 & D &
 \end{array}
 & \equiv &
 \begin{array}{ccc}
 & C & \\
 e \nearrow & & \searrow G \\
 1 & \xrightarrow{e} & C' \\
 e \searrow & & \nearrow L \\
 & D &
 \end{array}
 \end{array}$$

which proves the claim.

Proving compatibility with the multiplication

$$\begin{array}{ccc}
 \begin{array}{c}
 \begin{array}{ccc}
 C' \times C' & \xrightarrow{\otimes} & C' \\
 \downarrow L \times L & \searrow \mu_X & \downarrow L \\
 \begin{array}{ccc}
 & X \times X & \\
 u \times u \xrightarrow{\quad} & & \\
 & D \times D & \xrightarrow{\otimes} D
 \end{array}
 \end{array}
 & \equiv &
 \begin{array}{ccc}
 C' \times C' & \xrightarrow{\otimes} & C' \\
 \downarrow L \times L & \searrow \mu_L & \downarrow L \\
 \begin{array}{ccc}
 & L & \\
 & \xrightarrow{u} & \\
 & D \times D & \xrightarrow{\otimes} D
 \end{array}
 \end{array}
 \end{array}
 \quad (B.6)$$

The image contains two commutative diagrams, each preceded by an equivalence symbol \equiv .

The top diagram shows a square with vertices $C \times C$ (top-left), C (top-right), $D \times D$ (bottom-left), and D (bottom-right). Horizontal arrows are labeled \otimes . A curved arrow μ_F goes from $C \times C$ to D . A curved arrow F goes from $C \times C$ to $D \times D$. A curved arrow $F \times F$ goes from $D \times D$ to D . A curved arrow G goes from C to C' . A curved arrow λ goes from C to C' . A curved arrow L goes from C' to D . A curved arrow X goes from C' to D . A curved arrow u goes from C' to D .

The bottom diagram shows a square with vertices $C \times C$ (top-left), C (top-right), $D \times D$ (bottom-left), and D (bottom-right). Horizontal arrows are labeled \otimes . A curved arrow μ_G goes from $C \times C$ to C . A curved arrow $G \times G$ goes from $C \times C$ to $C' \times C'$. A curved arrow $\lambda \times \lambda$ goes from $C \times C$ to $C' \times C'$. A curved arrow $L \times L$ goes from $C \times C$ to $C' \times C'$. A curved arrow μ_L goes from $C' \times C'$ to D . A curved arrow $F \times F$ goes from $D \times D$ to D . A curved arrow G goes from C to C' . A curved arrow λ goes from C to C' . A curved arrow L goes from C' to D . A curved arrow X goes from C' to D . A curved arrow u goes from C' to D .

The natural transformation (B.2) is epic, so that $\lambda \times \lambda$, whiskered by $D \times D \rightarrow D$, can be canceled. μ_G is an isomorphism, so that it can be canceled as well. Finally $G \times G$ is essentially surjective, and therefore pre-whiskering by it can also be canceled. We are then left with (B.6).

Now suppose that η'_L and μ'_L give another monoidal structure on L . For λ to be monoidal, the equations (B.5) and (B.7) need to be satisfied. But now by (B.5) and the invertibility of η_G , we get $\eta'_L = \eta_L$. Similarly, by (B.7) and the fact that μ_G is an isomorphism, $\lambda \otimes \lambda$ is epic, and $G \times G$ is essentially surjective, we conclude that $\mu'_L = \mu_L$. ■

It may help to visualize these equations three-dimensionally, by interpreting every rewriting step as a globular 3-cell, and whiskering and composing these 3-cells so as to form a 3-dimensional pasting diagram. This way, (B.7) becomes a full cylinder, with the two caps formed by $\lambda \times \lambda$ and λ , and with the three multiplications forming the side surface. The same is true for equation (B.5), but with the $\lambda \times \lambda$ cap collapsed to a single point, so that one obtains a cone with λ on the base.

C. Convex combinations of metric spaces

In this section, we will establish that \mathbf{Met} is a pseudoalgebra for the simplex operad, which means that there is a notion of “convex combination” of metric spaces which behaves very similarly to convex combinations in a vector space. The n -fold uniform convex combination of a metric space X with itself is the power X^n from Section 3.1. This

motivates the definition of the latter, and in particular the rescaling of the metric by $\frac{1}{n}$, which may otherwise seem unnatural. Very informally, by Theorem 3.21, we can therefore think of the Wasserstein space PX as a uniform infinitary convex combination of X with itself.

C.1. THE SIMPLEX OPERAD. Standard simplices, or equivalently spaces of probability measures on finite sets, or also equivalently convex combinations of arbitrary finite arity, naturally form an operad.

C.2. DEFINITION. [e.g. [Lei11, BF14]] *The simplex operad Δ , or convex combination operad has a set of operations of arity n given by the $(n - 1)$ -simplex¹¹,*

$$\Delta^n := \left\{ (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n \mid \lambda_i \geq 0, \sum_i \lambda_i = 1 \right\} \tag{C.1}$$

The unit of the operad is the unique element $1 \in \Delta^1$. The composition is defined as

$$\begin{aligned} \Delta^n \times \Delta^{m_1} \times \dots \times \Delta^{m_n} &\longrightarrow \Delta^{m_1 + \dots + m_n} \\ (\nu, \lambda_1, \dots, \lambda_n) &\longmapsto (\nu_1 \lambda_{11}, \dots, \nu_1 \lambda_{1m_1}, \dots, \nu_n \lambda_{n1}, \dots, \nu_n \lambda_{nm_n}). \end{aligned}$$

The symmetry of the operad is given by permutations of the vertices of the simplex: for $\sigma \in S_n$,

$$\begin{aligned} S_n \times \Delta^n &\longrightarrow \Delta^n \\ (\sigma, \lambda) &\longmapsto (\lambda_{\sigma(1)}, \dots, \lambda_{\sigma(n)}). \end{aligned}$$

For example, every convex space (Definition 5.2) is an algebra of Δ in **Set**. It can be shown that Δ is the operad generated by binary operations c_λ with $\lambda \in [0, 1]$ subject to the relations encoded by the parametric symmetry and associativity of Definition 5.2. In other words, the concept of Δ -algebra in **Set** is a relaxation of the notion of convex space without the requirements 5.2(a)–(b).

C.3. PSEUDOMETRIC SPACES ARE A PSEUDOALGEBRA. We can turn **Met** into a *pseudotalgebra* (or *weak algebra*) in **Cat** of the simplex operad. To do so, it is more convenient to work with a similar category **PMet**, whose objects are *pseudometric spaces*, i.e. sets with a distance function satisfying $d(x, x) = 0$, symmetry, and the triangle inequality, but where $d(x, y) = 0$ may not imply $x = y$. As morphisms, we again choose short maps. In the following, we define a Δ -pseudotalgebra structure on **PMet**, where it is more explicit. Below, we will use this to turn **Met** into a Δ -pseudotalgebra as well. Since all operations turn out to map the full subcategory **CMet** to itself, we also have a pseudotalgebra structure on **CMet**.

In the following, in contrast to the main text, we use subscript notation X_i to refer to members of a sequence or family of spaces.

¹¹Usually Δ^n stands for the n -simplex of dimension n , but in this context it is less confusing to index the standard simplices by the number of vertices rather than by dimension.

C.4. DEFINITION. For any n and $\lambda \in \Delta^n$, the functor

$$k_\lambda : \mathbf{PMet}^n \rightarrow \mathbf{PMet}$$

takes a tuple of pseudometric spaces (X_1, \dots, X_n) to the pseudometric space $k_\lambda(X_1, \dots, X_n) \in \mathbf{PMet}$, where:

- The underlying set is the Cartesian product $X_1 \times \dots \times X_n$;
- The pseudometric is given by the convex combination of the pseudometrics,

$$d((x_1, \dots, x_n), (y_1, \dots, y_n)) := \lambda_1 d(x_1, y_1) + \dots + \lambda_n d(x_n, y_n). \quad (\text{C.2})$$

This strongly suggests that \mathbf{PMet} is indeed a pseudoalgebra of Δ in a canonical way. For the technical details, we use [Web15, Definition 4.1] as the definition of pseudoalgebra. First, (C.2) shows why it is important to work with pseudometric spaces, since this metric is guaranteed to be nondegenerate only when λ has full support; thus setting up the same structure on \mathbf{Met} directly would be more cumbersome, since then the underlying set would also have to vary with λ .

In order to construct the necessary coherence maps and prove the required equations, it is most convenient to use the fact that the Cartesian products make \mathbf{Set} into a symmetric monoidal category, which is the same thing as a pseudoalgebra of the terminal operad \mathbf{Comm} [Web15, Example 4.5]. (This gives an unbiased definition of symmetric monoidal category, in the spirit of [Lei04, Definition 3.1.1].) Thanks to the unique operad morphism $\Delta \rightarrow \mathbf{Comm}$, the category \mathbf{Set} becomes a pseudoalgebra of Δ as well, where the action by $\lambda \in \Delta^n$ is given by $(X_1, \dots, X_n) \mapsto X_1 \times \dots \times X_n$. This corresponds to the k_λ from Definition C.4 under the forgetful functor $\mathbf{PMet} \rightarrow \mathbf{Set}$. Since the latter functor is faithful, there is at most one way to lift the coherence isomorphisms to \mathbf{PMet} , where the required equations then automatically hold. Proving that such liftings exist boils down to showing that the coherence isomorphisms are isometries, but this is straightforward to see from the definition (C.2).

It remains to transport this pseudoalgebra structure to \mathbf{Met} . To this end, we first consider the category \mathbf{PMet}_\sim , which is the quotient of \mathbf{PMet} where $f, g : X \rightarrow Y$ are considered equivalent if $d(f(x), g(x)) = 0$ for all $x \in X$. Considering a metric space as a pseudometric space gives a fully faithful functor $\mathbf{Met} \rightarrow \mathbf{PMet}_\sim$. By identifying all points that have distance zero, every pseudometric space has a metric quotient to which it is isomorphic in \mathbf{PMet}_\sim , using the axiom of choice to construct a section of the quotient map. Therefore the functor $\mathbf{PMet}_\sim \rightarrow \mathbf{Met}$ is an equivalence of categories, and it is enough to construct a Δ -pseudoalgebra structure on \mathbf{PMet}_\sim . But this follows immediately from the fact that the functors k_λ from Definition C.4 respect the equivalence relation on morphisms and therefore descend to functors $\mathbf{PMet}_\sim \rightarrow \mathbf{PMet}_\sim$.

In summary, we have turned the category \mathbf{Met} into a pseudoalgebra of the simplex operad Δ . The n -ary convex combination with weights $\lambda \in \Delta^n$ of metric spaces X_1, \dots, X_n

is the metric space with underlying set $X_1 \times \dots \times X_n$, but where those X_i with $\lambda_i = 0$ are omitted, and metric (C.2). The power X^n is the special case where $\lambda = (\frac{1}{n}, \dots, \frac{1}{n})$ is the uniform distribution, since then one gets from C.2,

$$d((x_1, \dots, x_n), (y_1, \dots, y_n)) = \frac{1}{n} \sum_{i=1}^n d(x_i, y_i). \tag{C.3}$$

which is exactly the metric on the power X^n defined by 3.1.

Strangely, the category of finite probability spaces and measure-preserving maps, **FinProb**, is also a pseudoalgebra of Δ [BF14, Appendix B], but in a very different way, where a convex combination of spaces is based on the disjoint union of sets with the corresponding convex combination of probability measures, supported on these disjoint sets. At present we do not know whether to regard this as coincidental or deep.

C.5. INTERNAL ALGEBRAS AND THE MICROCOSM PRINCIPLE. The concept of monoid makes sense internally to any monoidal category. Similarly, commutative monoids can be considered internally to any symmetric monoidal category. More in general, there are situations where one expects an internal structure to be naturally defined within a categorified version of that same structure. This phenomenon has been called “microcosm principle” by Baez and Dolan [BD98], and they made this statement precise in terms of operads: if a category **C** is equipped with a pseudoalgebra structure for an operad O , then an internal algebra in **C** is specified by a lax morphism of pseudoalgebras $\mathbf{1} \rightarrow \mathbf{C}$. In the following, we will show that the internal Δ -algebras in **CMet** are closely related to the P -algebras of Section 5.

In our case, we have the simplex operad Δ , and the pseudoalgebra structure on **Met**, or equivalently on **PMet**_~. The terminal category $\mathbf{1}$ is also trivially an algebra for Δ . A lax morphism between these pseudoalgebras is then a functor $A : \mathbf{1} \rightarrow \mathbf{PMet}_\sim$, selecting an object $A \in \mathbf{PMet}_\sim$, and for each n and $\lambda \in \Delta^n$ a natural transformation

$$\begin{array}{ccc} \mathbf{1}^n & \xrightarrow{A^{\times n}} & \mathbf{PMet}_\sim^n \\ \downarrow ! & \swarrow c_\lambda & \downarrow k_\lambda \\ \mathbf{1} & \xrightarrow[A]{} & \mathbf{PMet}_\sim \end{array} \tag{C.4}$$

or equivalently a morphism $c_\lambda : k_\lambda(A, \dots, A) \rightarrow A$, satisfying certain coherence conditions which we specify below. Before doing so, it is instructive to look at what it means for c_λ to be short: the metric on $k_\lambda(A, \dots, A)$ is given by

$$d((x_1, \dots, x_n), (y_1, \dots, y_n)) = \sum_i \lambda_i d(x_i, y_i). \tag{C.5}$$

Thus shortness of c_λ means that

$$d(c_\lambda(x_1, \dots, x_n), c_\lambda(y_1, \dots, y_n)) \leq \sum_i \lambda_i d(x_i, y_i), \tag{C.6}$$

which is the metric compatibility inequality from Section 5 in the generalized form (5.9).

Following [Web15, Definition 4.10], the c_λ 's have to satisfy the following conditions in order to define a lax morphism:

- Unit condition. For $1 \in \Delta^1$, the composite

$$\begin{array}{ccc}
 & A & \\
 & \Downarrow \iota & \\
 \mathbf{1} & \xrightarrow{k_1(A)} & \mathbf{PMet}_\sim \\
 & \Downarrow c_1 & \\
 & A &
 \end{array} \tag{C.7}$$

must be equal to the identity transformation of A , where ι is one of the coherences of the pseudoalgebra structure of \mathbf{PMet} . So explicitly, this means that the map $c_1 : k_1(A) \rightarrow A$ must be inverse to the canonical isomorphism $A \rightarrow k_1(A)$. This is the same as the unit condition for convex spaces when defined as algebras of the convex combinations monad.

- Composition condition. We consider arities n with $\lambda \in \Delta_n$ and m_1, \dots, m_n with $\mu_i \in \Delta_{m_i}$, and write $m := \sum_i m_i$. The composite 2-cell

$$\begin{array}{ccccc}
 & & \mathbf{PMet}_\sim^m & \xrightarrow{\prod_i k_{\mu_i}} & \mathbf{PMet}_\sim^n & \\
 & \nearrow^{A^{\times m}} & \Downarrow \prod_i c_{\mu_i} & \nearrow^{A^{\times n}} & \downarrow k_\lambda & \\
 \mathbf{1}^m & \xrightarrow{\quad} & \mathbf{1}^n & & & \\
 & \searrow & \downarrow & \swarrow c_\lambda & & \\
 & & \mathbf{1} & \xrightarrow{A} & \mathbf{PMet}_\sim &
 \end{array} \tag{C.8}$$

has to be equal to the composite

$$\begin{array}{ccccc}
 & & \mathbf{PMet}_\sim^m & \xrightarrow{\prod_i k_{\mu_i}} & \mathbf{PMet}_\sim^n & \\
 & \nearrow^{A^{\times m}} & \Downarrow c_{\lambda \circ \mu} & \nearrow & \downarrow k_\lambda & \\
 \mathbf{1}^m & \xrightarrow{\quad} & \mathbf{1} & \xrightarrow{A} & \mathbf{PMet}_\sim & \\
 & \searrow & & & & \\
 & & & & &
 \end{array} \tag{C.9}$$

where all unlabelled morphisms are the obvious ones. This amounts to commuta-

tivity of

$$\begin{array}{ccc}
 k_\lambda(k_{\mu_1}(A, \dots, A), \dots, k_{\mu_n}(A, \dots, A)) & \xrightarrow{k_\lambda(c_{\mu_1}, \dots, c_{\mu_n})} & k_\lambda(A, \dots, A) \\
 \downarrow & & \downarrow c_\lambda \\
 k_{\lambda \circ \mu}(A, \dots, A) & \xrightarrow{c_{\lambda \circ \mu}} & A
 \end{array} \tag{C.10}$$

- Equivariance condition. Since all symmetries of $\mathbf{1}$ are trivial, this is basically only an invariance condition. Let $\sigma \in S_n$. The composite 2-cell

$$\begin{array}{ccccc}
 & & \mathbf{PMet}_\sim^n & & \\
 & \nearrow^{A \times n} & & \searrow^\sigma & \\
 \mathbf{1}^n & & & & \mathbf{PMet}_\sim^n \\
 \downarrow ! & \Downarrow c_\lambda & \searrow^{k_\lambda} & \Leftarrow & \swarrow^{k_{\sigma(\lambda)}} \\
 \mathbf{1} & \xrightarrow{A} & \mathbf{PMet}_\sim & &
 \end{array} \tag{C.11}$$

has to be equal to the composite

$$\begin{array}{ccccc}
 & & \mathbf{PMet}_\sim^n & & \\
 & \nearrow^{A \times n} & \parallel & \searrow^\sigma & \\
 \mathbf{1}^n & \xrightarrow{\sigma} & \mathbf{1}^n & \xrightarrow{A \times n} & \mathbf{PMet}_\sim^n \\
 \downarrow ! & \Leftarrow & \swarrow^{c_{\sigma(\lambda)}} & \Downarrow c_{\sigma(\lambda)} & \swarrow^{k_{\sigma(\lambda)}} \\
 \mathbf{1} & \xrightarrow{A} & \mathbf{PMet}_\sim & &
 \end{array} \tag{C.12}$$

where all unlabelled morphisms are the obvious ones. This amounts to commutativity of the following equivariance diagram,

$$\begin{array}{ccc}
 k_{\sigma(\lambda)}(A, \dots, A) & \xrightarrow{c_{\sigma(\lambda)}} & A \\
 \downarrow & & \parallel \\
 k_\lambda(A, \dots, A) & \xrightarrow{c_\lambda} & A
 \end{array} \tag{C.13}$$

where the left arrow corresponds to the coherence isomorphism associated to the pseudoalgebra structure of \mathbf{PMet}_\sim , i.e. to permutation of the factors of the underlying cartesian product.

Taken together, these conditions state precisely that the underlying set of A must be an algebra of the simplex operad, with the additional condition that the maps $c_\lambda : k_\lambda(A, \dots, A) \rightarrow A$ must be short. The latter is exactly the metric compatibility inequality in the generalized form (5.9). So by Theorem 5.4, there is a fully faithful functor from

P -algebras to internal Δ -algebras in \mathbf{PMet}_\sim . However, this functor is not essentially surjective, since the unitality and idempotency conditions 5.2(a)–(b) cannot be expressed in operadic terms. For example, we can choose any fixed $y \in A$ and define $c_\lambda(x_1, \dots, x_n) := y$ for all $\lambda \in \Delta^n$ of arity $n \geq 2$. This defines an internal Δ -algebra in \mathbf{PMet}_\sim which is not in the essential image of the forgetful functor from \mathbf{CMet}^P if $|A| \geq 2$.

In order to improve on this and to obtain a complete characterization of P -algebras as internal algebras, it may therefore be necessary to consider \mathbf{PMet}_\sim as a lax algebra of a suitable 2-monad. We have not done this yet.

Bibliography

- [AR01] J. Adámek and J. Rosický, *On sifted colimits and generalized varieties*, Theory Appl. Categ. **8** (2001), 33–53.
- [AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, ETH Zürich, Birkhäuser Verlag, 2005.
- [BD98] John C. Baez and James Dolan, *Higher-Dimensional Algebra III. n -Categories and the Algebra of Opetopes*, Advances in Mathematics **135** (1998), no. 2. [arXiv:q-alg/9702014](https://arxiv.org/abs/q-alg/9702014).
- [BF14] John C. Baez and Tobias Fritz, *A Bayesian Characterization of Relative Entropy*, Theory Appl. Categ. **29** (2014), no. 16, 421–456. [arXiv:1402.3067](https://arxiv.org/abs/1402.3067).
- [Bas15] Giuliano Basso, *A Hitchhiker’s guide to Wasserstein distances*, 2015. Available at: <http://n.ethz.ch>.
- [CF13] Valerio Capraro and Tobias Fritz, *On the axiomatization of convex subsets of Banach spaces*, Proc. Amer. Math. Soc. **141** (2013). [arXiv:1105.1270](https://arxiv.org/abs/1105.1270).
- [Dob06] Ernst-Erich Doberkat, *Eilenberg-Moore algebras for stochastic relations*, Inform. and Comput. **204** (2006), no. 12, 1756–1781.
- [Edg98] Gerald A. Edgar, *Integral, Probability, and Fractal Measures*, Springer, 1998.
- [Fre06] D. H. Fremlin, *Measure theory. Vol. 4*, Torres Fremlin, Colchester, 2006. Topological measure spaces. Part I, II, Corrected second printing of the 2003 original.
- [FS95] Peter Freyd and Ross Street, *On the size of categories*, Theory Appl. Categ. **1** (1995), No. 9, 174–178.
- [Fri09] Tobias Fritz, *Convex Spaces I: Definition and Examples*, 2009. [arXiv:0903.5522](https://arxiv.org/abs/0903.5522).
- [FP18] Tobias Fritz and Paolo Perrone, *A criterion for kan extensions of lax monoidal functors*, 2018. [arXiv:1809.10481](https://arxiv.org/abs/1809.10481).
- [FKM16] Soichiro Fujii, Shin-ya Katsumata, and Paul-André Melliès, *Towards a formal theory of graded monads*, Foundations of software science and computation structures, 2016, pp. 513–530.
- [FJ15] Robert Furber and Bart Jacobs, *From Kleisli Categories to Commutative C^* -Algebras: Probabilistic Gelfand Duality*, Log. Methods Comput. Sci. **11** (2015), no. 2. [arXiv:1303.1115](https://arxiv.org/abs/1303.1115).
- [Gir82] Michèle Giry, *A Categorical Approach to Probability Theory*, Categorical aspects of topology and analysis, 1982.
- [GL17] Jean Goubault-Larrecq, *Complete Quasi-Metrics for Hyperspaces, Continuous Valuations, and Previsions*, 2017. [arXiv:1707.03784](https://arxiv.org/abs/1707.03784).

- [Gud73] Stan Gudder, *Convex structures and operational quantum mechanics*, *Comm. Math. Phys.* **29** (1973), 249–264.
- [Jac17] Bart Jacobs, *From Probability Monads to Commutative Effectuses*, *Journ. of Logical and Algebraic Methods in Programming* (2017). In press. Available at <http://www.cs.ru.nl/B.Jacobs/PAPERS/probability-monads.pdf>.
- [Kei08] Klaus Keimel, *The monad of probability measures over compact ordered spaces and its Eilenberg-Moore algebras*, *Topology Appl.* **156** (2008), no. 2, 227–239.
- [Kou15] Seerp R. Koudenburg, *Algebraic Kan extensions in double categories*, *Theory and Applications of Categories* **30** (2015), no. 5. arXiv:1406.6994.
- [Kou] Seerp Roald Koudenburg, *A double-dimensional approach to formal category theory*. arXiv:1511.04070.
- [Law62] William Lawvere, *The category of probabilistic mappings*, 1962. Available at <https://ncatlab.org/nlab/files/lawvereprobability1962.pdf>.
- [Law73] ———, *Metric spaces, generalized logic and closed categories*, *Rendiconti del seminario matematico e fisico di Milano* **43** (1973).
- [Law86] ———, *Taking Categories Seriously*, *Revista Colombiana de Matematicas* **20** (1986).
- [Lei04] Tom Leinster, *Higher Operads, Higher Categories*, *London Mathematical Society Lecture Note Series*, vol. 298, Cambridge University Press, 2004. arXiv:math/0305049.
- [Lei11] ———, *An Operadic Introduction to Entropy*, 2011. *n-Category Café* blog post, golem.ph.utexas.edu/category/2011/05/an_operadic_introduction_to_entropy.html.
- [LW17] Rory B. B. Lucyshyn-Wright, *Functional distribution monads in functional-analytic contexts*, *Advances in Mathematics* **322** (2017), 806–860.
- [MPP16] Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin, *Quantitative Algebraic Reasoning*, *Proceedings of the Thirty first Annual IEEE Symposium on Logic in Computer Science (LICS 2016)*, 2016, pp. 700–709.
- [MT08] Paul-André Melliès and Nicolas Tabareau, *Free models of T-algebraic theories computed as Kan extensions*, 2008. hal.archives-ouvertes.fr/hal-00339331/document.
- [Pat12] Ross Paterson, *Constructing applicative functors*, *Mathematics of Program Construction*, 2012, pp. 300–323.
- [PP02] Gordon D. Plotkin and John Power, *Notions of Computation Determine Monads*, *International Conference on Foundations of Software Science and Computation Structures* (2002).
- [Sem73] Zbigniew Semadeni, *Monads and their Eilenberg-Moore algebras in functional analysis*, *Queen’s University, Kingston, Ont.*, 1973. *Queen’s Papers in Pure and Applied Mathematics*, No. 33.
- [Sto49] Marshall H. Stone, *Postulates for the barycentric calculus*, *Ann. Mat. Pura Appl.* (4) **29** (1949), 25–30.
- [vB05] Franck van Breugel, *The Metric Monad for Probabilistic Nondeterminism*, 2005. Available at <http://www.cse.yorku.ca>.
- [vBHMW05] Franck van Breugel, Claudio Hermida, Michael Makkai, and James Worrell, *An Accessible Approach to Behavioural Pseudometrics*, *Proceedings of the 32nd International Colloquium on Automata, Languages, and Programming (ICALP)*, 2005, pp. 1018–1030.

- [Vil09] Cédric Villani, *Optimal transport: old and new*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, 2009.
- [Web15] Mark Weber, *Operads as polynomial 2-monads*, Theory Appl. Categ. **30** (2015), no. 49, 1659–1712. [arXiv:1412.7599](https://arxiv.org/abs/1412.7599).
- [Web16] ———, *Algebraic Kan extensions along morphisms of internal algebra classifiers*, Tbilisi Mathematical Journal **9** (2016), no. 1. [arXiv:1511.04911](https://arxiv.org/abs/1511.04911).
- [Świ74] Tadeusz Świrszcz, *Monadic functors and convexity*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **22** (1974).

*Perimeter Institute for Theoretical Physics
Waterloo, ON, Canada*

*Max Planck Institute for Mathematics in the Sciences
Leipzig, Germany*

Email: tfritz@perimeterinstitute.ca
paolo.perrone@mis.mpg.de

This article may be accessed at <http://www.tac.mta.ca/tac/>

THEORY AND APPLICATIONS OF CATEGORIES will disseminate articles that significantly advance the study of categorical algebra or methods, or that make significant new contributions to mathematical science using categorical methods. The scope of the journal includes: all areas of pure category theory, including higher dimensional categories; applications of category theory to algebra, geometry and topology and other areas of mathematics; applications of category theory to computer science, physics and other mathematical sciences; contributions to scientific knowledge that make use of categorical methods. Articles appearing in the journal have been carefully and critically refereed under the responsibility of members of the Editorial Board. Only papers judged to be both significant and excellent are accepted for publication.

SUBSCRIPTION INFORMATION Individual subscribers receive abstracts of articles by e-mail as they are published. To subscribe, send e-mail to tac@mta.ca including a full name and postal address. Full text of the journal is freely available at <http://www.tac.mta.ca/tac/>.

INFORMATION FOR AUTHORS \LaTeX 2 ϵ is required. Articles may be submitted in PDF by email directly to a Transmitting Editor following the author instructions at <http://www.tac.mta.ca/tac/authinfo.html>.

MANAGING EDITOR. Robert Rosebrugh, Mount Allison University: rrosebrugh@mta.ca

TEXNICAL EDITOR. Michael Barr, McGill University: michael.barr@mcgill.ca

ASSISTANT TEX EDITOR. Gavin Seal, Ecole Polytechnique Fédérale de Lausanne: gavin_seal@fastmail.fm

TRANSMITTING EDITORS.

Clemens Berger, Université de Nice-Sophia Antipolis: cberger@math.unice.fr

Julie Bergner, University of Virginia: jeb2md@virginia.edu

Richard Blute, Université d' Ottawa: rblute@uottawa.ca

Gabriella Böhm, Wigner Research Centre for Physics: bohm.gabriella@wigner.mta.hu

Valeria de Paiva, Nuance Communications Inc: valeria.depaiva@gmail.com

Richard Garner, Macquarie University: richard.garner@mq.edu.au

Ezra Getzler, Northwestern University: getzler@northwestern.edu

Kathryn Hess, Ecole Polytechnique Fédérale de Lausanne: kathryn.hess@epfl.ch

Dirk Hoffman, Universidade de Aveiro: dirk@ua.pt

Pieter Hofstra, Université d' Ottawa: phofstra@uottawa.ca

Anders Kock, University of Aarhus: kock@math.au.dk

Joachim Kock, Universitat Autònoma de Barcelona: kock@mat.uab.cat

Stephen Lack, Macquarie University: steve.lack@mq.edu.au

F. William Lawvere, State University of New York at Buffalo: wlawvere@buffalo.edu

Tom Leinster, University of Edinburgh: Tom.Leinster@ed.ac.uk

Matias Menni, Conicet and Universidad Nacional de La Plata, Argentina: matias.menni@gmail.com

Ieke Moerdijk, Utrecht University: i.moerdijk@uu.nl

Susan Niefield, Union College: niefiels@union.edu

Robert Paré, Dalhousie University: pare@mathstat.dal.ca

Kate Ponto, University of Kentucky: kate.ponto@uky.edu

Jiri Rosicky, Masaryk University: rosicky@math.muni.cz

Giuseppe Rosolini, Università di Genova: rosolini@disi.unige.it

Alex Simpson, University of Ljubljana: Alex.Simpson@fmf.uni-lj.si

James Stasheff, University of North Carolina: jds@math.upenn.edu

Ross Street, Macquarie University: ross.street@mq.edu.au

Tim Van der Linden, Université catholique de Louvain: tim.vanderlinden@uclouvain.be

R. J. Wood, Dalhousie University: rjwood@mathstat.dal.ca